Depth-Profiling of Surface-Segregated Composition Spread Alloy Films

Gary Klein

June 15, 2010

Abstract

It is common for the components in a homogeneous metallic alloy to separate at its surface. This phenomenon is known as surface segregation. For example, under certain conditions, the surface layer of atoms in a $Cu_{85}Pd_{15}$ is pure copper, while the second layer of atoms is 50% copper. The bulk retains its 85/15 composition. A depth-profile characterizes the surface segregation by describing the atomic fraction of each element as a function of depth.

To study the effects that each alloy composition has on the degree of surface segregation, researchers have developed a composition spread alloy film, a material that contains all possible alloy compositions on a single substrate. We use angle-resolved x-ray photospectroscopy data collected from a copper-palladium composition spread alloy film to estimate the depth-profiles for sampled alloys using both a constrained linear regression and a full Bayesian model. We extend these models to predict the depth-profile for any alloy composition while allowing for measurement error given the novelty of the composition spread alloy film.

1 Introduction

It is common for the components in a homogeneous metallic alloy to separate at its surface. This phenomenon, known as *surface segregation*, results in a surface composition that can be quite different than the deeper depths of the alloy known as the bulk. For example, Newton, et al. showed that under certain conditions, the surface layer of atoms in a 85% copper, 15% palladium alloy is pure copper, while the second layer of atoms is 50% copper.[18][17] See Figure 1. Hereafter, we will use the standard notation for binary alloys: Y_xZ_{100-x}, where Y and Z are the symbols of the elements found in the alloy and x and

$100 - x$ are the atomic fractions for those respective elements. The above 85% copper, 15% palladium alloy is denoted $Cu_{85}Pd_{15}$. The degree of surface segregation is affected by a number of factors including: the component metals, including their proportions in the alloy, their atomic size, and how they interact; any foreign components that may have adsorbed into the system; and environmental factors including temperature and pressure.

Surface segregation can be advantageous and thus has practical uses. Industrial uses of surface segregation include catalysis, where a surface-segregated bimetallic alloy is a superior catalyst to its monometallic counter-parts.[6] Of particular interest is that of *functional coating*, where surface segregation induces a protective barrier of the alloy if the component metal that migrates to the surface is particularly resistant to corrosion.

An important application of functional coatings is in the creation of palladium-based membranes. A palladium membrane can be used to separate hydrogen from any one of a number of hydrogen mixtures, including hydrogen-carbon monoxide. By exploiting the surface-segregation of a $Pd_{70}Cu_{30}$ alloy, Miller, et al. use the predominantly copper surface as a functional coating to protect a palladium-membrane interior from hydrogen sulfide, H_2S, a corrosive gas commonly found in fossil fuel plants.[15][16] Adhikari, et al. also report the use of copper as a functional coating and further note that a copper-palladium alloy in the correct proportions has a higher hydrogen permeation rate and is more thermally and chemically stable than palladium alone, allowing the alloy to separate more hydrogen than a pure palladium membrane.[1]

Understanding the impact that different alloy compositions has on the degree of segregation is not trivial. Previous research required scientists to create a series of alloys with specific compositions then analyze them with a preferred method to investigate their surface and bulk compositions. To advance this process, Gellman, Miller and their group in the Department of Chemical Engineering at Carnegie Mellon University have developed a method to create a *Composition Spread Alloy Film* (CSAF). A CSAF is a multimetallic substance that contains all possible combinations of an alloy, in this case, Pd_XCu_{100-X}, on a single substrate. Figure 2 depicts a copper-palladium CSAF, where the ratio of copper to palladium increases continuously and linearly from 0:1 to 1:0, that is, every alloy composition is represented including pure palladium ($Pd_{100}Cu_0$) and pure copper (Pd_0Cu_{100}) on the left and right sides of the material, respectively.

Angle-resolved x-ray photospectroscopy (ARXPS) is a technique to extract compositional information at near surface depths. The result is a *depth-profile* which gives the atomic concentration of each element as a function of depth. Depth-profiles can then be used to examine surface segregation. With ARXPS data collected from numerous locations on a copper-palladium CSAF, **the primary goal of this research is to build a model which predicts the depth-profile for any composition of a copper-palladium alloy.** Our initial model will be a high-order polynomial fixed-effects regression model which includes the alloy composition as a predictor. If an examination of residuals shows a lack of fit, then the data may be fit using splines.

If the gradient in the alloy compositions of the CSAF was exactly linear, then we

would know *which* alloy composition was sampled if we know *where* on the CSAF the data was collected. However, because of the variation that can occur when creating a CSAF, it is unclear as to the nature of this gradient. Consequently, although we can be fairly certain as to where on the CSAF the data is collected, we cannot be sure as to which alloy composition was sampled. Therefore, **the second goal of this research is to extend this regression model to allow for measurement error of the alloy compositions in the CSAF**.

We also note that estimating the depth-profiles from raw ARXPS data is not trivial. Many of the current methods are based on older ARXPS machines which were limited in the amount of data they could collect. Consequently, many of these methods are iterative and may be time-consuming. Gellman, Miller and their group have acquired a Theta-Probe®, a new type of an ARXPS machine capable of rapidly collecting numerous sources of data. Because of the large amount of additional data, we believe a constrained least squares approach and/or a Bayesian model will be at least as accurate as the standard methods in estimating depth-profiles. **Our third goal is to compare one current method, maximum entropy, with our linear and Bayesian models. Our fourth goal is to use our estimates and variances, and incorporate it into the measurement error model as described above.**

2 Angle-Resolved X-Ray Photospectroscopy

2.1 Electron Binding Energy

Binding energy is the amount of energy necessary to remove an electron from its orbital.[20] An electron orbital is the probability distribution of the location of a single or pair of electrons around its nucleus. The shape and location of the orbitals are distinguished by their three quantum numbers: the *principal* number which gives its main location or shell (also characterizes its overall size) - this is given as $n = 1, 2, 3, \ldots$; the *azimuthal* number which characterizes its distribution shape - this is given as $l = 0, 1, 2, 3, \ldots$, or more commonly expressed as $s, p, d, f, \ldots$ (with $l \leq n-1$); and its *magnetic* number which gives the orientation of the distribution shape - this is given as $-l, -l+1, \ldots, l-1, l$. For example, the 2p orbital is in the second shell and has "shape" p, or $l = 1$ (p is usually described as dumbbell shaped). This "dumbbell" can be oriented in three different directions, -1, 0, and 1.[30] Electrons fill in the orbitals in an orderly manner following *Madelung's rule*, where an electron occupies an orbital in order of increasing $n + l$. When two orbitals have the same value of $n+l$, the electron occupies the orbital with the smaller n.[33] The order for the first few orbitals is thus:

$$1s, 2s, 2p, 3s, 3p, 4s, 3d, 4p, \ldots$$

The binding energy for any electron is largely a function of the number of protons of the

atom as well as the location of the electron, i.e., which orbital the electron occupies.[1][20] Because of the attractive properties of protons and electrons, the more protons the atom has, the more energy necessary to remove the electron. Similarly, the closer the electron is to the nucleus and thus to the protons, the more binding energy the electron has. For example, an electron in the $3s$ orbital will have less binding energy than an electron in the $2p$ orbital, since the $2p$ electron is closer to the atom's nucleus. Similarly, a $2p$ electron in palladium will have a higher binding energy than a $2p$ electron in copper because of palladium's additional 17 protons.

Because each element has a specific number of protons and due to the location difference between orbitals, the binding energy for an electron is unique given its element and orbital.[30] For example, the binding energies for palladium electrons residing in the $2p\frac{1}{2}$, $2p\frac{3}{2}$, $3d\frac{3}{2}$, and $3d\frac{5}{2}$ orbitals are 3330, 3173, 340.5, and 335.2 eV, respectively. For copper electrons in the $2p\frac{1}{2}$ and $2p\frac{3}{2}$ orbitals, the binding energies are 952.3 and 932.7 eV.[32][31] The two binding energies seen in a single orbital are from the two electrons that reside in that orbital, with the difference in energies given by the interaction between the electron spin and the orbital, denoted by the fraction given after the azimuthal number.[30]

2.2 X-ray Photospectroscopy

If a photon hits an electron, that electron, known as a *photoelectron*, can be removed from its orbital, provided that the photon's energy is greater than the binding energy of the electron. Thus, a large number of photons with sufficient energy in an x-ray can be used to remove a group of electrons from a substance. If it is assumed that the x-ray has constant energy given by $h\nu$ (e.g., for an Al K_α x-ray, $h\nu = 1486.6$eV), then the kinetic energy of the removed electron will be given by:

$$\mathrm{KE} = h\nu - \phi - \mathrm{BE} \tag{1}$$

where KE is the kinetic energy of the electron, ϕ is a constant given by a work function of the x-ray and analyzer, and BE is the binding energy of the electron. It follows that the kinetic energy of the removed electrons also acts like a signature, identifying the element and the shell the electron once occupied.[30]

An electron energy analyzer such as a concentric hemispheric analyzer can be used to determine the kinetic energies of the removed electrons and count the number of electrons at those kinetic energies. The result is a spectrum, with either kinetic energy or binding energy on the horizontal axis and electron counts on the vertical axis. By examining the spectrum of a substance, we can identify the elements present in the substance by determining the binding/kinetic energies associated with large electron counts (smaller

[1] Other factors can substantially effect binding energy as well, including the interaction between the electron spin and the shape of the orbital it occupies, as well as other elements present, especially for electrons in the outer valence shells.

counts are usually noise). The counts are also known as the *intensity, I,* at a given kinetic or binding energy. See Figure 3(a). Further, by calculating the ratios of the intensities for each element, we can quantify the atomic fractions of each element in the sample.

The process of removing electrons through the use of an x-ray, then identifying the present elements is sometimes referred to as *electron spectroscopy chemical analysis* (ESCA) or *x-ray photospectroscopy* (XPS).[30] It should be noted that the binding energy of valence electrons, those electrons on the outermost shells of an atom which are capable of bonding with the valence electrons of other atoms, are not constant and depend not only the element, but on the local chemical environment, e.g., the properties of the atom in which the electron is bound. Therefore, many of the XPS techniques used to identify elements are only concerned with core electrons, those electrons located in the inner shells of the atom. Typically, the specific electrons selected for XPS analysis can vary for different elements but will have the same binding energy order of magnitude.

Although the photons from an x-ray can penetrate numerous atomic layers, the probability that a photoelectron will travel through atomic layers to the surface of a substance without losing energy decays exponentially with how deep within the substance the electron lies. This is because the electron loses energy each time it collides with another electron shell. This concept is characterized by the attenuation length, λ, of an electron, which is determined by a variety factors including the size of the atoms in the substance and the kinetic energy of the electron. The parameter λ is defined as the mean distance an electron travels before its kinetic energy decays to e^{-1} of its initial value. Attenuation length is typically calculated using the Seah and Dench's formula[22]:

$$\lambda = 538aKE^2 + 0.41\sqrt{a^3 KE}$$

where KE is the electron kinetic energy and a^3 is the volume in nm of one atom in a solid state.

Using the attenuation length and the *Beer-Lambert Law* , we can estimate the percentage of the spectral intensity that is given by the electrons at each depth. The Beer-Lambert Law is given by:

$$I = I_0 \exp\left\{\frac{-d}{\lambda}\right\} \tag{2}$$

where I is the intensity of electrons, d is depth within the sample orthogonal to the surface and I_0 is the intensity of electrons of an infinitely thick sample. Figure 5a, shows a standard setup where the electron energy analyzer is parallel to the surface of the sample. Under this setup, the Beer-Lambert Law shows that 63% of the electron intensity found in the sample's spectrum comes from within one attenuation length, λ of the surface, 86.5% comes from within two attenuation lengths, and 95% comes from within three attenuation lengths (Figure 5b).

Without experimental noise, there should be a single spike in the spectrum only at the binding energy that coincides with an element's electron and no counts elsewhere. However,

the XPS results from a copper-palladium alloy as seen in Figure 3(b) and (c) shows wide peaks around the two binding energies as well as low-level background noise of smaller electron counts away from the binding energies. The wide peaks are a consequence of a number of factors, including the time in which the gap created by the emitted photoelectron is filled by another electron, variance in x-ray energy which is taken as a constant in Eq. (1), and analyzer/experiment effects. The result is a line shape that can be approximated by a convolution of a Cauchy (Lorentzian) and Gaussian curve. See Ratner and Castner for details.[20]

The background noise comes from a variety of sources. The first is from those photoelectrons that escaped from the sample's interior but lost energy due to collisions, or inelastic scattering, with other atoms on their way to the surface. Further spectral counts are from photoelectrons from outer shells, either excited by a photon or by a photoelectron from an inner shell. This latter type is known as a "shake-off satellite." Alternatively, "shake-up satellites" result from the loss of energy of the core photoelectron when it does lose energy to a valence electron, whether that valence electron gained enough energy to be emitted itself (shake-off satellite) or just enough to be moved to a higher energy level. Further, when an outer-shell electron fills in the hole left by of an emitted inner-photoelectron, energy is released and can be transferred to another outer electron which can then escape from its orbital. This is known as an Auger electron. Photoelectrons can also interact with banded valence electrons giving a discrete energy loss, known as a plasmon loss. Finally, background counts can arise from x-rays, either with impurities in the x-ray's anode (known as x-ray ghosts), or lower energy photons resulting from unlikely electron transitions in the anode (x-ray satellites). For details, again see Ratner and Castner[20], Briggs[2], and Watts.[30]

The intensity or counts used for quantification from XPS data is usually the area under the fitted peak after the background is subtracted. Although there are many algorithms to both fit an XPS peak and/or subtract a background, the preferred methods are to use a convolution of a Lorentzian and Gaussian curve to fit the peak, and to use the "Smart" method to subtract the background.[3] The Smart method is a modified Shirley method, an iterative method which adjusts the intensity of the background until it is proportional to the peak area.[23][26] The results can be seen in Figure 4.

2.3 Angle-Resolved X-ray Photospectroscopy

Using the Beer-Lambert Law, scientists have developed *angle-resolved* x-ray photospectroscopy. By increasing the angle, θ of the analyzer normal to the surface (Figure 5c), electrons at a depth d within the surface of the sample must travel $\frac{d}{\cos\theta}$ to reach the surface with their original kinetic energy and be detected by the analyzer. The Beer-Lambert Law can then be given by:

$$I(\theta) = I_0 \exp\left\{\frac{-d}{\lambda \cos\theta}\right\}. \tag{3}$$

Clearly, the effect of varying the angle of the analyzer affects the percentage of the spectrum intensity given by electron depth, with larger angles corresponding to intensities that are given mostly by the surface of the sample.

3 Depth-Profile Estimation

3.1 Brief Literature Review

With collected data and the Beer-Lambert Law, researchers can develop a *depth-profile* of the sample. A depth-profile gives the concentration(s) of each element in a sample at selected depths. See Figure 6.

There are various methods of estimating the depth-profile from angle-resolved data. The first group of methods are based on the model:

$$I_\ell(\theta) = k \int_0^\infty \rho(d)\, c_\ell(d) \exp\left\{\frac{-d}{\lambda_\ell \cos\theta}\right\} \mathrm{d}d \tag{4}$$

where k is a constant based on instrumentation, $\rho(d)$ is the total mass density of the sample at depth d, and $c_\ell(d)$ is the concentration depth-profile of element ℓ (subject to the constraint $\sum_\ell c_\ell(d) = 1$ for all d). Monte Carlo or numerical inversion methods, usually the Laplace transform, are then applied to solve for $\rho(d)\, c_\ell(d)$. Cumpson notes that these methods, like many inversion methods, are highly unstable, with estimates varying widely for small errors in measurements.[5]

Other methods based on this model have avoided this problem by simplifying $\rho(d)\, c_\ell(d)$ using additional and possibly spurious assumptions. For example, Fischer et al. suppose specific geometries, setting the depth-profile to be trapezoidal and solving for various parameters of the trapezoid (see Figure 7) using a least squares method, minimizing:

$$\sum_j \sum_\ell \left(I_\ell^{\mathrm{calc}}(\theta_j) - I_\ell^{\mathrm{obs}}(\theta_j)\right)^2 \tag{5}$$

over each angle θ_j and each element ℓ.[7] I_ℓ^{calc} and I_ℓ^{obs} are the calculated and observed intensities, respectively. Grabherr et al. as well as Siuda had a similar approach, but assume the concentration profile follows a polynomial.[10][24]

Another widely used group of methods is based on what is sometimes referred to as a *multilayer model*. This model assumes that the sample is composed of discrete atomic layers of equal thickness. See Figure 8. The intensity, $I_\ell(\theta)$ of element ℓ from angle θ is calculated by a weighted sum as:

$$I_\ell(\theta) = k_\ell \left[n_{\ell,1} + n_{\ell,2} T_\ell(\theta) + n_{\ell,3} T_\ell(\theta)^2 + n_{\ell,4} T_\ell(\theta)^3 + \ldots\right] \tag{6}$$

where $n_{\ell,i}$ is the atomic fraction of element ℓ at layer i, the entirety of which makes up the concentration depth profile; $T_\ell(\theta) = \exp\left\{\frac{-t}{\lambda_\ell \cos\theta}\right\}$ is the *layer transmission function* given by the Beer-Lambert Law with t the thickness of each layer; k_ℓ is a constant such that for a pure substance, $I_\ell(\theta) = 1$. Therefore k_ℓ is based on few factors including element-specific information such as the cross-section of the atom, the number of layers in the multi-layer model, and instrument/experiment factors such as time the x-ray is used on the sample, the size of the area used in the analysis, and the *pass energy* in the analyzer. The pass energy is the window of energy accepted into each energy bin of the spectrum. For example, any electron with a binding energy between 860 and 900 eV will be counted in the 880 eV spectrum if the pass energy of the analyzer is set to 40 eV. It follows that larger pass energies correspond to more data, but wider peaks (more noise).

To solve for $n_{\ell,i}, i = 1, 2, 3, \ldots$, researchers have largely relied on methods based on least squares or regularization.[19][11][29][4] However, Smith and Livesy and later Scorciapino et al. used a method known as *maximum entropy*.[13][21] Maximum entropy defines two quantities:

$$S = \sum_\ell \sum_i n_{\ell,i} - m_{\ell,i} - n_{\ell,i} \log\left(\frac{n_{\ell,i}}{m_{\ell,i}}\right) \tag{7}$$

and

$$C = \sum_j \sum_\ell \frac{\left(I_\ell^{\text{calc}}(\theta_j) - I_\ell^{\text{obs}}(\theta_j)\right)^2}{\sigma_{\ell,j}^2}. \tag{8}$$

S in Eq. (7) is the entropy, a measure of the lack of information we have in the system. $n_{\ell,i}$ is again, the proportion of element ℓ in layer i. $m_{\ell,i}$, however, is our initial guess of that proportion (prior to any data collection), allowing the researcher to impart some prior knowledge she may have of the system. Because the data is noisy, the goal is to maximize the entropy, S, i.e., allow for the greatest amount of missing information. This is equivalent to ensuring that there is no overfitting of the data and fitting a structure that may be a consequence of a noisy measurement.

However, S is only maximized under the constraint that the estimated depth profile $n_{\ell,i}$ is consistent with the data. This is satisfied by minimizing C in Eq. (8). Here, $\sigma_{\ell,j}^2$ is the variance of the jth measurement of element ℓ. We note that C follows a χ^2 distribution and $C \leq N$, the number of independent observations.

Livesey and Smith in a later paper discussed some solutions to maximizing S while minimizing C. This is typically done either by using an iterative approach put forth by Skilling and Bryan or alternatively by maximizing the weighted sum of Eqs. (7) and (8)

$$Q = \alpha S - \frac{1}{2} C, \tag{9}$$

where α is reduced until $C = N$.[25][14]

In the same paper, Livesey and Smith also showed the solution as an application of Bayesian inference, where

$$\mathbb{P}\left(n_{\ell,i}|D\right) = \frac{\mathbb{P}\left(n_{\ell,i}\right)\mathbb{P}\left(D|n_{\ell,i}\right)}{\mathbb{P}\left(D\right)}$$

where $\mathbb{P}\left(n_{\ell,i}|D\right)$ is the probability of the depth-profile $n_{\ell,i}$ given the data, D; $\mathbb{P}\left(n_{\ell,i}\right)$ is the prior distribution of the profile; $\mathbb{P}\left(D|n_{\ell,i}\right)$ is the probability of measuring the data D as observed given the depth-profile $n_{\ell,i}$; and $\mathbb{P}\left(D\right)$ is the prior probability of measuring D.

If the likelihood function $\mathbb{P}\left(D|n_{\ell,i}\right)$ can be derived from D by:

$$\begin{aligned}
\mathbb{P}\left(D|n_{\ell,i}\right) &= \exp\left\{\frac{-\chi^2}{2}\right\} \\
&= \exp\left\{-\sum_j \sum_\ell \frac{\left(I_\ell^{\text{calc}}\left(\theta_j\right) - I_\ell^{\text{obs}}\left(\theta_j\right)\right)^2}{\sigma_{\ell,j}^2}\right\} \text{ (from Eq. (8))}
\end{aligned}$$

and the prior distribution of the profile $n_{\ell,i}$ can be given by:

$$\begin{aligned}
\mathbb{P}\left(n_{\ell,i}\right) &\propto \exp\left\{\alpha S\left[n_{\ell,i}\right]\right\} \\
&= \exp\left\{\alpha \sum_\ell \sum_i n_{\ell,i} - m_{\ell,i} - n_{\ell,i}\log\left(\frac{n_{\ell,i}}{m_{\ell,i}}\right)\right\} \text{ (from Eq. (7)).}
\end{aligned}$$

then letting $Q = \log \mathbb{P}\left(n_{\ell,i}|D\right)$ gives us back our original maximum entropy equation, Eq. (9): $Q = \alpha S - \frac{1}{2}C$.

Although the most probable α can be found by maximizing the joint probability $\mathbb{P}\left(\alpha n_{\ell,1}|D\right)$, Gull and Skilling determined an analytical solution for the optimal value of α. Again, see Livesey and Smith for details.

Livesey and Smith pointed out that the Bayesian analysis of the maximum entropy data provide numerous benefits. The most important are a) it allows the calculation of credible intervals for the constructed depth profile; b) the amount of noise in the data can be determined; and c) the value of α that is found "gives a closer fit to the data than the historic fitting criterion [where] the weighted sum of squares error...equals the number of independent measurements."

3.2 Numerous-Angle Approaches

Common instruments that perform angle-resolved x-ray photospectroscopy require the researcher to manually change the angle of the sample with respect to the analyzer to gather angle-resolved data. Because of these manual procedures as well as the secondary processes instrumental in collecting data including maintaining an ultra-high vacuum in

the chamber where the sample is located, the number of angles and the amount of data that can be collected is limited.

Gellman, Miller and their group in the Department of Chemical Engineering at Carnegie Mellon University have acquired a Theta-Probe®. This ARXPS device manufactured by ThermoFisher Scientific uses a magnetic field to draw the electrons into as many as 96 different angle regions simultaneously, allowing for greater data acquisition in less time.[28] See Figure 9.

3.2.1 Constrained Least Squares

Because of the large amount of data that can be collected at the numerous angles, we believe a constrained least-squares approach based on the multilayer model as described above will give us reasonable estimates of depth profiles.

Restating Eq. (6) in matrix notation for a two element, M layer model with a t layer thickness:

$$\mathbf{I} = \begin{bmatrix} \mathbf{I_1} \\ \mathbf{I_2} \end{bmatrix} = \begin{bmatrix} k_1\mathbf{T_1}\mathbf{N_1} \\ k_2\mathbf{T_2}\mathbf{N_2} \end{bmatrix} \tag{10}$$

k_ℓ is still the element/instrumentation constant; $\mathbf{I}_\ell$ is a vector of length p of the measured intensities for element ℓ at various angles (we assume p is the same for all elements):

$$\mathbf{I}_\ell = \begin{bmatrix} I_\ell(\theta_1) \\ I_\ell(\theta_2) \\ \vdots \\ I_\ell(\theta_p) \end{bmatrix} \tag{11}$$

$\mathbf{T}_\ell$ is a $p \times M$ matrix with columns corresponding to the transmission layer functions for each layer given by the Beer-Lambert Law and rows corresponding to the p measurements:

$$\mathbf{T}_\ell = \begin{bmatrix} 1 & \exp\left\{\frac{-t}{\lambda_\ell \cos\theta_1}\right\} & \exp\left\{\frac{-2t}{\lambda_\ell \cos\theta_1}\right\} & \exp\left\{\frac{-3t}{\lambda_\ell \cos\theta_1}\right\} & \cdots & \exp\left\{\frac{-(M-1)t}{\lambda_\ell \cos\theta_1}\right\} \\ 1 & \exp\left\{\frac{-t}{\lambda_\ell \cos\theta_2}\right\} & \exp\left\{\frac{-2t}{\lambda_\ell \cos\theta_2}\right\} & \exp\left\{\frac{-3t}{\lambda_\ell \cos\theta_2}\right\} & \cdots & \exp\left\{\frac{-(M-1)t}{\lambda_\ell \cos\theta_2}\right\} \\ \vdots & & \ddots & & & \vdots \\ 1 & \exp\left\{\frac{-t}{\lambda_\ell \cos\theta_p}\right\} & \exp\left\{\frac{-2t}{\lambda_\ell \cos\theta_p}\right\} & \exp\left\{\frac{-3t}{\lambda_\ell \cos\theta_p}\right\} & \cdots & \exp\left\{\frac{-(M-1)t}{\lambda_\ell \cos\theta_p}\right\} \end{bmatrix} \tag{12}$$

and $\mathbf{N}_\ell$ is a vector of the concentrations for the M layers:

$$\mathbf{N}_\ell = \begin{bmatrix} n_{\ell,1} \\ n_{\ell,2} \\ \vdots \\ n_{\ell,M} \end{bmatrix}. \tag{13}$$

$\mathbf{T}_\ell$ is clearly fixed since the measurement angle is known and t, the layer thickness, is given. Because $\mathbf{I}_\ell$ is measured, we can thus estimate the depth profile by solving for $\mathbf{N_1}$ and $\mathbf{N_2}$ by minimizing the constrained sum of squares:

$$\underset{\mathbf{N_1},\mathbf{N_2}}{\operatorname{argmin}} \left[\mathbf{I} - \begin{bmatrix} k_1\mathbf{T_1N_1} \\ k_2\mathbf{T_2N_2} \end{bmatrix}\right]' \left[\mathbf{I} - \begin{bmatrix} k_1\mathbf{T_1N_1} \\ k_2\mathbf{T_2N_2} \end{bmatrix}\right],$$

under the equality constraint $\mathbf{N_1} + \mathbf{N_2} = \mathbf{1}$ and the inequality constraint $\mathbf{0} \leq \mathbf{N_1} \leq \mathbf{1}$ (it follows that $\mathbf{0} \leq \mathbf{N_2} \leq \mathbf{1}$).

We can reparameterize the linear constraints as

$$\mathbf{HN} - \mathbf{C} = \mathbf{0}$$

where $\mathbf{H}$ is the partitioned matrix, $[\mathbf{I_m}\mathbf{I_m}]$, $\mathbf{I_m}$ being the identity matrix with m columns and rows, $\mathbf{N} = [\mathbf{N_1}\mathbf{N_2}]$, and $\mathbf{C}$ is a column vector of $2m$ 1s. For example, for a three layer model with two elements,

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{N} = \begin{bmatrix} n_{1,1} \\ n_{1,2} \\ n_{1,3} \\ n_{2,1} \\ n_{2,2} \\ n_{2,3} \end{bmatrix}$$

and

$$\mathbf{C} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

The general solution to $\mathbf{N}$ with only linear constraints is:

$$\hat{\mathbf{N}} = \mathbf{A^{-1}}\left[\mathbf{kT'I} - \left[\mathbf{H'}\left(\mathbf{HA^{-1}H'}\right)^{\mathbf{-1}}\left[\mathbf{HA^{-1}kT'I} + \mathbf{HA^{-1}H'C} - \mathbf{C}\right]\right] + \mathbf{H'C}\right] \tag{14}$$

where $\mathbf{A} = \mathbf{H'H} + \mathbf{kT'Tk}$. The variance-covariance of $\hat{\mathbf{N}}$ is

$$\operatorname{VAR}\left(\hat{\mathbf{N}}\right) = \mathbf{G}'\sigma^2\mathbf{G} \tag{15}$$

where $\mathbf{G} = \left(\mathbf{A^{-1}kT'} - \mathbf{A^{-1}H'}\left(\mathbf{HA^{-1}H'}\right)^{\mathbf{-1}}\mathbf{HA^{-1}kT'}\right)$ and $\sigma^2 = \operatorname{VAR}(\mathbf{I})$. Further, under the Karush-Kuhn-Tucker conditions, the necessary conditions for an optimal solution

under constraints, if the optimal solution to a parameter lies outside the inequality constraints, the constraint is binding and can therefore be set equal to the bound using an equality constraint.[12]

3.2.2 Bayesian Analysis

We believe a Bayesian model may also be appropriate here for reasons to be explained in the following section. We again follow the multilayer model for an alloy and assume the observed intensities, I, given the layer concentrations, n, the element, l, and the angle of measurement, θ follow independent normal distributions from the following model:

$$
\begin{aligned}
I|n,\ell,\theta &\sim N\left(\mu_{n,\ell,\theta},\sigma^2\right) && (16)\\
\frac{1}{\sigma^2} &\sim \text{Gamma}\left(\gamma_1,\gamma_2\right) && (17)\\
\mu_{n,\ell,\theta} &= k_\ell \sum_i^m n_{\ell,m} T_\ell\left(\theta\right)^{i-1} && (18)\\
n_{\ell,m} &\sim \text{Beta}\left(\alpha_m,\beta_m\right) && (19)\\
\alpha_m &\sim \text{Gamma}\left(\alpha_{0.1},\alpha_{0.2}\right) && (20)\\
\beta_m &\sim \text{Gamma}\left(\beta_{0.1},\beta_{0.2}\right) && (21)
\end{aligned}
$$

where $\gamma_1, \gamma_2, \alpha_{0.1}, \alpha_{0.2}\beta_{0.1}$, and $\beta_{0.2}$ are fixed, and again we constrain $n_{1,m} = 1 - n_{2,m}$. Our model for $\mu_{n,\ell,\theta}$ from (28) follows directly from Eqn. (6) above. And since $n_{\ell,m}$ is the atomic concentration or percent of element ℓ in layer m, we assume that each layer is an independent sample from a Beta distribution. Because we don't assume any dependence of concentrations between each layer, we assume that the α_m and β_m hyperparameters are independent samples from a Gamma distribution.

3.3 Model Comparison

Using simulated noisy intensities with underlying known depth-profiles, the estimates and confidence/credible intervals of computed concentrations from the two numerous-angle approaches will be compared to that given by the maximum entropy method.[2] Although the maximum entropy method can be computed using the **Avantage** software which comes standard with the Theta-Probe, we believe it uses the Skilling and Bryan approach rather than the Bayesian method and therefore does not provide any estimates of error.[27] We are in the process of creating software that computes the maximum entropy using the Bayesian method so a credible interval can be calculated. Further, due to the number of

[2]At this time, we are not clear how to characterize the noise from the intensities. However, we believe this can be resolved with repeated experiments.

depth-profiles to be estimated, we will investigate the time of computation for the different methods.

It should also be noted that we have not seen any literature regarding the choice of M and t, the number of layers and the thickness of each layer. We note that the Beer-Lambert Law implies that 99.9% of the intensity is given by electrons that are within seven attenuation lengths of the surface when the analyzing angle is orthogonal to the sample surface. So although M and t are somewhat arbitrary, we believe a reasonable choice of M and t to approximate the multilayer model are such that $t \times M \approx 7\lambda_\ell$, with specific M and t to be chosen using a Bayes factor for the Bayesian and maximum entropy approach and/or AIC for the numerous angle approach.

4 Composition Spread Alloy Film Depth-Profile Prediction

4.1 Fixed-Effects Model

Recall that a depth-profile gives the concentration(s) of each element in a sample at any (shallow) depth. To estimate depth-profiles using either the maximum entropy or the two numerous-angle approaches, we used a multilayer model which assumes that the composition of the sample is comprised of discrete atomic layers of equal thickness. Although we believe this model to be adequate, depending on the number of layers estimated, it may only *approximate* the actual structure of the sample.

To find what we believe to be a more realistic depth-profile that provides concentrations of the elements for any depth at or just below the surface of the sample, we assume the discrete estimates are samples from an underlying smooth continuous depth-profile. That is, we wish to find some smooth, continuous function f that fits the discrete points estimated from our numerous-angle methods, such that

$$\text{conc} = f\left(\text{depth}\right),$$

where conc is the atomic concentration. We also note that the information about the atomic concentrations decreases for estimates deeper within the sample. This is reflected by the increasing standard errors for those estimates, as seen in Figure 10.

To find $f\left(\text{depth}\right)$ and to account for the decreasing information as a function of depth, we will use a weighted least squares regression on a high-order polynomial with weights determined by the standard errors of the estimates. Although just a starting point, we fit individual regression functions to each simulated profile that may be sampled from a composition spread alloy film, as seen in Figure 11. Under these simulated conditions, sixth order polynomial regression functions fit the data rather well.

In order to model profiles for different alloy compositions, comp_0, in a CSAF, we add as a predictor a high order polynomial of composition and its interaction with depth to

our model. That is, we will attempt to model:

$$\text{conc} = \alpha + \sum_{i=1}^{m} \beta_{i,0} \text{depth}^i + \sum_{j=1}^{n} \beta_{0,j} \text{comp}_0^j + \sum_{i=1}^{m} \sum_{j=1}^{n} \beta_{i,j} \text{depth}^i \times \text{comp}_0^j. \tag{22}$$

The order of the two polynomials and interactions will be determined with some goodness-of-fit criteria, such as AIC. However, due to the collinearity within the polynomials as well as numerical stability issues in estimating $\beta_{i,j}$, we will transform the predictors into orthonormal polynomials first, using the *stabilized* Gram-Schmidt process.[34] The results from our simulated data can be seen in Figure 12(a) which we found using a 6th order polynomial of depth, a 5th order polynomial of bulk, and all of their interactions. **The important thing to note is that using Eq. (22) above, we can predict the depth profile for any alloy composition.** The fitted simulated alloys as well as predicted alloys can be seen in Figure 12(b). Also, if we find the order of the polynomials it too large, we will investigate the use of a spline basis.

4.2 Measurement-Error Model

The above model in Eqn. (22) assumes the compositions, comp_0 from the composition spread alloy film are known and fixed. As stated earlier, we are wary of this assumption. The process of making the CSAF is new and it is unclear if the properties are understood. Further, with only a few CSAFs made, the differences between "identically" made CSAFs are unknown.

Therefore, we wish to model the concentration function using a "errors-in-variables" or "measurement-error" model. These types of models allow for additional variance in the explanatory variable, in this case, the composition. See Fuller and Gilks et al. for further details.[8][9]

Following Gilks et al., we can redefine our model of the concentration in Eq. (22) in hierarchical terms:

$$\text{conc} = \alpha + \sum_{i=1}^{m} \beta_{i,0} \text{depth}^i + \sum_{j=1}^{n} \beta_{0,j} \mu_{\text{comp}}^j + \sum_{i=1}^{m} \sum_{j=1}^{n} \beta_{i,j} \text{depth}^i \times \mu_{\text{comp}}^j. \tag{23}$$

with the assumed composition imparting information about μ_{comp} with an additional variance term. For now, we will assume the "measured" composition follows a normal distribution given by the mean of the assumed composition, and an unknown variance, that is,

$$\mu_{\text{comp}} \sim N\left(\text{comp}_0, \sigma_{\text{comp}}^2\right) \tag{24}$$

We can model this variance by:

$$\frac{1}{\sigma_{\text{comp}}^2} \sim \text{Gamma}\left(.001, .001\right) \tag{25}$$

where the choice of parameters for the Gamma distribution were selected to be diffuse.

Using Markov Chain Monte Carlo methods on a large number of depths from our concentration functions, we believe we can find reasonable estimates not only for βs, but also σ^2_{comp} which will give us some notion regarding how precise our assumptions about the true composition are.

4.3 Bayesian Analysis

Modeling the concentration depth profile of the CSAF from the Bayesian perspective should be a natural extension to the model laid out in Eqs. (16) - (21) in Section 3.2.2. There, our goal was to find $\mathbb{P}(n_m|I_\ell)$ for an m layer model where we assumed n_m came from a Beta distribution. We extend this by instead modeling the joint distribution $\mathbb{P}(n_{m,\kappa}, \psi|I_\ell)$, where $n_{m,\kappa}$ is the atomic concentration in layer m for the κ composition measure. $n_{m,\kappa}$ is modeled using a logit link with mean given by ψ, a coefficient of the composition. Our model thus becomes:

$$
\begin{aligned}
I|n,\ell,\theta,\kappa &\sim N\left(\mu_{n,\ell,\theta,\kappa}, \sigma_0^2\right) && (26)\\
\frac{1}{\sigma_0^2} &\sim \text{Gamma}\left(\gamma_{0,1}, \gamma_{0,2}\right) && (27)\\
\mu_{n,\ell,\theta,\kappa} &= k_\ell \sum_i^m n_{\ell,m,\kappa} T_\ell(\theta)^{i-1} && (28)\\
\text{logit}\left(n_{\ell,m,\kappa}\right) &= \psi_{\ell,m}\text{comp}_\ell && (29)\\
\text{comp}_\ell &\sim N\left(\text{comp}_0, \sigma^2_{\text{comp}}\right) && (30)\\
\frac{1}{\sigma^2_{\text{comp}}} &\sim \text{Gamma}\left(.001, .001\right) && (31)
\end{aligned}
$$

Allowing for a distribution on comp_ℓ allows us again to estimate the degree of measurement error of the "true" composition of the alloy, σ^2_{comp}. Further, using the joint distribution of the θ_m, we can predict the concentrations for compositions not originally sampled.

5 Summary of Proposed Work

The use of the composition spread copper-palladium alloy films may be invaluable in determining the correct ratio of copper to palladium in order to create a hydrogen purification membrane. We believe that with enough ARXPS data from various locations on the CSAF along with the models laid out in this proposal, we can not only estimate the depth-profile for any alloy composition, but gain some insight into the properties of the CSAF. To this end, we propose to:

1. Determine an appropriate atomic concentration model for a discrete number of layers for a single set of observations from a given alloy (Sections 3.2.1 and 3.2.2).
2. Extend these models to predict continuous depth profiles for any given alloy given the data sampled from a composition spread alloy film (Sections 4.1 and 4.3).
3. Incorporate measurement-error in these models and investigate the variance that may be associated with the "true" compositions (Sections 4.2 and 4.3).

References

[1] Sudhil Adhikari and Sandun Fernando. Hydrogen membrane separation techniques. *Industrial and Engeineering Chemistry Research*, 45:875–881, 2006.

[2] D. Briggs and J.C.Riviere. Spectral interpretation. In D. Briggs and M. P. Seah, editors, *Practical Surface Analysis by Auger and X-ray Photoelectron Spectroscopy*, pages 87–138. John Wiley and Sons, Inc., 1983.

[3] Esteban Broitman. personal communication, 2010. [In person; 04-March-2010].

[4] G. Yu. Cherkashinin. Inverse problem: the concentration depth profile of elements from arxps data. *Journal of Electron Spectroscopy and Related Phenomena*, 74(1):67 – 75, 1995.

[5] P. J. Cumpson. Angle-resolved xps and aes: Depth-resolution limits and a general comparison of properties of depth-profile reconstruction methods. *Journal of Electron Spectroscopy*, 73:25–52, 1995.

[6] Peter A. Dowben and Allen Miller, editors. *Surface Segregation Phenomena*. CRC Press, Inc., 1990.

[7] H. Fischer, R. Svagera, H. Ebel, M. F. Ebel, and B. Schobmann. Depth profiling by arxps in surface analysis. *Fresenius' Journal of Analytical Chemistry*, 353(3-4):473–477, 1995.

[8] Wayne A. Fuller. *Measurement Error Models*. John Wiley and Sons, Ltd., 1987.

[9] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, 1996.

[10] M. G. Grabherr, H. Ebel, M.F. Ebel, R. Svagera, and G. Baron. Investigation of concentration depth profiles by means of angle-resolved xps: A polynomial model. *Journal of Electron Spectroscopy and Related Phenomena*, 63:43–52, 1993.

[11] L. B. Hazell, I. S. Brown, and F. Freisinger. A model for determining the composition of layer structured samples using xps electron take-off angle experiments. *Surface and Interface Analysis*, 8(1):25–31, 1986.

[12] H.W. Kuhn and A.W. Tucker. Nonlinear programming. *Proceedings of 2nd Berkeley Symposium*, pages 481–492, 1951.

[13] A. K. Livesey and G. C. Smith. Maximum entropy: A new approach to non-destructive deconvolution of depth profiles from angle-dependent xps. *Surface and Interface Analysis*, 19(1-12):175–180, 1992.

[14] A. K. Livesey and G. C. Smith. The determination of depth profiles from angle-dependent xps using maximum entropy data analysis. *Journal of Electron Spectroscopy and Related Phenomena*, 67(3):439 – 461, 1994.

[15] James B. Miller, Christopher Matranga, and Andrew J. Gellman. Surface segregation in a polycrystalline Pd70Cu30 alloy hydrogen purification membrane. *Surface Science*, 602:375–382, 2008.

[16] James B. Miller, Bryan D. Morreale, and Andrew J. Gellman. The effect of adsorbed sulfur on surface segregation in a polycrystalline Pd70Cu30 alloy. *Surface Science*, 602:1819–1825, 2008.

[17] M.A. Newton, S.M. Francis, and M. Bowker. Copper-palladium alloy surfaces: Ii. equilibrium surface compositions of dilute Pd/Cu alloys from a simple segregation model. *Surface Science*, 259(1-2):56 – 64, 1991.

[18] Mark A. Newton, Stephen M. Francis, Yongxue Li, Donald Law, and Michael Bowker. Copper-palladium alloy surfaces: I. Cu/Pd[85:15]110, surface structure and reactivity. *Surface Science*, 259(1-2):45 – 55, 1991.

[19] M. Pijolat and G. Hollinger. New depth-profiling method by angular-dependent x-ray photoelectron spectroscopy. *Surface Science*, 105(1):114 – 128, 1981.

[20] Buddy D. Ratner and David G. Castner. Electron spectroscopy for chemical analysis. In John C. Vickerman, editor, *Surface Analysis - The Principal Techniques*, pages 43–93. John Wiley and Sons, Inc., 1997.

[21] Andrea M. Scorciapino, Gabriele Navarra, Bernhard Elsener, and Antonella Rossi. Nondestructive surface depth profiles from angle-resolved x-ray photoelectron spectroscopy data using the maximum entropy method. I. a new protocol. *The Journal of Physical Chemistry C*, 113(51):21328–21337, 2009.

[22] M. P. Seah and W. A. Dench. Quantitative electron spectroscopy of surfaces: A standard data base for electron inelastic mean free paths in solids. *Surface and Interface Analysis*, 1(1):2–11, 1979.

[23] D. A. Shirley. High-resolution x-ray photoemission spectrum of the valence bands of gold. *Phys. Rev. B*, 5(12):4709–4714, Jun 1972.

[24] R. Siuda. A procedure for non-destructive depth profiling of smooth profiles with ARXPS. *Vacuum*, 48(3-4):391 – 394, 1997. Proceedings of the 18th International Seminar on Surface Physics.

[25] J. Skilling and R.K. Bryan. Maximum entropy image reconstruction: General algorithm. *Monthly Notices of the Royal Astronomical Society*, 211(1):111–124, 1984.

[26] ThermoFisher Scientific. *Avantage*, 1999-2009. Online Help Manual.

[27] ThermoFisher Scientific. *Avantage Data Sysetm*, 3rd edition, 02 2007. Document Number: HA600085.

[28] ThermoFisher Scientific. *Sigma Probes: System Overview*, 3rd edition, 07 2007. Document Number: HA600093.

[29] B. J. Tyler, D. G. Castner, and B. D. Ratner. Regularization: A stable and accurate method for generating depth profiles from angle-dependent xps data. *Surface and Interface Analysis*, 14(8):443–450, 1989.

[30] John F. Watts and John Wolstenholme. *An Introduction to Surface Analysis by XPS and AES*. John Wiley and Sons, Ltd., 2003.

[31] WebElements. Webelements: the periodic table on the web — copper orbital properties, 2010. [Online; accessed 11-March-2010].

[32] WebElements. Webelements: the periodic table on the web — palladium orbital properties, 2010. [Online; accessed 11-March-2010].

[33] Wikipedia. Electron configuration — Wikipedia, the free encyclopedia, 2010. [Online; accessed 11-March-2010].

[34] Wikipedia. Gram-schmidt process — Wikipedia, the free encyclopedia, 2010. [Online; accessed 14-March-2010].

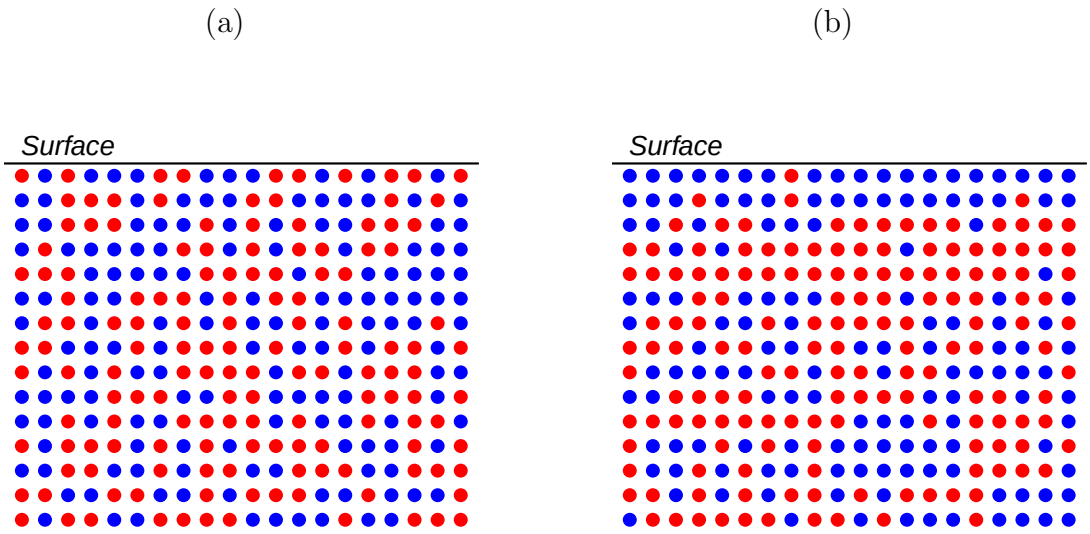


Figure 1: (a) A 50-50 alloy before surface-segregation. The different components shown as red and blue atoms are evenly distributed throughout the material. (b) The same alloy after surface-segregation. The blue component has migrated to the surface while pushing the red component just underneath the surface. The bulk of the alloy, that part of the material that is deep below the surface, remains a 50-50 evenly-distributed alloy.

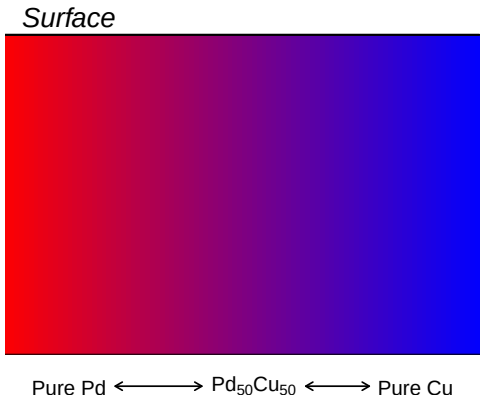
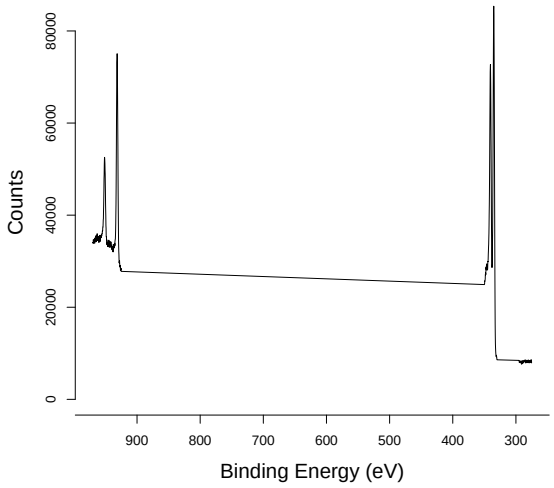


Figure 2: A copper-palladium composition spread alloy film. The CSAF consists of all possible combinations of alloys, with pure palladium on the left with a linear increase in the copper percentage as it moves to the right. The far right side is pure palladium.

(a)



(b) (c)

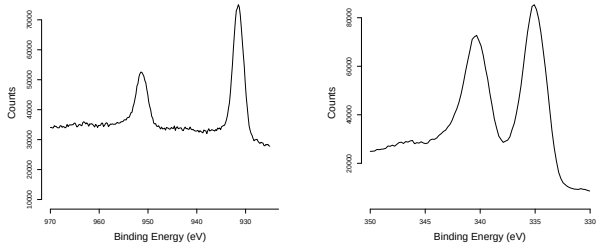


Figure 3: (a) A typical full spectrum resulting from an XPS experiment on a copper palladium alloy. With large counts at known binding energies, the exact elements in a sample can be identified. A close-up of the same spectrum for copper (b) and palladium (c).

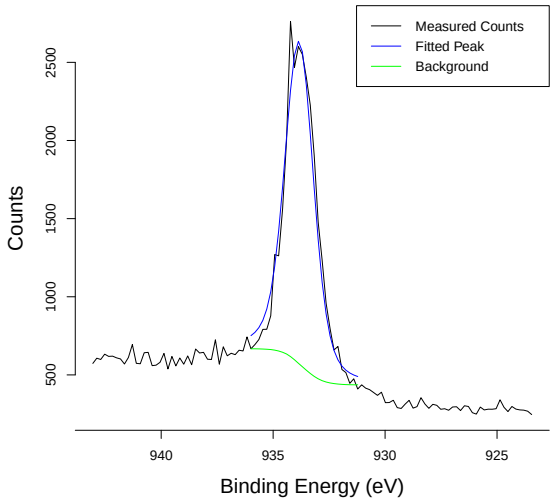


Figure 4: A spectrum resulting from an XPS experiment on a pure copper sample. The black line is the measured counts, the blue line is the peak fit using a Lorenzian-Gaussian convolution, the green line is the background calculating using the "Smart" method. The fiitted peak and background were calculated using the Avantage software.

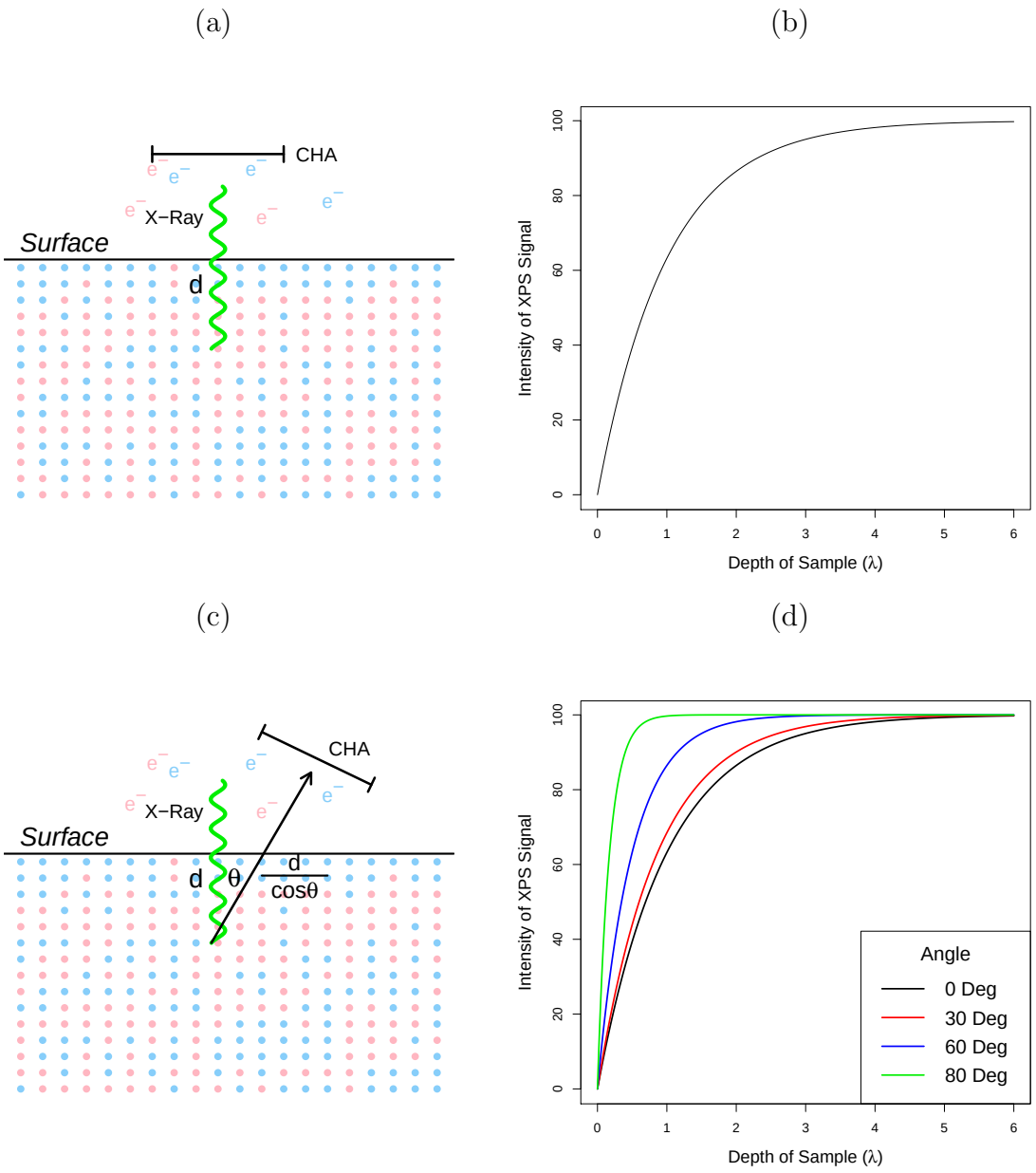


Figure 5: (a) Typical x-ray photospectroscopy where the concentric hemispherical analyzer is normal to the sample surface. (b) The percentage of the intensity of the XPS signal in this setup that comes from electrons that are within a given depth as predicted by the Beer-Lambert law. (c) Heuristically, in angle-resolved x-ray photospectroscopy, the angle of the analyzer with respect to the sample is changed as shown above. In practice, the sample's angle is changed with respect to the analyzer. (d) Steeper angles result is an XPS signal which is more surface sensitive.

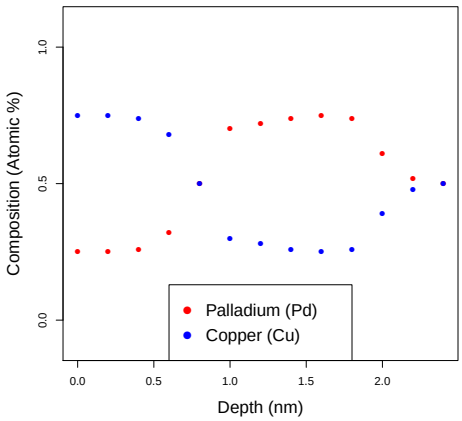


Figure 6: A sample depth profile. The atomic percent of each element in the sample is given at various depths.

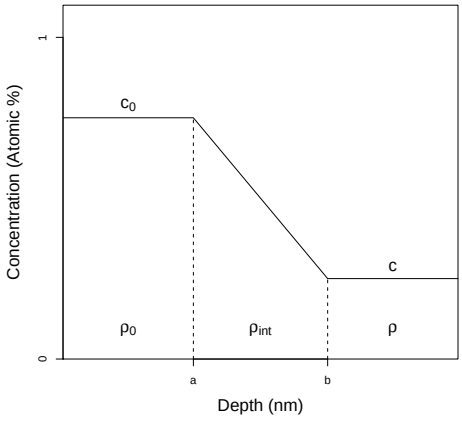


Figure 7: The depth-profile is assumed to be trapezoidal as seen above. A least squares method is used to solve for the free parameters $a, b,$ and c_0 with mass densities $\rho_0, \rho_{\text{int}},$ and ρ_b.

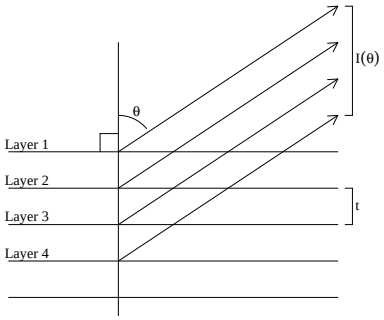


Figure 8: The multilatyer model assumes discrete layers of equal thickness. The XPS signal is given by the weighted sum of the electron energies from each layer, with weights equal to the Beer-Lambert law, the probability of the electron reaching the surface with its original energy.

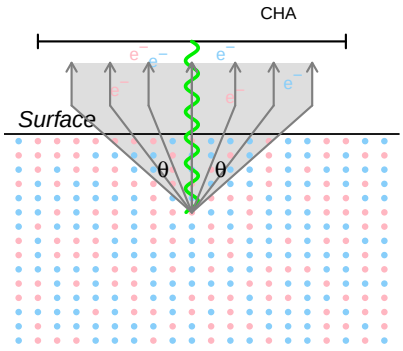


Figure 9: Using a magnetic field to draw the electrons into the analyzer, the Theta-Probe® is capable of collecting angle-resolved data from multiple regions of angles simultaneously.

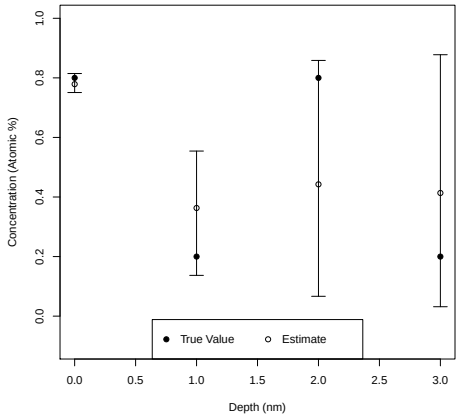


Figure 10: Concentrations estimated from the posterior mean of a Bayesian model. Error bars are given by the 2.5 and 97.5 sample quantiles. The error bars increase with deeper depths reflecting the decreasing information we have within the sample.

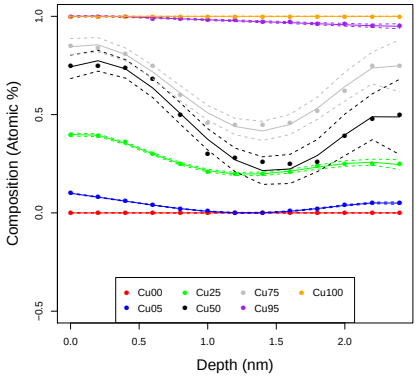


Figure 11: A sample of the possible copper observations from a composition spread alloy film made from , where each line corresponds to a different alloy (e.g., Cu25 is from a $Pd_{75}Cu_{25}$). We can estimate the depth profile using the multilayer model estimates with weighted least squares regression functions of a high order polynomial, in this case, a 6th order polynomial of depth. The larger confidence bands in the deeper depths is a result of the weighted least squares.

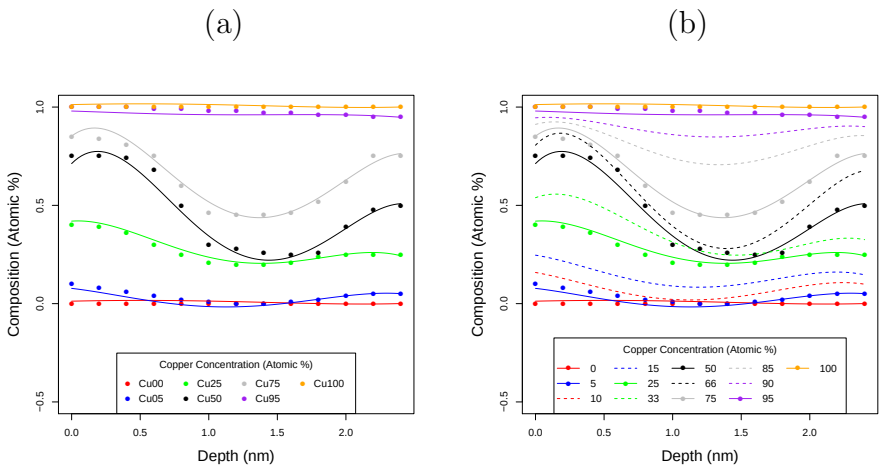


Figure 12: (a) The simulated depth profile estimated using a polynomial regression function with a sixth order polynomial of depth, a 5th order polynomial of original composition and their interactions. (b) The same regression function used to predict new compositions not sampled in the CSAF.