Regularization methods for the extraction of depth profiles from simulated ARXPS data derived from overlayer/substrate models

R.W. Paynter*

INRS Énergie Matériaux Télécommunications, 1650 boul. Lionel-Boulet, Varennes, Québec, Canada

ARTICLE INFO

ABSTRACT

Starting from posited input depth profiles of silicon oxide on silicon, 100 sets of noisy simulated ARXPS data were created for each oxide layer thickness of 3, 6, 9, 12, 15, 18, 21, 24 and 27 Å. Oxygen depth profiles were then recovered from the noisy simulated data using regularized inversion methods, including maximum entropy and Tikhonov regularization. Three regularization parameters were used: one determined by the S-curve method, one determined by the L-curve method and a third corresponding to the closest correspondence between the input and extracted profiles. The various regularization schemes evaluated were ranked with respect to their ability to reproduce the input profile.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

The extraction of depth profiles from ARXPS data, a mathematical manipulation equivalent to an inversion of the Laplace transform, is an "ill-conditioned" problem, as the shape of the profile is highly sensitive to random counting noise in the data [1]. The main problem is "over-fitting", in which unphysical spikes and steps appear in the derived profile in order to fit the noise in the data. A procedure known as "regularization" [2–4] is generally adopted to deal with this issue, and we have investigated this approach in some detail in a previous study [5], by means of synthetic data generated from a posited profile resembling an exponential decay with respect to depth.

In a "regularized" inversion the value of a mathematical expression known as the "joint function" is minimized by optimizing the parameters defining the extracted profile shape. The joint function has the form {residual norm + α solution norm}. The residual norm is a measure of the misfit between the experimental data and the values calculated from a proposed depth profile. A smaller residual norm corresponds to a better fit to the ARXPS data. The solution norm quantifies the complexity of the proposed depth profile, the most widely used example being the profile cross-entropy, used in the so-called maximum entropy (MaxEnt) method [6]. A smaller solution norm generally corresponds to a smoother profile shape. The "regularization parameter" α balances the contribution of each

norm to the joint function, with the aim of achieving the best overall fit to the data with the smoothest overall profile.

In a recent paper [5] we generated synthetic data, with different levels of noise added, from an "exponential decay" input profile shape. We then inverted these noisy data using various mathematical functions for the solution norm (regulators), and compared the extracted profiles with the input profile, as a function of the amount of added noise. Because the result obtained from any one particular set of ARXPS data is really just anecdotal, we ran the calculation 100 times per noise level, and averaged the results. We drew several conclusions from that study:

(1) The regularization schemes contributed to the extraction of profiles that resembled the input profile more closely than those obtained from a calculation that was not regularized.
(2) The regularization schemes stabilized the extracted profile shape – profiles extracted using a regularized calculation were more like one another than profiles extracted by means of a calculation that was not regularized.
(3) The operand in the regulator (concentrations, slopes or curvatures in the profile) was more important to the result than whether or not the regulator calculated a profile entropy.
(4) The extent to which statements (1) and (2) were true depended however on the particular regulator; regulators based on slopes or curvatures in the profile performed better than those based on concentrations.

Because these observations were made for a particular input profile shape, an obvious question arises: do they hold true for other

* Fax: +1 450 929 8102.
  E-mail address: Royston_Paynter@emt.inrs.ca

input profiles as well? In particular, it would be interesting to conduct a similar study using an input profile more closely resembling a classic "overlayer/substrate" structure. We report the results of such a study in this paper.

Using Mathematica software (Wolfram), we synthesize multiple sets of noisy ARXPS data from known input depth profiles, intended, within the limitations imposed by the spacing of profile points in the model, to represent oxide layers of various thicknesses on a substrate. We then invert these data using various regularization schemes in an attempt to recover the original profiles. Profiles composed of discrete layers in the past been a popular test for the inversion of ARXPS data [7–11]. As the input profiles are known, we can rank the recovered profiles, and hence the regularization schemes used to generate them, in a quantitative manner.

In the previous paper we used the "S-curve method" to determine an optimum value for the regularization parameter α from the noisy data themselves, stating that this procedure was intended as a replacement for the well-known but less tractable "L-curve method". In this paper we again employ the S-curve method but also describe and deploy a procedure intended to automate a true L-curve method for the extraction of a value for α from the data. We then demonstrate the equivalence of the S-curve and L-curve methods employed in terms of the results obtained.

Even if the L-curve method is well-known, and the S-curve method equivalent to it, the question remains: is the profile extracted using the value for α determined via the L-curve criterion the best profile that could have been extracted from those data? Could we have extracted a better profile using a different value for α? Because we know the exact parameters of the input profiles used to generate the noisy data, we can address this question in this paper. We will show that for two regulators based on concentrations, including the popular maximum entropy regulator, the optimum profile is extracted about half the time, whereas for a regulator based on slopes, better profiles could have been extracted in 90% of cases.

2. Generation of noisy simulated data

In order to limit the level of abstraction in this study, we will label the two atom types included in the simulations "oxygen" and "silicon", so as to nominally simulate a silicon oxide layer on a silicon substrate, however, we choose to ignore the elastic scattering of photoelectrons [12,13] in our simplified physical model. In the case of a real measurement, elastic scattering cannot usually be neglected at photoemission angles higher than 60–65°. In this work however we are only interested in mathematical aspects of the inversion problem and not in the physics of a real measurement, so the starting point for our calculations is based on the familiar Beer–Lambert equation (i.e. the "straight-line" approximation)

$$I(\theta) = sK(\theta)F\eta \left[1 + \frac{1}{2}\beta \left(\frac{3}{2}\sin^2\gamma - 1\right)\right] \int_0^\infty c(z)\exp\left(\frac{-z}{\lambda\cos\theta}\right)\,dz \tag{1}$$

$I(\theta)$ is the peak intensity at the photoemission angle θ, defined as the angle between the normal to the sample surface and the axis of the photoelectron collection optics. s is a scale factor, analogous to the total counting time in a real measurement. $K(\theta)$ represents a combination of instrumental and geometric factors, such as the X-ray flux and the analysis area; K is taken to be the same for all peaks at a given photoemission angle. F contains the analyzer transmission function, whose value is particular to the photoelectron kinetic energy, and the detector efficiency. The value of η, the

photoionization cross section, is particular to a given atomic orbital. $c(z)$ is the composition depth profile of the element giving rise to the peak; in this work, we will assume a constant total atom density as a function of depth and express the concentration of each constituent element in atomic percent. z is the depth into the sample perpendicular to the surface, and λ is the photoelectron inelastic mean free path, whose value is particular to the photoelectron kinetic energy. The value of λ will also be specific to the sample composition; in this paper, however, we make the simplifying assumption that the value of λ is negligibly sensitive to variations in the atom concentrations. β is the asymmetry parameter of the angular distribution of photoelectrons in the dipole approximation (for free atoms) and γ is the angle between the direction of the exciting photon beam and the direction of the photoelectrons detected [14].

We make several more simplifying assumptions in our simulation of a spectrometer with a "parallel collection" geometry [15]. The solid angle of collection of photoelectrons is taken to be negligibly small, and sixteen θ values evenly distributed between 24.875° and 81.125° were used. The function F is assumed to be flat with respect to energy and equal to unity at all collection angles. The Scofield photoionization cross-sections η were 0.955 for silicon (Si 2s peak) and 2.93 for oxygen (O 1s peak) [16]. The value of β was 2 for both these peaks, and γ was set to $\theta+15°$. The photoelectron inelastic mean free paths were taken to be 23 Å (O 1s) and 32 Å (Si2s) calculated from the NIST IMFP database software [17]. As K will be eliminated by cancellation when we combine the simulated peak intensities to obtain apparent concentrations in at.%, we do not need to consider it further. We employ a ten-point linear segment model for the input depth profile, which is modeled in terms of ten pairs of depth-concentration coordinates (z_i,c_i), with the profile continuing to infinite depth at the concentration value of the tenth profile point beyond the tenth profile point. The total atom density (silicon plus oxygen) is assumed to be constant, so that $c_i(\text{silicon}) = 100\% - c_i(\text{oxygen})$. The z_i were uniformly spaced at 3 Å intervals with $z_1 = 0$ being the sample surface. The synthetic peak intensity is calculated as [18]

$$I(\theta)^{synth} = s\eta\frac{3}{2}\sin^2(\theta+15)\left\{c_1\lambda\cos\theta + \lambda^2\cos^2\theta\sum_{i=1}^{i=9}\left(\frac{c_{i+1}-c_i}{z_{i+1}-z_i}\right)\right.$$
$$\left.\times\left[\exp\left(\frac{-z_i}{\lambda\cos\theta}\right) - \exp\left(\frac{-z_{i+1}}{\lambda\cos\theta}\right)\right]\right\} \tag{2}$$

using the appropriate values for silicon or oxygen for η, λ and the (z_i,c_i). The input oxygen profiles used in this paper, expressed as ten (z_i,c_i) pairs, were flat at a value of 67% to a depth of 3, 6, 9, 12, 15, 18, 21 or 24 Å, dropping to 0% oxygen at the next profile point 3 Å deeper. The "thickness" of this oxide layer is defined at the depth at which the oxygen concentration first equals zero. To the "noiseless" synthetic data, $I(\theta)_{silicon}^{synth}$ and $I(\theta)_{oxygen}^{synth}$, random noise is added, to give 50 or 100 sets of noisy synthetic data (depending upon the investigation, see below). The noise was made to obey Poisson statistics [19] with respect to the value of I. The noisy synthetic data I' is then combined to give a simulated set of ARXPS data in terms of apparent concentration as a function of the photoemission angle, at.%$(\theta)^{synth}$, for example for oxygen

$$\text{at.%}(\theta)_{oxygen}^{synth} = \frac{100 I'(\theta)_{oxygen}^{synth}/\eta_{oxygen}}{I'(\theta)_{oxygen}^{synth}/\eta_{oxygen} + I'(\theta)_{silicon}^{synth}/\eta_{silicon}} \tag{3}$$

The Poisson noise, in terms of the standard deviations on the $I'(\theta)^{synth}$, i.e. $\sqrt{I'(\theta)^{synth}}$, was propagated through this calculation following Harrison and Hazell [19], to give the standard deviation on at.%$(\theta)^{synth}$. Because of the inclusion of the asymmetry