Journal of Electron Spectroscopy and Related Phenomena 178–179 (2010) 433–448

Contents lists available at ScienceDirect

Journal of Electron Spectroscopy and Related Phenomena

journal homepage: www.elsevier.com/locate/elspec

XPS and angle resolved XPS, in the semiconductor industry: Characterization and metrology control of ultra-thin films

C.R. Brundle[a,*], Giuseppina Conti[a,1], Paul Mack[b]

[a] C.R. Brundle and Associates, 4215 Fairway Drive, Soquel, CA 95073, USA
[b] Thermo Fisher Scientific, The Birches Industrial Estate, Imberhorne Lane, East Grinstead, West Sussex, RH19 1UB, United Kingdom

ARTICLE INFO

Article history:
Available online 12 March 2010

Keywords:
XPS
ARXPS
Semiconductor wafer
Metrology
Materials characterization

ABSTRACT

This review discusses the development of X-ray photoelectron spectroscopy, XPS, used as a characterization and metrology method for ultra-thin films in the semiconductor wafer processing industry. After a brief explanation of how the relative roles of XPS and Auger electron spectroscopy, AES, have changed over the last 15 years or so in the semiconductor industry, we go into some detail as to what is implied by metrology, as opposed to characterization, for thin films in the industry, and then describe how XPS, and particularly angle resolved XPS, ARXPS, have been implemented as a metrology "tool" for thickness, chemical composition, and non-destructive depth profiling, of transistor gate oxide material, a key requirement in front-end processing. We take a historical approach, dealing first with the early use for SiO_2 films on Si(1 0 0), then moving to silicon oxynitride, SiO_xN_y in detail, and finally and briefly HfO_2-based material, which is used today in the most advanced devices (32 nm node).

© 2010 Elsevier B.V. All rights reserved.

Contents

1. Introduction 433
2. Characterization and metrology 435
   2.1. Characterization 435
   2.2. Metrology 435
   2.3. Precision and process tolerance 435
3. Development of XPS for characterization and metrology of ultra-thin films 436
   3.1. Background 436
   3.2. SiO_2 gate oxide film thickness as the prototype example 436
   3.3. SiO_xN_y gate oxide films 438
      3.3.1. Background 438
      3.3.2. Initial attempts at thickness and N dose determination for SiO_xN_y 438
4. Depth profiling by XPS in general and the determination of the N profile in SiO_xN_y in particular 440
   4.1. The RDP method 440
   4.2. Modeling the ARXPS response 441
   4.3. Simultaneous thickness, N dose, and N distribution wafer metrology by ARXPS 447
5. Beyond SiO_xN_y 447
6. Conclusions 448
   Acknowledgements 448
   References 448

1. Introduction

This review covers the development of XPS within the semiconductor wafer processing industry for materials characterization and materials processing metrology. Specifically, it concentrates on the use of Angle Resolved XPS, ARXPS, for the characterization and metrology of ultra-thin films. As such, much of what is discussed

* Corresponding author. Tel.: +1 8314791399.
E-mail address: brundle@attglobal.net (C.R. Brundle).
[1] Formerly at Applied Materials, Inc., Santa Clara, CA, USA.

0368-2048/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.elspec.2010.03.008

here could also be applicable to the disk drive data storage industry, where ultra-thin films appeared in manufacturing processes before they did in semiconductor technology, and in any other industry where control of ultra-thin films is needed. What these "high tech" industries have in common then, is that they all involve growth or etching of films, which have to be well-characterized and precisely controlled. The uses of XPS here are in the areas of process development, process control (particularly metrology), quality assurance, and failure analysis.

What this review is not about, but about which it is necessary to say a few words, is the traditional use of XPS as a surface analysis tool in the chemical, chemical engineering, and metallurgical industries. XPS has been used in these industries, such as oil refining, fine and heavy chemicals, and polymers, ever since it was originally established as being highly surface sensitive some 40 years ago [1]. For many years, however, its role was largely confined to research and development areas, where XPS capability might be found as a part of the traditional "materials laboratory". The XPS instrumentation in the materials laboratory at that time did not differ significantly from that in academia, and its effective use required a similar level of specialized expertise. Owing to its extreme surface sensitivity, its chemistry speciation capability, and its relatively easy quantification for surface species (*note*: not necessarily for bulk, or for a non-uniform depth distribution), which were all recognized early on [2], XPS naturally became used where the needs for these attributes was undisputed, such as surface contamination issues (leading to adhesion or corrosion failures, for instance), surface reactions and poisoning in heterogeneous catalysis, and the modification or degradation of polymer surfaces. XPS instrumentation however, had essentially no lateral resolution, and depth profiling was both awkward and slow to perform (alternate cycles of Argon sputtering and spectral acquisition), and often introduced artifacts of change in chemistry [3]. These two factors, no lateral resolution and slow depth profiling capability are part of the reason that XPS did not gain much early traction in the semiconductor device production industry, whereas e-beam Auger electron spectroscopy, AES, did. E-beam AES had sufficient lateral resolution to work at the device level, primarily for failure analysis [4]. This and the seemingly mundane attribute that elemental depth profiling (no chemical state information, and therefore no obvious evidence of the possible sputter induced chemical changes) was done rapidly through micrometer thick films, were very influential in the early adoption of AES in the semiconductor industry materials labs. Also the materials used then in the industry were rather simple and relatively inert (Si, SiO_2, Al, W), meaning that chemical speciation was of lesser importance.

So, from 40 years ago to about 15 years ago, XPS was more widely used for true surface issues in the chemistry-based industries, primarily in the R&D laboratories, whereas AES was not (charging of insulators, no chemical state information as practiced, poor quantification). AES was used more widely in the semiconductor industry and other industries where thin films were deposited as a basis of the technology. AES was not being used primarily for real surface analysis though and certainly not for chemistry. It was primarily a qualitative thin film elemental profiling tool, with excellent lateral resolution, if needed.

Over the last 15 years or so this balance has changed gradually, so that now XPS is much more used in the thin film arena, though, of course, for high spatial resolution, AES must still be used. The main driving force for this change has been the change in the semiconductor industry itself. There has been a steady decrease in the thickness of many layers termed "thin films" down to, in some cases, sub-nanometer, which has made XPS a highly suitable technique for materials characterization of the whole film, because of the match in thickness to XPS probing depth. In addition the surface of a layer, or interface between layers, on the nanometer scale was previously often technologically irrelevant when film thickness approached micrometer dimensions. A few Angstrom of reaction was of no concern for the bulk properties of the film (though it could be a practical showstopper because of delamination for instance). Today, with many films of only nanometer's thickness, the surface/interface reaction may consume a considerable fraction of the film and strongly affect the overall film properties. This is exacerbated by the recent trend to more exotic and reactive materials. The instability and reactivity of HfO_2 at a Si interface, where processing conditions can result in interdiffusion and a complex phase mixture, is a perfect example – yet HfO_2-based sub-20 Å films are in current production as the gate dielectric material in the most advanced transistors, and therefore must be very precisely controlled. Another example is the switch from Al to Cu for the interconnect metallurgy. Cu diffusion into the active device region is catastrophic to device performance, which leads to the need to introduce an additional effective barrier layer, such as Co. This, however, can only be a few nanometers thick, or the conductivity benefit of moving from Al to Cu for reduced dimension structures is compromised (the barrier layer takes up thickness that should be available to the Cu interconnect wire). Likewise, there are issues of controlling a few Angstrom of oxidation on the Cu seed layer which is deposited prior to electroplating the interconnect lines. The dielectric insulation between the interconnect levels has also changed drastically, and there may be up to nine interconnect levels in advanced product. Traditionally the insulation was SiO_2. Now exotic organic/inorganic material mixtures are also used, which can have homogeneity and stability issues (towards humidity, for instance).

Summarizing, the industry has evolved to the point where:

(a) Layers can be so thin that there is little distinction between a surface or interface region and the bulk of the film,
(b) XPS is an appropriate approach to analysis for the whole film thickness, and
(c) The wide range of materials now used requires much more chemistry understanding and control.

Hence the growth of XPS as an analytical tool, not (just) for surface issues, but for the whole film.

While the above changes were occurring in the semiconductor industry, the XPS instrumentation was also evolving. Lateral resolution has gradually improved for lab-based equipment to the point where it is quite feasible to work on the scribe lines and test areas between dies on patterned wafers, thus allowing measurement at any stage during the wafer processing steps. This requires current lateral resolution of about 10 μm, with enough signal strength to allow measurement in minutes rather than hours. Of course current lab XPS equipment does not (and probably never will) have the capability to work at the device structure level, where 10s nm resolution is needed.

The instrumentation intended for use in the industry also started seriously diverging in design from that intended for general purpose usage about 15 years ago. Instruments capable of taking large samples, and eventually full 200 mm and 300 mm wafers became available. They also became more automated, initially in terms of loading wafers and acquiring data in different locations (mapping), and later in automating the data reduction process to produce what might be termed the XPS analyst's "standard table"; that is a table of percentage elemental compositions, based on Scofield's atomic photoionization cross-sections [5], or empirically determined cross-sections for the particular instrument design, plus some information on chemical speciation of the elements present from resolved chemical shifts. Thus, the need for expert involvement to get to this stage – the standard table – was greatly reduced and therefore access to the technique greatly widened.

Of course the expert does not necessarily consider this development a good thing! First there is the concern that the expertise is no longer needed. In our experience this is never the case, partly because of the second concern, which is that a "standard table" of percentage compositions will only approximate the truth if the material composition is (a) constant with depth over the depth range probed and (b) if there are no significant differences in depth probed for each element (which, in turn, depends on which XPS lines are being used for the analysis). If there is any communication at all between the expert and the customer receiving such automated analysis (not always the case!), the customer quickly realizes the expert is needed to explain and mitigate the caveats of the "standard table".

However, it is often the case that, in comparison of sample A to sample B, use of the "standard table" is straightforward, because one is often looking for change, not correctness of absolute values. The expert's role here is to establish when this approach is appropriate and when it is not.

A specialized development of XPS instrumentation was to produce a system which could automatically provide a metrology for thin films. We will explain what is implied by "metrology" for the semiconductor industry in the next section. Here we just want to make clear what it means for the nature of the instrumentation. It means either having an instrument designed and dedicated entirely to the particular measurement in question, e.g. thickness, usually with hidden layers of proprietary software, or having a somewhat more flexible instrument which can also be switched between a dedicated mode for metrology (operator use) and a more general analytical mode (engineer, or expert use). Both types of XPS now exist in the semiconductor industry. The former [6] is found in chip-making fabrication facilities, FABS, and in companies developing processing tools for chip-making FABS, and the latter [7] is found both in FABS and in material labs supporting such operations.

2. Characterization and metrology

2.1. Characterization

Materials and process characterization should take place during the development stage of a product. In the case of Si wafer processing this implies characterizing everything about a deposited layer one needs to know in order to reliably produce the end product. This could mean, and often did mean, in the early days of semiconductor devices, as little as measuring film electrical resistance with a four point probe [8], or it could be as extensive as determining physical and electrical thickness, density, porosity, composition, interface roughness, stability towards ambient conditions (e.g. humidity), stability towards subsequent process conditions (e.g. elevated temperature), and diffusion and reaction properties towards the substrate or a subsequent layer. The idea is to perform enough characterization so that when a viable product emerges from development, it is known what parameters must be controlled, and to what degree of precision.

2.2. Metrology

In principle, then, once "enough" has been established by characterization during the development stage, tolerance limits can be set for the critical parameters, and measurement procedures (the metrology) put into place as an integral part of the manufacturing process in order to stay within the limits (statistical process control, SPC).

A dictionary definition of metrology is simply "the science of measurement". The science community understanding is "measurement to a known precision". The industrial/technological understanding is "measurement to a known precision, capable of meeting process tolerance requirements (upper and lower control limits)". The Si wafer processing industry expectation is that a viable metrology must also be rapid (seconds or minutes per measurements), robust (e.g. 95% metrology tool uptime) and FAB-compatible (no expert involvement on a day to day basis).

Since production wafers are currently 200 mm and 300 mm, metrology tools must handle these, and be capable of mapping the wafer (anywhere from 5 points to 100% mapping may be required, depending on what is being measured). If the metrology tool is in-line it must also be fitted to accept automated wafer cassette and wafer transfer operation, a very expensive proposition. If it is off-line, or in the support materials laboratory, automated single wafer handling may be acceptable.

The process engineer's view of metrology is likely to be quite different from that of the typical physical scientist, who will usually be concerned with understanding exactly what is being measured, and whether it is accurate. The process engineer simply wants a correlation between the parameter value being produced by the metrology measurement, at that particular point in the wafer processing sequence, and the final electrical performance of an individual device after completion of fabrication. As such, it is usually precision he/she is interested in. Taking gate oxide thickness as an example, it serves no purpose to determine "thickness" with insufficient precision to detect changes which affect final electrical performance. On the other hand it would be equally useless to determine "thickness" to a high precision, if change in thickness were not a controlling factor of electrical performance. In this example of gate oxide, the process engineer does not particularly care whether the absolute value of the gate oxide thickness is determined correctly, or whether it is even a physically meaningful concept, as long as changes in the value determined correlate with changes in device performance. This is the reason that the absolute accuracy of a metrology measurement is likely to be much less important than the precision of the measurement, i.e. the ability to detect change.

2.3. Precision and process tolerance

To have confidence to some defined level (e.g. 1σ, 3σ or 6σ) that a parameter is within upper and lower process limits, the metrology tool must have considerably greater measurement precision capability than the process tolerance sought. The "Watchmakers Golden Rule" [9] is that the metrology tool should have 10 times better measurement precision capability than the tolerance sought for the process under test. Thus:

$$\frac{P^{6\sigma}}{t} = 0.1 \quad (1)$$

In practice, in the wafer industry this arbitrary standard has been relaxed to:

$$\frac{P^{6\sigma}}{t} \leq 0.3 \quad (2)$$

So, again taking film thickness as the example, where the upper and lower control limits (at 6σ) are often set at $\pm 4\%$, i.e. a tolerance, T, of 8%, for a $P(6\sigma)/T$ of 0.3 on a 10 Å film:

$$P(6\sigma) = 0.08 \times 0.3 \times 10\,\text{Å} = 0.24\,\text{Å}, \text{ or } 2.4\%$$

$$\text{or } P(1\sigma) = 0.24\,\text{Å}/6 = 0.04\,\text{Å} \text{ or } 0.4\%$$

Instrument vendors usually quote the precision of their instruments to 1σ or 3σ.

Table 1
Some ultra-thin films used in the wafer processing industry.

Material/technology	Mature production	Advanced production
SiO_2 in gate (Å)	20–40	–
Si/O/N gate (Å)	15–25	6–10
Metal gate (Å)	–	Gate + 20–30
High *k* gate (Å)	–	20–50
Shallow implant (Å)	∼100	20–30
Liner/barrier (Ta/TaN) (Å)	100s	10–30
Cu seed layer (Å)	1500	100
Oxide on seed layer (Å)		≤5
Co passivation (Å)		100s

Table 2
Examples of precision capabilities requested from metrology tool vendors.

	Minimum film thickness measurable	1σ precision
Gate (Si/O/N)	10 Å	0.5%
High *k* (HfO_2)	10 Å	0.3%
Barriers (TaN)	10 Å	0.3%
Gate N dose (for Si/O/N)	10 Å	1% at 1E15 atoms/cm^3
Implant dose (B, F, As, P)	50 Å	1% at 1E15 atoms/cm^3 5% at 1E14 atoms/cm^3

3. Development of XPS for characterization and metrology of ultra-thin films

3.1. Background

Table 1 lists the thickness for some of the key films involved in both mature and advanced semiconductor FABS. We give below a brief explanation of their roles for those not familiar with this industry.

SiO_2 was the industry standard for the transistor gate oxide until a few years ago. It cannot be used for 90 nm node technology and beyond because, with the device scaling (shrinkage) required at this node, it would have to be <10 Å thick for the required capacitance, and it is not possible to engineer reliably such a "bulk" SiO_2 film. This was demonstrated by Muller et al. [10] using transmission electron microscopy/electron energy loss spectroscopy, TEM/EELS, who showed that, in a processed gate stack, a 20 Å "SiO_2" layer actually consisted of about 50% interface region (top and bottom) where the material did not have the required bulk SiO_2 properties. At 12 Å nominal thickness no bulk SiO_2 remained. Such a film would be very electrically leaky and would have a dielectric constant, *k*, lower than bulk SiO_2.

Silicon oxynitride films, SiO_xN_y, were introduced into production originally as a "temporary" fix for this problem, after a lot of development work concerning the processing method; the total nitrogen content (known as the "dose") and the way the nitrogen should be distributed through the film. The usual manufacturing procedure is to start with a thermally grown SiO_2 layer and thermally treat it (>900 °C) in NO or NO_2, or implant nitrogen by a plasma process [11]. The idea is to create a leakage free film with a modest increase in *k* (∼25%) compared to bulk SiO_2, allowing a modest increase (∼25%) in thickness for a given capacitance. This material is in current production for 90 nm node technology.

Metal gate is a follow-on technology where a metal layer is placed between the oxynitride gate oxide layer and the subsequent α-Si [12].

High *k* gate [13] uses a dielectric layer with a considerably higher dielectric constant than that of SiO_2, allowing considerably greater physical thickness for an equivalent SiO_2 electrical thickness (known as the equivalent oxide thickness, EOT). HfO_2-based material has become the choice here, but an acceptable structure has been hard to achieve because of the reactivity of HfO_2 towards Si under processing conditions. This issue led to a major development program within the industry, and much academia study also. This program had a large need for characterization of the H_fO_2/Si interface and XPS has played a significant role in this [14]. Metal gate is now also incorporated into the HfO_2 gate oxide for the 32 nm node [13].

Shallow implant [15] is a plasma based process which is replacing traditional high energy (many keV) ion implantation for source and drain dopants such as As and B, so that the dopant is confined in a much thinner surface layer. The much thinner layer also means the dopant concentration must be much higher to produce the same electrical performance. Both factors – thinner layers and higher dopant concentration make the use of XPS characterization possible [16].

The remaining layers listed in Table 1 are deposited at the "back end" of the production line and are related to the interconnect wiring, which, in the most advanced product, is nine layers deep and Cu based.

All of the films listed in Table 1 have been the subject of intense characterization studies during development, with XPS playing a significant role. However, owing to the lack of sufficient lateral resolution, the XPS work is confined to films deposited on flat wafers ("monitor wafers") and test and scribe regions on patterned product wafers.

Table 2 lists the metrology tool precisions actually being specified by the industry. For thickness determination they are compatible with a 6σ tolerance of ±4%. At first sight it may seem absurd to ask for a precision in measurement capability of 0.5%, or 0.05 Å, for a 10 Å film, since this is a small fraction of an atomic size. XPS, however, determines the average thickness over the area sampled, using the ratio of overlayer to substrate XPS signal intensities [17], assuming a density and a mean free path length for inelastic scattering. A typical area analyzed is 10^4 μm^2. A 10 Å film thus corresponds to several 10^{11} atoms, and it is actually not difficult to achieve the required precision for a raw XPS peak intensity in a short time. The determined average thickness is then typically calibrated against some "direct" method, such as high resolution TEM cross-sectional sampling (caution is also needed here, though, as several things have become apparent through round robins: (a) for very high resolution cross-sections, only a very small length of the gate thickness is sampled in any given image, which may, or may not be typical of the average value needed; (b) the few Angstrom interface region in a TEM image of a given cross-section of a few 10s Å thick film can look quite different under different conditions of focus and contrast, leading to different operator placement for "the interface"; and (c) even for a given TEM image of a given cross-section there can be operator disagreement on exact placement of the interface). Of course, for XPS, if there were sufficient variations in surface roughness, or inhomogenity in density from measured site to site on the wafer, this would affect the ratio of XPS intensities and result in apparent variations in thickness. If such effects correlate with final device electrical performance changes in the same manner as a "real" thickness change, no one would know the difference (or care). If they caused different electrical performance changes, this could reduce the usefulness of XPS as a thickness metrology tool.

3.2. SiO_2 gate oxide film thickness as the prototype example

The standard, simplified, 2 layer XPS model [17], used for determining layer thickness for an SiO_2 film on Si substrate, results in the equation:

$$\ln\left[1+\frac{R}{R^{\infty}}\right]=\frac{d}{\lambda\cos\theta} \quad (3)$$

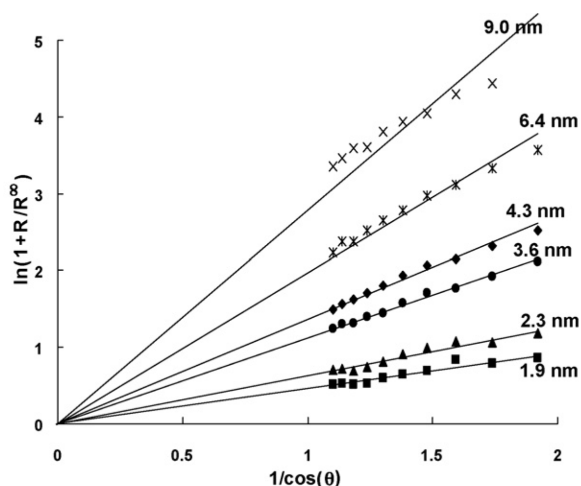


Fig. 1. Thickness measurement: testing model validity for ARXPS for SiO_2/Si, plot of $\ln[1+R/R^{\infty}]$ versus $1/\cos(\theta)$; gradient = d/λ, nominal thickness from ellipsometry.

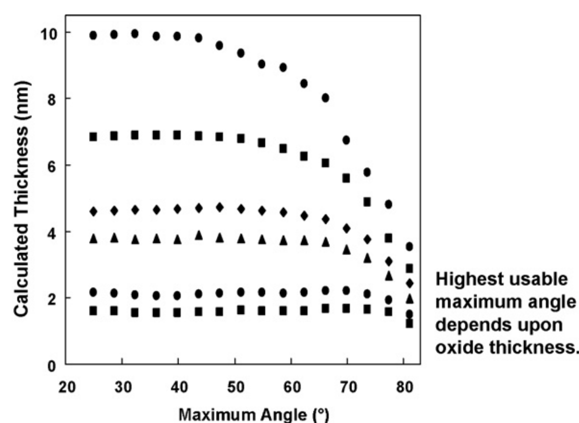
where R is the measured ratio of the Si^{4+}2p SiO_2 XPS intensity to the $Si^0$2p Si substrate intensity; R^{∞} is the same ratio, but for clean bulk SiO_2 and clean bulk Si; λ is the inelastic mean free length for Si2p electrons through SiO_2; and θ is the analyzer acceptance angle w.r.t. the surface normal. It is a simplified model because there are at least 2 assumptions:

(1) The λ value for Si^{2+}2p electrons is the same as for $Si^0$2p substrate electrons. Since their KE's are practically the same, this is not a problem. For determination of thickness on a different substrate, using a substrate XPS signal of significantly different KE from the overlayer XPS signal, using the simplified equation would introduce an error.
(2) It ignores the effect of elastic scattering (see later).

In addition, the value for R^{∞} may be obtained experimentally, or may be a theoretical value. Variation in the value used could be considered a calibration error (for SiO_2/Si the theoretical value differs significantly from the experimentally determined value, possibly because there is significant intensity in the plasmon loss structure, which varies from the Si^O to the Si^{4+} situation, and which is not included fully and correctly in the experimentally determined value).

A plot of $\ln[1+R/R^{\infty}]$ against $1/\cos\theta$ should be linear, passing through the origin, with a slope of d/λ. Fig. 1 is such a plot [18] for a range of SiO_2 thickness determined by ellipsometry (long a standard metrology for SiO_2 thickness [9], which works provided the SiO_2 layer possesses the bulk optical properties that go into modeling the ellipsometry data. This begins to break down when the interface width becomes significant compared to the full film thickness). The linearity holds well up to about 6 nm thickness, and then starts to break down. We know why. The elastic scattering effects, not included in Eq. (3), are accentuated at high θ and large layer thickness [17,19].

From a practical point of view, for SiO_2, not including elastic scattering in Eq. (3) simply results in a small absolute error in determined thickness provided data above about 60° is not included (for thickness up to about 6 nm). This is demonstrated in Fig. 2, where calculated thicknesses, using Eq. (3) and a straight line fit as in Fig. 1, are shown as a function of the maximum angle for which data is included [20]. The determined thickness remains sensibly constant until data above a critical angle is included, at which point it starts to decrease. It is above this angle where the importance of a rapidly changing contribution of elastic scattering starts to increase. Consequently we know not to trust the



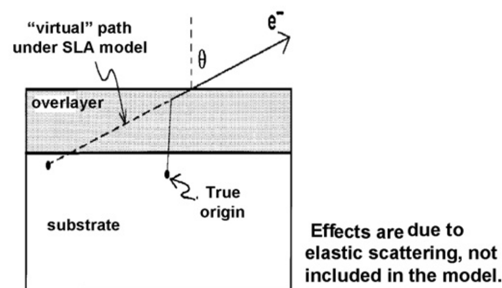


Fig. 2. (Top) ARXPS measurements of SiO_2 thickness: effect of angular range on thickness calculation determined using Eq. (3). (Bottom) Schematic of the effect of elastic scattering on apparent electron trajectory.

model when data above about 60° is included, for thicknesses above about 6 nm.

These are empirical observations, and the effect, on the absolute thickness determined, of leaving out the elastic scattering term for data below a 60° angle, can be corrected for by a TEM calibration point to yield an effective λ, if absolute accuracy is a concern.

The same general behavior of the effect of elastic scattering appears to hold for other material, such as HfO_2, but the elastic scattering effects are stronger for the higher Z material and consequently become significant for thickness determination at smaller angles and lower thickness.

Fig. 3 plots the extracted values of d, assuming an effective λ, against measured ellipsometric values [21]. There is an excellent linear correlation, but there is an offset of about 1 nm, which is due to the fact that there was significant hydrocarbon contamination on these films (they had been kept in ambient for a considerable time). The ellipsometry optical modeling tries to fit this as an additional SiO_2 component, but it has no effect on the XPS results because the presence of a hydrocarbon layer does not affect the ratio, R. As a function of time, and therefore increasing hydrocarbon presence, the ellipsometric values will therefore increase, but the XPS values will not be affected.

Based on the above approach, Fig. 4 is an XPS determined plot of the oxide thickness, as a function of radius for a 150 mm Si wafer, which has gone through a standard thermal oxidation process to produce a nominal 40 Å thick layer. The equivalent standard ellipsometry measurements are also shown. This wafer was provided by the present authors to VG Scientific (now Thermo, Inc.), where the data was acquired using a prototype Thetaprobe (see later), and analyzed by Seah et al. [22]. Without getting into the details of the statistics it is clear that ARXPS can do a creditable precision job at this thickness. This example was the first open literature analysis of this type on a real production wafer, though major semiconductor equipment and semiconductor chip-making companies had been pursuing such demonstrations with at least two XPS vendors for

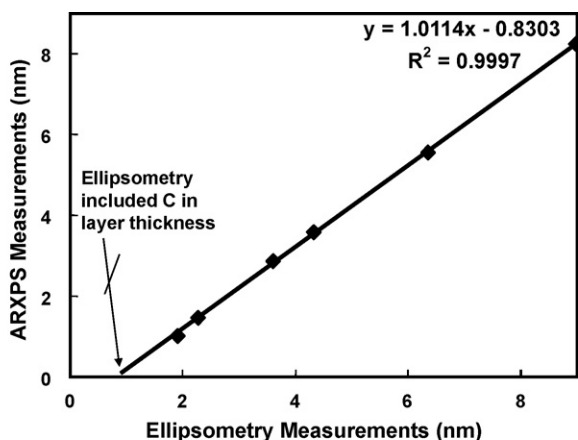
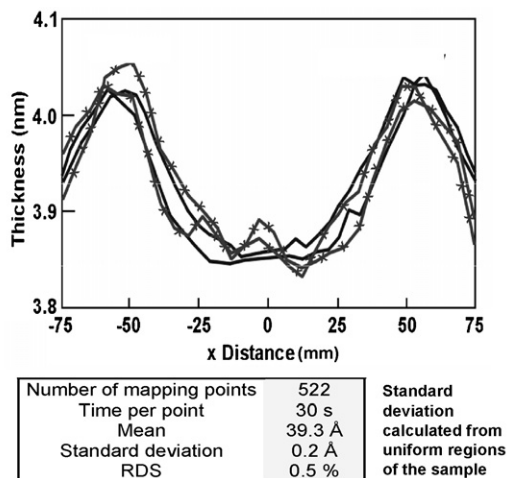


Fig. 3. Comparison of XPS SiO_2 thickness values from ARXPS (Fig. 1) to ellipsometry values.

some time before this. The real interest, of course, was not to pursue XPS as a potential metrology approach for thermal SiO_2 on Si, since ellipsometry remained a fast and adequate approach down to the minimum industry thickness of such layers (Table 1), but to investigate its potential for thinner and more complex material where ellipsometry modeling might not work.

Of course it is not strictly necessary to perform XPS as a function of θ to extract a value for thickness, d. Eq. (3) can be used with a single angle. This is the approach taken by ReVera in their



Number of mapping points	522
Time per point	30 s
Mean	39.3 Å
Standard deviation	0.2 Å
RDS	0.5 %

Standard deviation calculated from uniform regions of the sample

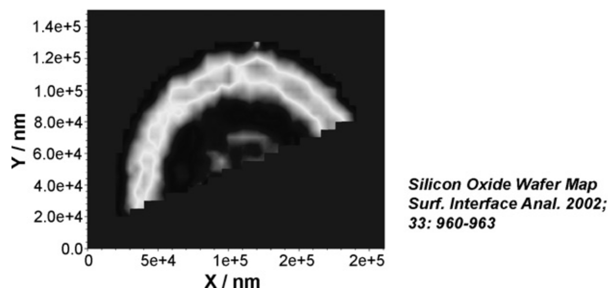


Fig. 4. Thickness wafer mapping by ARXPS: ~40 Å SiO_2 on a 150 mm Si(1 0 0) wafer, and comparison to SiO_2 (Ref. [22]). (x) ARXPS; (() ellipsometry.

FAB-compatible RVX1000 instrument [6], which is designed purely for fully automated metrology of SiO_xN_y and follow-on products. However, without checking that there is at least a linear fit to Eq. (3) (necessary but not sufficient), one cannot have confidence that the model used is appropriate, and if it is not, then a different value of d would result from fitting at each different angle.

3.3. SiO_xN_y gate oxide films

3.3.1. Background

The introduction of nitrogen content into a SiO_2 film may appear simple, but it greatly complicates the characterization and metrology requirements. With almost half a century of experience, industry was pretty comfortable it had adequate control over the SiO_2 gate oxide, at least down to 30–40 Å thickness, so a simple thickness metrology was all that was needed [9]. At smaller thickness the properties started to get non-bulk like, as demonstrated by Muller et al.'s work [10], resulting in the adoption of SiO_xN_y material as an alternative.

Below the 90 nm node, where SiO_xN_y thickness down to below 20 Å is being used, one now needs metrologies which will determine variation of thickness, and the total N content ("dose") (as well as dielectric constant at such a thickness). Also, since statements such as "... *with the result of creating an N-rich layer near the Si/SiO_2 interface*" are made by chip-making companies [23], it is very clearly implied that the N distribution, not just the total dose, through the SiO_xN_y layer is important – i.e. it affects electrical performance.

20 Å SiO_xN_y films can have thickness variations successfully monitored by ellipsometry, by fitting parameters to an optical equation model, provided the N dose, and its depth distribution, do not change, but it is harder to go thinner, and impossible to determine both thickness and N dose reliably this way to the required precision. XPS can, in principle, do this and, in addition it holds at least the promise of providing some N depth distribution information. It can also give information about any chemistry difference in the interface regions, compared to the "bulk" of the film. Owing to this promise a large effort has gone into evaluating and establishing the use of XPS for both characterization, and metrology (thickness, dose, and maybe N distribution) of gate SiO_xN_y films (and, after that, HfO_2-based films).

Much of the published work took place in academia and government laboratories, and often parallel work in industry was kept confidential. A problem in bringing all this information together, and evaluating it, has been in establishing whether differing results/conclusions in different laboratories represent differences in the films used, or limitations in the analytical approaches used. Here we simply review our own attempts, over a several year period, using real industrial wafer product, and give our opinion of what the current status is for this type of metrology using XPS.

3.3.2. Initial attempts at thickness and N dose determination for SiO_xN_y

3.3.2.1. Thickness. To a first approximation SiO_xN_y thickness can be determined using the same approach as for SiO_2, namely simple measurement of the Si^{4+} (overlayer) to Si^0 (substrate) ratio, and use of Eq. (3) with an SiO_2 λ value. This will introduce an obvious error: the λ appropriate for SiO_xN_y is not the same as for SiO_2, and will depend on x and y. Ignoring this for the moment, we show, in Fig. 5a the correlation between two sets of thickness measurements for a set of 60 SiO_xN_y samples. One set was derived using a single angle XPS metrology tool [24], the other using a Thermo Theta 300 ARXPS instrument [25]. The objective was to see if the instruments could be "tool matched", meaning could they be cross-calibrated to give the same answers. The Thermo 300 ARXP data, though taken in angular resolved mode, was collapsed to a single average angle to

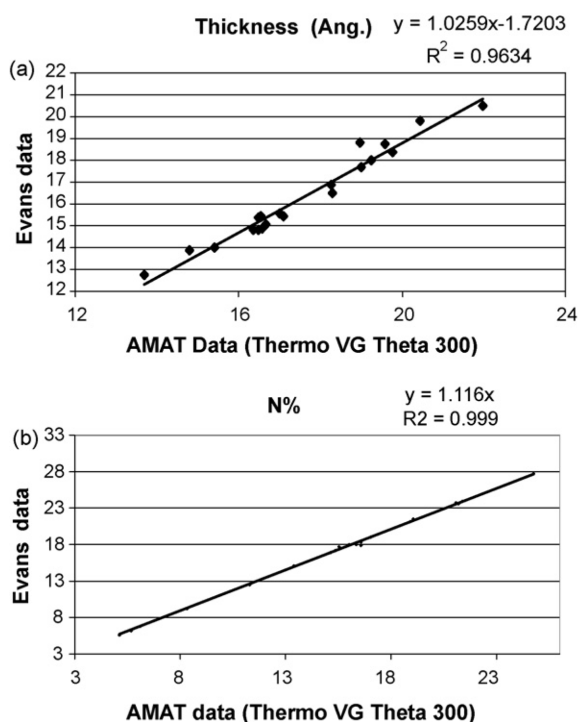


Fig. 5. Tool matching: single angle instrument results (Evans data) versus ARXPS instrument results (AMAT data) for a 60 sample set of SiO_xN_y wafers. (a) Thickness and (b) N atomic %.

be compared to the single angle instrument data. The correlation in Fig. 5a is linear. This indicated that, in principle, either instrument could be used for thickness determination, with Fig. 5a providing the necessary tool matching cross-calibration. The gradient is close to unity also (not actually important for tool matching), but there is an offset of 1.7 Å in the determined value. This is not important, since a calibration by TEM cross-section would be performed if a true physical thickness value was wanted. There are a couple of outliers in the plot. We do not know the origin of these.

3.3.2.2. N dose. The term "N dose", used in the industry, simply means the number of N atoms/cm^2 introduced into the SiO_2 film by the thermal or plasma treatments. What is actually measured in XPS is the intensity of the N(ls) signal. This is converted to the "standard table" N% value using an appropriate cross-section value [5], and, of course assuming complete homogeneity of concentration through the film. Using the assumption that the film can be represented, for stoichiometry consideration, as an $(SiO_2)_a/(Si_3N_4)_b$ mixture, no actual O(ls) measurement is needed and the N% can be determined using only the N(ls), and Si^{4+}(2p) signals. For the 60 samples, the N% was determined with the above assumption for the Theta 300 instrument data. Since no O(1s) data was used in the ReVera approach, we believe, but do not know for certain, that the same assumption was also made in their N% determination. The correlation is shown in Fig. 5b. It is linear, passes through the origin, and has tight precision. The slope deviates significantly from unity, however. We do not know why, and for tool matching purposes it is irrelevant. The important point is that, without knowing anything further an engineer can convert an answer obtained by one instrument to that which would have been obtained by the other. Which value is actually "correct" is also irrelevant, as long as everyone is on the same page.

If the composition is uniform throughout the depth of the film, then the N% can be converted to the required N dose by:

$$\text{N dose} = \text{N\%} \times d \times C \quad (4)$$

Table 3
Thicknesses and doses for a set of 6 Si/O/N wafers. The larger (wrong) N dose values were derived from a single angle measurement assuming uniform depth distribution. The smaller values were derived after determining the true N depth distribution by ARXPS.

Sample	Total thickness (A)	Apparent N dose (1e15/cm^2)	Corrected N dose (1e15/cm^2)
1	74.8	11.0	4.8
2	75.3	11.7	5.2
3	78.2	11.5	5.3
4	75.4	11.6	5.1
5	72.6	12.1	5.7
6	76.4	12.0	5.6

where *C* is a constant containing density and other terms. So, at this point we can see that, from a metrology point of view, either the single angle instrument or the Theta 300 instrument could be used for thickness and N dose metrology, with Fig. 5a and b being used for tool matching, despite the fact that we do not actually know how the single angle instrument data is processed.

If, however, the N concentration is not uniform with depth, but is actually decreasing with depth, then, since XPS probes with an exponentially decreasing sensitivity with depth, an assumption of uniform depth distribution will lead to an "apparent N dose", where:

$$\text{N apparent dose} > \text{N true dose}$$

If the N concentration is actually increasing with depth, then:

$$\text{N apparent dose} < \text{N true dose}$$

The extent of the discrepancy between apparent and true doses obviously depends on the magnitude of the concentration variation with depth over the thickness of the whole film. An extreme example is shown in Table 3, for a set of wafers which turned out to consist of roughly 50 Å of SiO_2 on Si(1 0 0), plus roughly 25 Å SiO_xN_y on top of the SiO_2, as in the schematic of Fig. 6. The outer few Angstrom of the stack had also re-oxidized in ambient atmosphere towards SiO_2. A "blind" metrology measurement, *assuming uniform N distribution* through the whole 75 Å, and calculating N dose from N% × *d*, gives values more than twice the values obtained using the correct physical structure (established using ARXPS, Table 3 – see Section 4). Though this may seem an unrealistically extreme example, it is in fact a real example of a gross overestimate of dose made by both using an inappropriate application (software assuming uniform N distribution), and applying it to a film that is much too thick (>2λ).

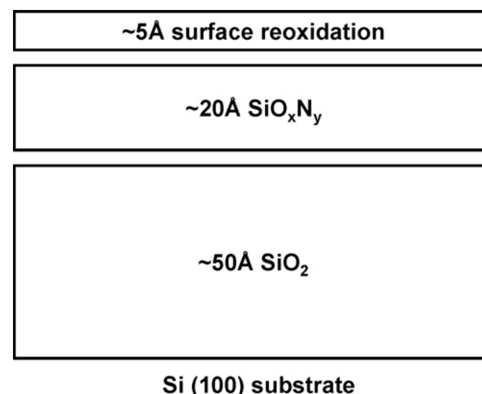


Fig. 6. Schematic of the actual layer structure for the nominal "75 Å layer" of SiO_xN_y on Si(1 0 0) of the samples in Table 3.

4. Depth profiling by XPS in general and the determination of the N profile in SiO_xN_y in particular

The above discussion makes it clear that even if the N depth distribution was not considered important to device function, an incorrect N dose would result from using an incorrect N distribution. The values of a and b in the assumed stoichiometry $(SiO_2)_a(Si_3N_4)_b$ are also directly related to N dose, so an incorrect dose value will also affect a λ value based on an interpolation from SiO_2 and Si_3N_4 λ values, which in turn affects the calculated thickness. This would, however, be a much smaller error than the equivalent dose error.

In fact the industry knows, by extensive electrical testing, and characterization of the N profile by destructive methods (TEM/EELS, time-of-flight secondary ion mass spectroscopy, TOF-SIMS), that the N profile does affect device performance – at least gross effects do. So it is important to understand the N depth profiling capabilities of XPS, including the limitations.

There have been three general approaches used in attempting to obtain non-destructive depth profiling information by XPS. One is to make use of angle resolved data and the changing depth probed. The second is to compare intensities of the XPS peak to the background on the low KE side, at a fixed angle. The increase in background is caused by inelastically scattered electrons which would otherwise have been in the XPS peak. The further below the surface an element is located, the smaller will be the unscattered XPS peak and the larger the inelastically scattered background following that peak. Modeling this situation is known as the Tougaard background approach, fully reviewed in the article in this volume by Tougaard [26], so we will not discuss it further. We should point out though, that this approach seems to be the basis for proprietary software that ReVera sells to attempt to correct their single fixed angle metrology measurement for any non-uniform depth distribution.

The third, more ambitious approach to the issue of depth information from XPS, is to attempt to model the complete XPS spectrum. Powell [27], and Werner [28] both have articles in this volume which discuss this.

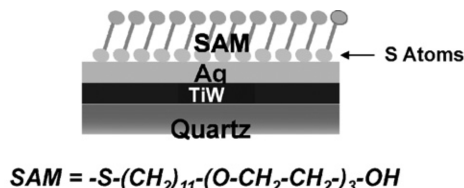
With angular resolved data one can pursue either a qualitative analysis, or attempt various degrees of quantification. The simplest version of a qualitative approach is sometimes called the relative depth profile, RDP, method.

4.1. The RDP method

This is a very simplistic approach intended to indicate which elements (or chemical states, if identified) in an acquired XPS spectrum are, on average, nearer the surface and which further away. If the RDP parameter for each line in an XPS spectrum is defined as:

ln(peak area at a high θ)/ln(peak area at low θ)

Then, to a first approximation, an element located closer to the surface will have a higher RDP value than one deeper into the bulk. There is no depth scale. The values can be considered simply as indicating a layer ordering. The approach is very similar to one originally proposed by Seah et al. [29]. As an example, we show a set of RDP parameter values for a self-assembled mono-layer, PEO-Thiol, on a Ag substrate, of Fig. 7. We chose this as an example because the mono-layer chain length is about 30 Å and it is well-known that the S atom is located at the interface, and that the O containing segment of the molecular chain is outermost. The RDP values immediately confirm this, with the Ag(3d) peak having the lowest value, followed by the S(2p), then the C(1s) of the hydrocarbon chain, and then the C(1s) and O(1s) of the ether unit with the highest values, Fig. 7. There are many other best case examples like this in the literature involving self-assembled mono-layers.



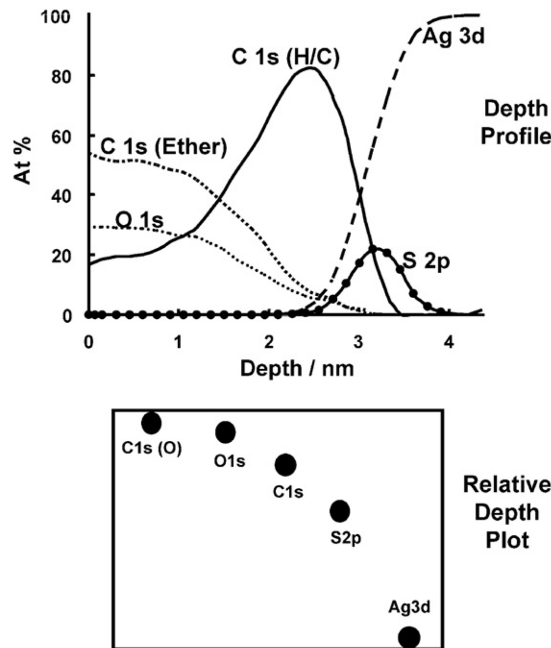


Fig. 7. RDP and maximum entropy profiles for a thiol SAM on Ag.

Fig. 8 is an RDP plot for one of the ~75 Å thick $Si/SiO_2/SiO_xN_y$ structures referred to earlier (Table 3 and Fig. 6). It is trivial to conclude that the N content is confined more to the outer region than the O content, and that an assumption of uniform depth composition will overestimate the N dose, as happened. Nothing further, however, can be extracted from the RDP concerning the depth profile.

Another example of the practical usefulness of the RDP approach in the SiO_xN_y processing development area is shown in Fig. 9, which shows the N(1s) XPS region for a film of about 20 Å thickness on bulk Si, made by a remote plasma deposition, RPD, process, but before a final annealing step. Data at high and low θ are shown, corre-

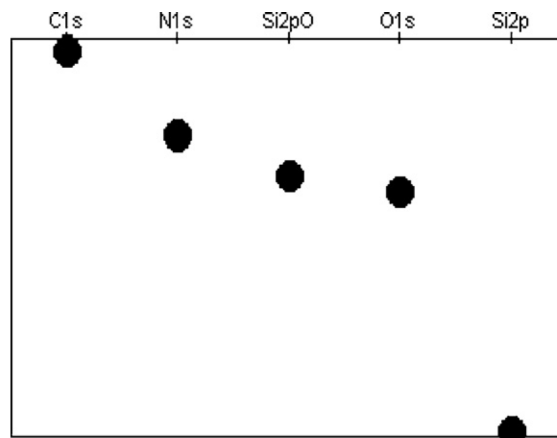


Fig. 8. RDP for the SiO_xN_y samples in Table 3 (all give the same result).

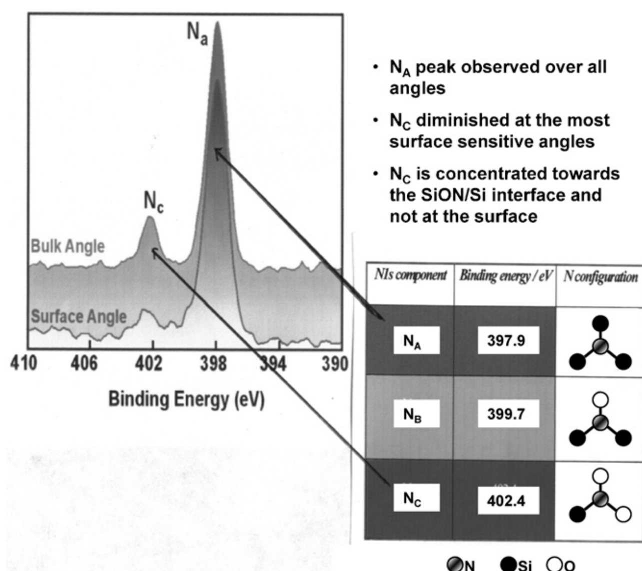


Fig. 9. N(1s) XPS of a ~20 Å SiO_xN_y film, at bulk and surface sensitive angles. The nitrogen, N_C, is clearly closer to the Si substrate interface than is N_A.

sponding to higher surface sensitivity and more bulk sensitivity, respectively. There are two N species present. N_a corresponds to the expected N(1s) position for the N of SiO_xN_y. N_c has a large positive chemical shift and represents a highly oxidized form, probably NO_2. The ratio of N_a/N_c doubles when going from the surface sensitive to the more bulk sensitive angle, showing unambiguously that the oxidized form, N_c, is concentrated strongly towards the SiO_xN_y/Si substrate interface.

Fig. 10, however, shows what happens when a traditional depth profiling using Ar sputtering is attempted. The oxidized N_c species disappears after the initial 1–2 Å of sputtering. This kind of data was obtained by several development groups in different companies, prior to the use of the angular resolved approach, and was initially erroneously interpreted, not surprisingly, as implying that the oxidized N_c species was present at the outer surface of the SiO_xN_y. This is an example of a serious sputtering artifact confusing the XPS interpretation. The oxidized form of the N is not thermodynamically stable, as is revealed by the subsequent annealing step, which converts it to bulk SiO_xN_y. This can be demonstrated by XPS (not shown here) where the N_C component disappears upon annealing, and the N_A increases slightly in intensity. What is actually happening during sputter profiling is that the Ar beam easily penetrates the whole 20 Å film and converts the unstable oxidized N_C species at the substrate interface to bulk SiO_xN_y type N_A, just as an anneal does.

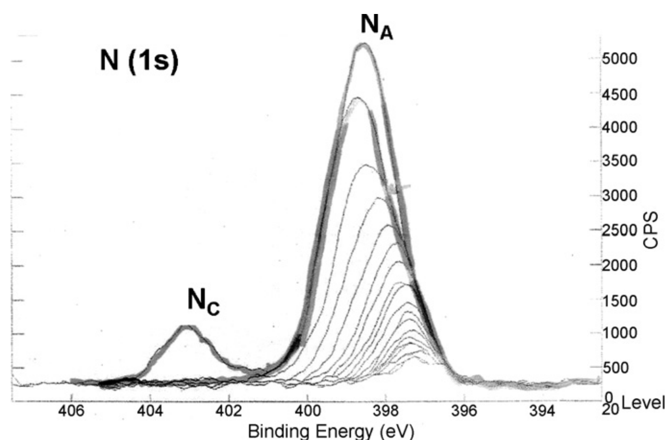


Fig. 10. N(1s) spectra during traditional Ar^+ sputter depth profiling of the ~20 Å SiO_xN_y sample from Fig. 9. Note the peak N_C disappears completely after the first brief sputter, even though Fig. 9 proves it is at the Si substrate interface, not at the outer surface.

Another example of how misleading an attempted sputter profile can be, occurred in the early development work for high *k* gate oxides, involving HfO_2. Depth profiling by ion sputtering appeared to indicate that metallic Hf was penetrating into the Si substrate subsurface, when, in fact the ion sputtering was reducing the HfO_2, leaving small clusters of metal Hf, which had a very low sputter rate and so remained on the (now very rough) surface after the Si interface had been reached and penetrated by the sputter process. Non-destructive angle resolved XPS showed no evidence of the presence of metallic Hf, and no evidence of any Hf increased concentration at the substrate surface. To convince the process engineers, however, it was necessary to polish in from the backside close to the interface, and then do a sputter profile into the HfO_2 from that direction, demonstrating that Hf metal was not present in the Si substrate near the interface [30].

The simple RDP approach can be considered a layer ordering method, as stated earlier. As such, it will *only work reliably* if the material under consideration does, in fact, have a layered structure. This is often the case in the technologies concerned here, as shown by the examples given above. If, however, the layered structure has the same species in more than one layer, the RDP approach is not capable of establishing this. It will average the layer position.

Another problem arises when comparing RDP values for XPS peaks with widely separated BE's. The difference in λ for the different KE photoelectrons will modify the simple RDP interpretation, but approximate correction can be made for this fairly easily. Cumpson [31] has raised these issues, pointing out that the RDP approach is neither truly relative (the λ difference problem), nor really a depth profile (no depth scale and no means of estimating the relative thickness of each layer). Practical experience, however, shows it to be an extremely valuable approach, if the constraints discussed above are kept in mind.

4.2. Modeling the ARXPS response

The alternative approach to RDP is to attempt to model the angular response to obtain a quantitative depth profile. This involves, as a primary step, generating trial profiles, calculating the change in peak relative intensities as a function of angle according to Eq. (3), and then minimizing the χ^2 fit of the relative intensities of the trial profile values, using an iterative procedure, against the experimental values.

There are major problems that limit the likely success of this procedure. The first is that there can never be a unique fit with this inversion process, only a plausible one. The second is that small changes in the experimental ratios can lead to large changes in the trial profile giving a best fit, i.e. the resulting profile is very sensitive to the quality of the data (not very robust).

To mitigate relying entirely on an over-sensitive χ^2 fit, the maximum entropy approach was originally suggested by Smith and Livesey [32], and was implemented several years ago for SiO_xN_y films [33]. This approach balances the quality of the χ^2 fit against a calculated entropy optimization term, i.e. the entropic probability of being in that concentration configuration at that depth. In this approach the joint probability function, *Q*, is minimized, instead of the χ^2 fit:

$$Q = \alpha S - 0.5\chi^2 \quad (5)$$

where *S* is the calculated entropy for the trial fit and α is an empirical term used to decide how important the entropy term is compared to the χ^2 fit. αS is basically a smoothing term and cannot have any unique value for a given system, because how much weighting

should be given to S, compared to χ^2, will depend on the quality of the experimental data being subjected to the χ^2 fit.

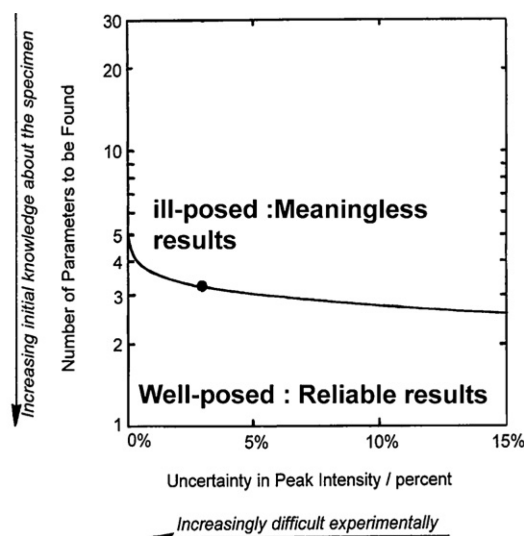
The Thermo Scientific Thetaprobe (or Theta 300 for full 300 mm wafers) allows simultaneous collection of angle resolved XPS data (θ = 23–73°) by using a wide angled input lens system and dispersion of the electron energies and electron trajectories orthogonally through the hemispherical analyzer [34]. The increased speed of data collection, plus simply not having to deal with tilting wafers and sequentially acquiring data is what has made the use of full ARXPS, in a metrology mode, acceptable to the wafer industry.

Initially, even though the modeling software was incorporated as an application with the Thetaprobe, actual usage, incorporating the maximum entropy approach, involved major operator interventions to establish a viable recipe (what quality of spectra to require; how many iterations of trial structures to run; what values to use for the constant needed to turn the profile into a scale with depth, and, in particular what to pick for α). More recently, automated procedures (applications) have been developed for use with defined appropriate situations, so that the expertise/experience needed is reduced or eliminated. We follow this evolution (or at least our evolution) below.

We have been involved in ARXPS for development and product SiO_xN_y gate material for about a decade, and HfO_2 gate material, TiN and TaN barrier layers, and other wafer processing films in the more recent years. Our earlier attempts on SiO_xN_y gate material used a full bore maximum entropy approach, and were useful in teaching us the limitations of the method, and establishing what not to do.

First one has to accept that a unique solution cannot be provided with this data inversion process, only a plausible one that has minimized Q for the trial profiles (apparent convergence may just represent a local minimum). Excellent S/N is required because, as mentioned earlier, the method is not robust and small fluctuations in signal can result in significant alteration to the best fit model profile. Even if there were no noise, N, the signal, S, has to be separated from the background, B, at each angle, and the ability to achieve this separation correctly is likely to be a limiting factor. Cumpson [31] has stated that even if we can reduce the uncertainty in the true peak intensity to 3%, the maximum number of independent parameters that can be reliably extracted, per element, is 3. To extract four parameters requires an uncertainty of 0.5% in true signal intensity to be achieved (see Fig. 11). A profile of the composition, with multiple elements present, and data acquired at multiple depth positions, involves a very large number of parameters if not constrained; one for each element at each depth, and this is the problem. The introduction of the entropy term, S, and minimization of Q instead of χ^2, is an attempt to "regularize" an ill-defined problem. Appreciation of this issue leads us towards constraining ARXPS modeling sensibly. In an entirely unknown complex situation (multiple species, no knowledge that there is even a layer structure, and no constraints on composition with depth), the full maximum entropy modeling approach should not be trusted, even though apparent fits may be obtainable. The corollary is that the more constraints that can be applied, the fewer are the parameters left to be determined, and the more reliable will be a modeled profile fit to the data. Of course any constraints applied have to be known or assumed to be realistic. Fortunately, for the technological areas we are discussing, film growth/deposition in the wafer processing industry, many constraints can be applied with some certainty.

First the layer model itself is usually appropriate (a linear relationship in Eq. (3), over the range of thickness and angle conditions where the equation is expected to be valid, is consistent with a layer model, and the lack of such a relationship may be an indication that a layer model is not appropriate). Second, for an Si(1 0 0) substrate one expects a sharp interface, at least for non-reactive material,



Peter Cumpson, IUVSTA Workshop on ARXPS, Cancun, March 2007

If both precision and accuracy allow separation of peak intensity from back-ground contribution to 3% uncertainty, 3 parameters can be determined for each species in the spectrum.

To determine 4 would require 0.5% uncertainty!

Fig. 11. Cumpson criteria (Ref. [31]) for meaningful parameter determination from ARXPS.

such as thermally grown SiO_2. If in doubt, TEM cross-section analysis may be used for confirmation. So in the maximum entropy modeling, the Si(1 0 0)\film interface should be constrained so that the entropy term cannot force film material into the substrate. This will happen if α is too large, and there is no interface constraint, since entropy always wants to disperse material. This artifact will become worse the deeper the interface is and the deeper one allows the model to run, because the contribution to the total XPS signal being χ^2 fit is exponentially decreasing with depth, and so cannot counteract the entropy term in Eq. (3) at large depths. What constitutes too deep? This is λ dependent and therefore material dependent. Anything beyond 3λ is hopeless and constraining to 2λ is recommended. If one is sure that the interface is sharp, then also constrain the model to allow only a small percentage of the film thickness for the interface width, and do not allow the modeling to run much deeper than the interface.

An inappropriately large α term, plus not constraining the interface width, will lead to an extremely smooth and slowly varying modeled distribution (S is over-emphasized), with the film components trailing into the bulk and never reaching 0% concentration, and the substrate Si never reaching 100%. At the other extreme, if α is much too small and/or the experimental data is of poor statistical quality, the χ^2 fit, without additional constraints, may find best solutions with physically unreasonable sharp spikes in the depth distribution. The maximum entropy modeled depth distribution of the SAM, PEO-Thiol, in Fig. 7, was done using very high quality data and has had α empirically adjusted according to the discussion above, and, as can be seen, it gives the expected (already known really) profile.

Earlier, we discussed the effects of not considering elastic scattering on thickness determination. These effects sometimes show up also in modeled depth profiles. In the case of films on Si(1 0 0), including the data up to the highest angles will sometimes result in a small apparent elemental Si^0 contribution appearing at the surface. This is because, without specifically including elastic scattering, the model is forced to interpret all signal emerging at high

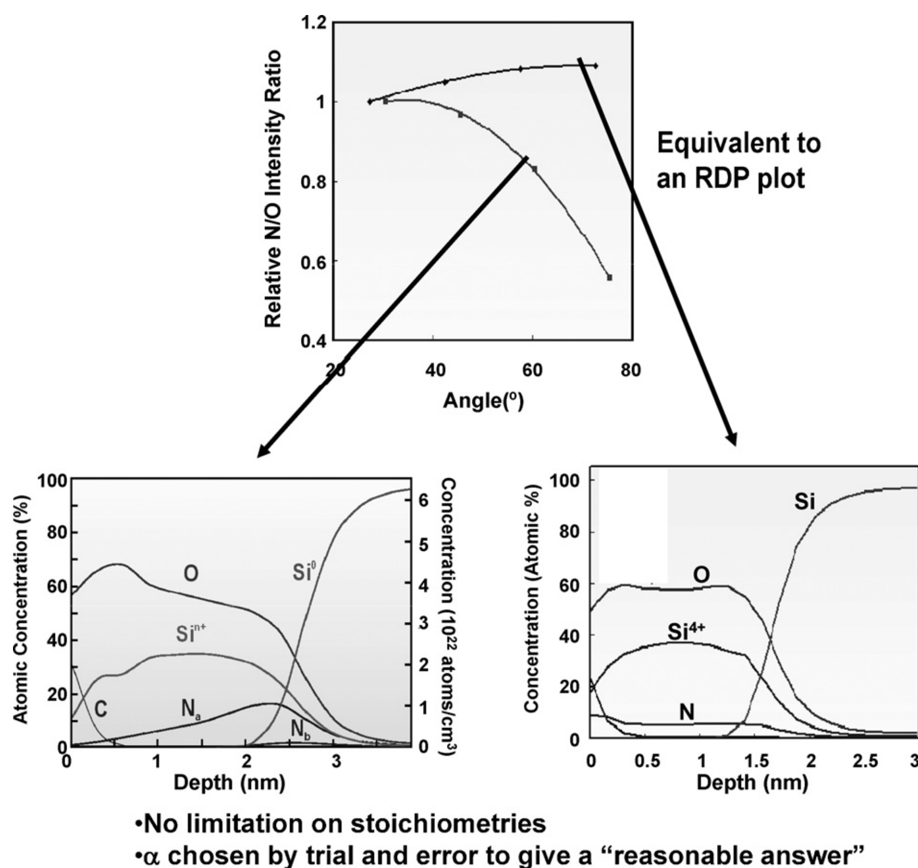


Fig. 12. Maximum entropy derived profiles for two very different N distribution SiO_xN_y samples.

angles with no inelastic scattering (i.e. in the XPS peak) as originating from close to the surface. However, an elastic scattering event taking place close to the surface, for an electron coming from the substrate at low angle (normal to the surface), can also lead to high angle emission from the surface (see Fig. 2, bottom). One can easily check whether such an apparent surface Si^0 contribution is genuine or not by comparing a model fit including high angles to one not including them. If the apparent surface Si^0 component goes away on excluding the high angles, it is an artifact. This, of course, does compromise our ability to maximize surface sensitivity, since we are discarding the most surface sensitive section of the data.

Fig. 12 shows one of our earliest maximum entropy attempts at modeling the depth distribution through two SiO_xN_y films prepared in very different ways, and having substantially different thickness. The Si interface half width was constrained to 10% of the depth modeled in both simulations, and the value for α was adjusted by trial and error in an attempt to get the Si elemental substrate value to reach 100% in the bulk (note that this was never quite achieved). Composition was not constrained in any way and the C1s of the substantial hydrocarbon contamination was included. The two samples are clearly quite different (one is the unannealed process situation referred to earlier in Fig. 9, with an oxidized N species located nearer the lower interface). What level of confidence can one put in the comparison between the two films? Our opinion now, though we were over-optimistic back then, is that beyond the statement that one film shows a strong bias of N towards the interface, and the other a more or less a flat distribution, there is nothing further that the quality of the actual data, and the relatively unconstrained nature of this particular maximum entropy simulation, can reliably establish. The valid conclusions about the differences between the two films can be equally well established directly from the behavior of the N1S/O1s ratio as a function of θ, Fig. 12, which is simply the equivalent of a RDP plot done at more than two angles, without any extra attempt at quantitative modeling.

In practical wafer processing examples there will always be adsorbed hydrocarbon on XPS examined material, since it is ambient exposed. This constitutes an additional species, and layer, to model (more parameters), which is often irrelevant to the information being sort in the depth profile. If one is sure that there is no relevant carbon species present (e.g. carbide) why include C(1s) in the modeling? Or even spend acquisition time collecting it? This is quite different from the situation with ellipsometry, where the hydrocarbon layer will contribute in an unknown way to the optical fitting, leading to errors.

A serious problem with much of the maximum entropy modeling early on, and largely ignored, was that if all elemental percentages were allowed to float freely, stoichiometries that made no chemistry sense could result from the fitting procedure. For example, for SiO_xN_y, if the Si^0(2p) substrate, Si^{4+}(2p) oxidized, O(1s) and N (1s) peak intensities were allowed to float freely, large amounts of free Si^{4+}, O, or N, might appear in the best fit model depth distribution, at some particular depth, which is clearly nonsense chemically. This led, eventually, to more sensible constraints on stoichiometries; essentially not modeling the profile in elemental form, but in chemical species form. This has resulted in much more reproducible and believable profiles (reduction of parameters), but it is essentially moving the analysis in the direction of writing individual algorithms (applications or recipes) for very specific situations. This is quite typical in the wafer processing industry, where often a new application, or recipe, must be developed every time there is even a minor change in product material.

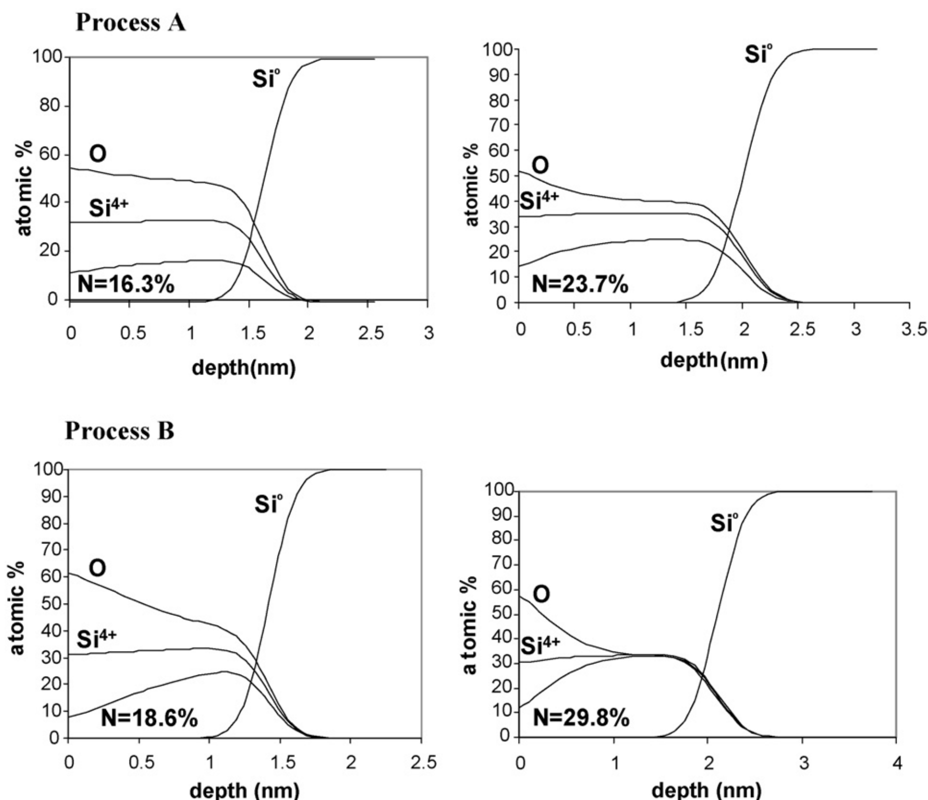


Fig. 13. Maximum entropy profiles for SiO_xN_y films made by slightly different processes A and B.

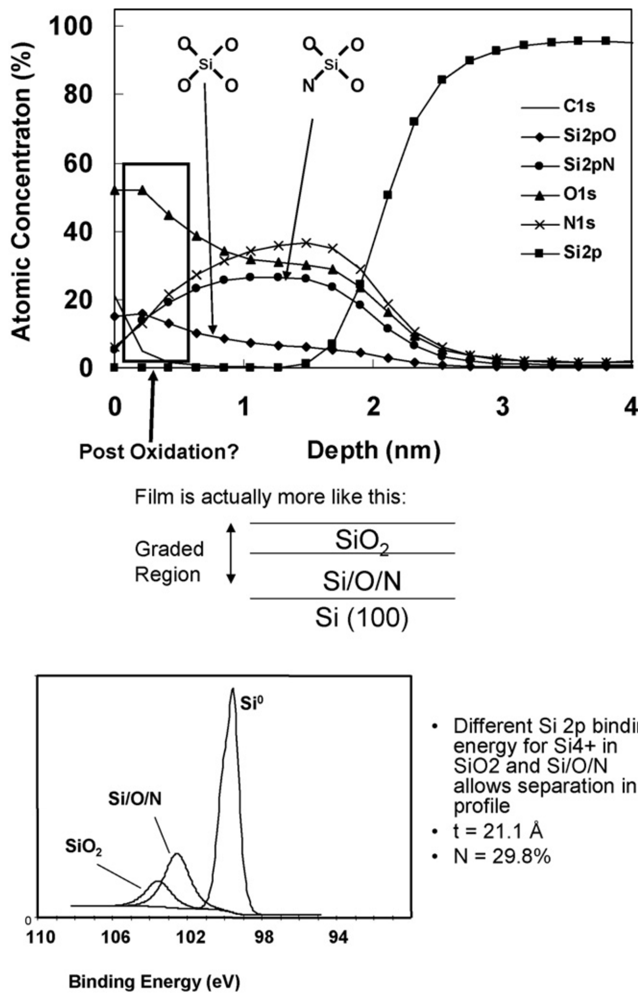
Fig. 13 shows the maximum entropy simulation results for two samples prepared by a process A, and two by a modified process B. This modeling was done a couple of years later than that in Fig. 12 above (unrelated samples), and additional constraints were applied. Only the Si oxidized, Si elemental, and N(1s) data were modeled. The C(1s) from the hydrocarbon was not included and the O(1s) content was back fit assuming a stoichiometry of mixed SiO_2/Si_3N_4 composition for the SiO_xN_y. An α was used which now allowed the substrate Si^0 elemental signal to actually reach 100%. The main believable conclusion here, from this modeling, is that the near surface region of both samples is more oxidized than the deeper region. This is completely unsurprising, since the films had been air exposed for some time and we know from time dependent studies that re-oxidation, with loss of N, does occur as a function of time. The next conclusion would be that the higher the N content of the film, for both processes, the more pronounced is the surface re-oxidation; again not surprising from a chemistry point of view.

Is process B producing a measurably different N depth distribution from A? Trying to answer this question was actually the reason for the study in the first place. There is certainly more surface re-oxidation for the B process samples, which is about all you can say, but this may have nothing to do with the actual deposition process itself. The detailed differences in behavior of concentration versus depth for process A versus B should not be trusted, however, since there are far too many floating parameters.

For the highest N% case, the surface re-oxidation can actually be directly observed from the Si^{4+}(2p) XPS region, because Si(2p) in SiO_2 has a slightly higher BE than that in SiO_xN_y, and there is sufficient intensity and statistics to be able to deconvolute the two signals, Fig. 14. The maximum entropy profile for the Si^{4+} can then be modeled as its two components, Fig. 14, and one can see that the SiO_2 component follows the O(1s) trend and the SiO_xN_y component follows the N(1s) trend.

An additional reason for studying this set of samples was that a TOF-SIMS destructive depth profiling analysis for a parallel set of wafers indicated large additional N concentration spikes right at the surface (the first 2 Å). Such concentration spikes should certainly show up in XPS. We suspect that, in this case, the presence of these in TOF-SIMS, even using the well defined Douglas protocol [35], was a SIMS surface artifact, possibly representing trace surface contamination by a N containing molecular species with very high SIMS cross-section, or reaction of the N in SiO_xN_y at the surface to yield an N species with a high SIMS cross-section. Whereas SIMS, of any variety, has the highest probability for artifacts in the first few Angstrom of profiling, for a non-sputtering XPS analysis, material located right at the surface is actually the easiest to quantify because of the maximum sensitivities there, and cross-sections that are known and largely unaffected by matrix effects.

Earlier (Section 3.3.2.1) we used the data in Table 3 to illustrate how an incorrectly assumed uniform depth distribution led to a very incorrect N dose estimate, and how a simple RDP could have caught the problem (Section 4.1). We now return to examine Table 3 in more detail. Samples 1–3 represent product using one manufacturing process, samples 4–6 another. The thickness and dose measurements reported in Table 3 were originally made to establish any differences between the two products, using a single angle XPS metrology tool. The samples were then passed to the Theta 300 laboratory for “verification”, but the expected actual gross structure, represented in the schematic of Fig. 6, was not known to the Theta 300 XPS analyst prior to the measurements (though presumably it was known to the process engineers, since it looks like a deliberate two step process). We showed earlier how the RDP approach, using the Theta 300, was sufficient to clearly demonstrate that the N, on average, was closer to the surface than the O. The question now is can an attempt at a quantitative profile add anything useful? Fig. 15 shows a modeled depth profile determined by an entirely automated algorithm designed specifically for



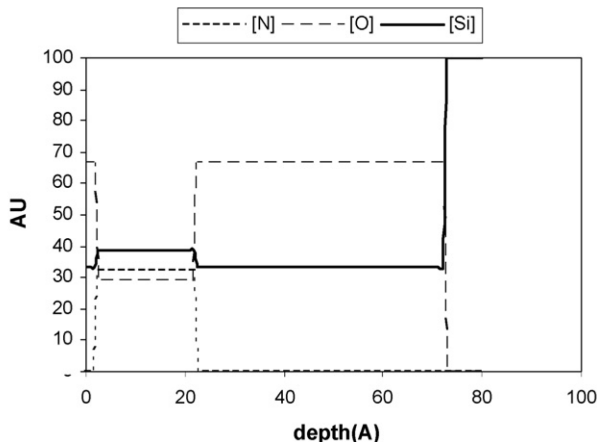


Fig. 14. Example of chemical depth profiling: distinction of Si in SiO_2 from Si in SiO_xN_y by using Si(2p) chemical shifts.

SiO_xN_y (Trinity application [36]), for one of the samples in Table 3. It consists of a box model constrained to three layers with zero interface widths, where the layers are SiO_2 (outer), SiO_xN_y (center), and SiO_2 (lower interface). The SiO_xN_y layer is further constrained to have the stoichiometry of an SiO_2/Si_3N_4 mixture. There are no floating element concentrations; everything is modeled as stoichometric compounds. This is not a maximum entropy procedure (i.e α

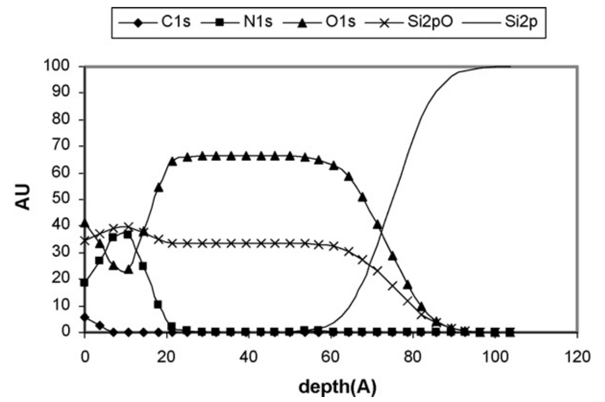


Fig. 16. Automated maximum entropy depth profile for Sample 4 of Table 3.

is 0), but just the best χ^2 fit to the constrained model. The modeling has therefore actually been reduced to a three parameter fit; two independent thicknesses (the total thickness is an input parameter determined in the standard way, Section 3.2 but using a λ value interpolated from the values for SiO_2 and Si_3N_4 and the XPS measured N%), and one composition variable, the SiO_2/Si_3N_4 ratio in the SiO_xN_y layer. As such, it has a chance of meeting the Cumpson requirement for validity, provided the statistical quality of the data is high enough, and should be reliable. So, in this case, where the goal of the modeling has been very much reduced, we actually do obtain additional information compared to a qualitative RDP plot. The N dose in the film can now be determined directly from the area under the SiO_xN_y box in Fig. 15. These were the values shown in the "ARXPS corrected N dose" column of Table 3.

Fig. 15. Trinity algorithm fit to ARXPS data for sample 4 of Table 3.

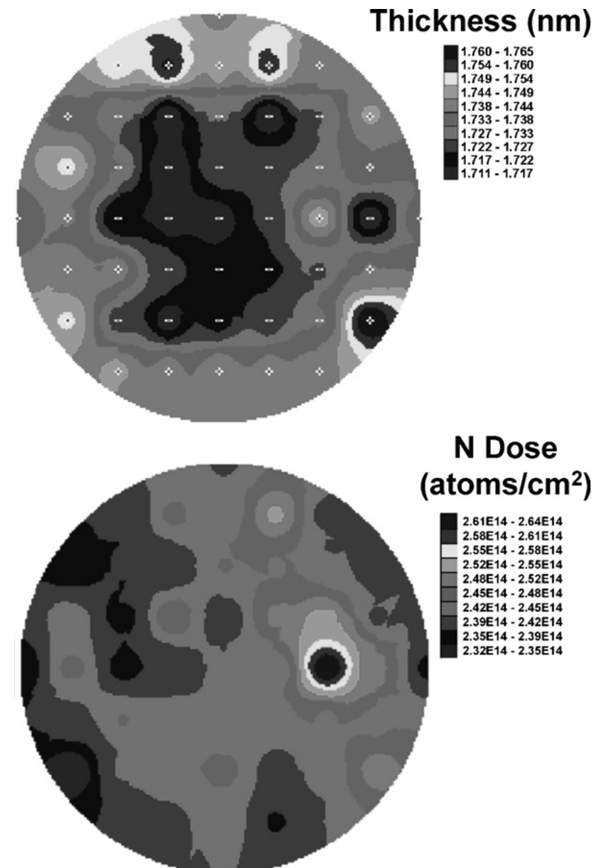
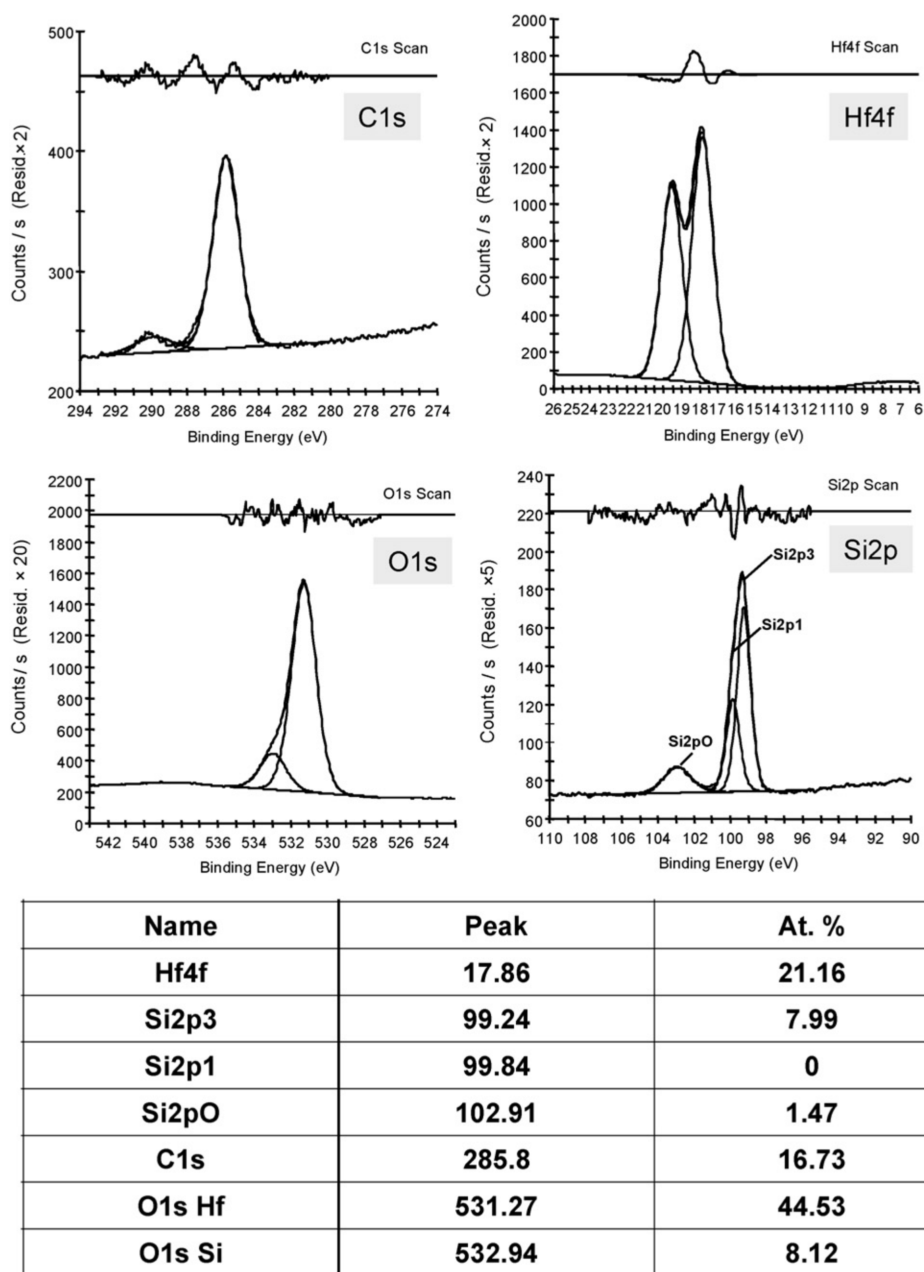


Fig. 17. Simultaneous nitrogen dose and thickness metrology for a 300 mm wafer (Ref. [36]).



Name	Peak	At. %
Hf4f	17.86	21.16
Si2p3	99.24	7.99
Si2p1	99.84	0
Si2pO	102.91	1.47
C1s	285.8	16.73
O1s Hf	531.27	44.53
O1s Si	532.94	8.12

Fig. 18. Angle integrated XPS for a ~25 Å HfO_2 film.

Fig. 16 is a maximum entropy simulation of sample 1 from Table 3, using a completely automated application from Thermo. As for the Trinity algorithm results in Fig. 14, the total thickness is an input parameter. The chemical species present are constrained to be SiO_2, Si_3N_4, Si elemental (substrate), hydrocarbon, and free O. The value of α is calculated by the algorithm after examination of the statistical quality of the collected data. The interface half widths are constrained to 10% of the depth modeled, but unlike the Trinity model there is no a priori three layer constraint. Interestingly, the resulting maximum entropy profile is quite similar, but not identical, to the Trinity result. The SiO_xN_y layer is somewhat thinner, and the outer re-oxidation layer is not completely SiO_2, but just shows an increased percentage of SiO_2 compared to the SiO_xN_y layer. We do not really know how much real additional information there is here, or whether the differences partly reflect the blurring of the interface introduced by restoring an α term, but the important advance is that a known layer structure was not assumed, and the algorithm is generic and not specific to SiO_xN_y. It has also been used successfully for other gate material, such as HfO_2-based material (see later).

The above discussion, whether using a three layer, zero α model, or an entirely automated maximum entropy algorithm, shows that, with sufficient constraints and sufficient quality of data, it is certainly possible to extract useful information reliably beyond the RDP approach. The resulting profile, however, can only ever be an approximation, not a fully quantitative representation, and the best way to present the results, without leaving them open to gross over interpretation, remains a problem.

The results in Table 3 for N dose, obtained by either assuming a uniform N distribution or using a modeled profile approach (whether an α term is included or not) raise an interesting point related to the relative importance of accuracy and precision in

the metrology of thin films, and the different effects that different assumptions make. The N dose numbers in Table 3, using a uniform N distribution approach, from a single fixed angle measurement (from the ReVera instrument) are very inaccurate, being a factor of approximately 2, too large. The spread in the determined (inaccurate) values for the uniform N distribution approach, however, is smaller than the spread in the (more accurate) ARXPS depth corrected approach. Since the actual nominal N dose applied to the wafers was the same in all cases, this might suggest that the increased spread in the ARXPS depth corrected case is due to small variations in the model depth profile which are not, in fact, real. So, it may be that, for dose *accuracy* the full depth profile is needed (but small variations should be dismissed as artifacts), whereas for dose *precision* (necessary to correctly distinguish small changes in dose from sample to sample having nominally the same, but unknown profile) the assumed uniform N distribution approach is better. This situation also comes up in the next section.

4.3. Simultaneous thickness, N dose, and N distribution wafer metrology by ARXPS

Understanding the differences between using a uniform N depth distribution assumption (possible poor dose accuracy; but good precision) and using a modeled N depth distribution to calculate dose (poorer dose precision, but better accuracy) raises the possibility of using a combined approach for providing simultaneous metrology for thickness, N dose, and characterization of N depth distribution, which can meet customer demands, in terms of speed and reliability. The 49 point simultaneous mapping of thickness and N dose, at high precision, across a wafer is shown in Fig. 17 using the Theta 300. It was determined using the following protocol [36].

Angle integrated N(1s) and Si(2p) spectra are collected at all points in non-angular resolved snapshot mode. The thickness is calculated from the Si oxidized/Si elemental ratio, using prior work on SiO_2 gate layers, where thickness was determined by the full ARXPS approach, which allowed a calibration for angle integrated snapshot data, so that thickness could then be determined directly from the latter. The same calibration was used here for the SiO_xN_y. An uncorrected N dose is then determined using this determined thickness and the measured N% from the XPS standard table at each of the 49 sites, *assuming uniform N distribution*. To obtain a relative standard deviation, RSD, of approximately 1% in the calculated N dose this way requires about 50 s acquisition time for the Si(2p) region, and 75 s for the N(1s) region, for the thickness and dose value ranges used in this technology.

Next, at one point on the wafer (usually the center) a full ARXPS set of data is obtained and modeled using the Trinity algorithm. A corrected N dose is then derived from the N atomic percentage in, and thickness of, the SiO_xN_y middle layer of the Trinity profile. About 3 min acquisition time is required for acceptable statistics. This same correction factor is then applied to all the other points, which therefore assumes that the N distribution is constant over the wafer.

Using the above procedures the precisions and speed of the measurement are optimized. The precision of the thickness measurement is simply determined by the statistics of the measurement of the ratio of Si oxidized/elemental, and that of the N dose by the statistics of both the Si oxidized/elemental ratio, and the N% determination. The N dose precision is therefore always poorer than the thickness precision.

Of course, if there are really variations in N profile, for a constant N dose, as a function of the position on the wafer, these will show up as apparent variations in N dose using this procedure. If this is suspected, full profile modeling can be done at the positions concerned. It is likely, however, because of the restricted ability of the modeling to reliably detect small variations in depth distribution, that this procedure will only be able to distinguish the two possibilities (change in dose, fixed distribution; change in distribution, fixed dose) if there are substantial differences. In addition, as in the case of the data from Table 3, the assumption of differing depth distributions at each point and subsequent attempted modeling will very likely seriously compromise precision of the N dose measurement.

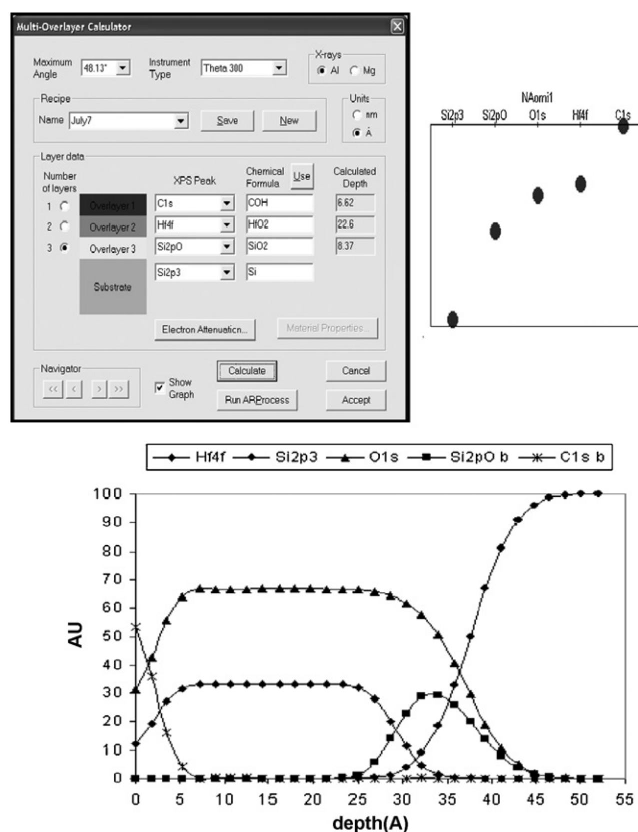


Fig. 19. RDP, thickness calculations, and fully automated maximum entropy profile for the HfO_2 film giving the spectra of Fig. 18.

5. Beyond SiO_xN_y

The forefront in gate technology currently involves the use of HfO_2. The chemistry issues here include controlling interface diffusion between deposited HfO_2 and a passive layer of SiO_2 layer at the Si substrate interface, possibly forming Hf_2SiO_4, for instance. Discussion of this is beyond the scope of this review, but in Figs. 18 and 19 we show a typical example. Fig. 18 shows the angle integrated C1s, Hf4f, O1s and Si2p spectra, with peak assignments and atomic % values. Fig. 19 shows an RDP plot and an automated ARXPS maximum entropy modeled depth distribution from angle resolved data. The algorithm used is the same as above for SiO_xN_y, but the modeled species are constrained to be COH, HfO_2, SiO_2, and the Si substrate. The total thickness, used as an input parameter, is no longer simply obtainable from a Si^{4+}/Si^0 ratio, but has to be obtained from more complex multilayer equations [17], assuming a COH contamination layer, a layer of HfO_2, and an interface layer of SiO_2.

The roughly 10 Å thick SiO_2 layer at the interface is clearly indicated by the modeling. A check on the reliability of this profile is also available directly from the O(1s) spectrum, since the peaks from O in HfO_2 and SiO_2 are separable (Fig. 18), and the O(1s) 532.95 eV peak profile follows that of the Si^{4+} profile, if modeled separately (not shown).

Successful applications of the same algorithm, to other HfO_2-based film structures, have also been reported by Conard [37].

6. Conclusions

XPS is being used today in the semiconductor industry not just in its traditional mode of providing surface chemical state information, but also for full compositional analysis of these films. This is because the films concerned are now often of comparable thickness to the probing depth of the technique. Since that probing depth can be varied by changing the detection angle with respect to the surface normal, it is also, in principle, possible to provide non-destructive depth profiling, though in practice one has to be very careful about how to go about this. Very high quality data is needed and constraints must be applied to reduce the number of parameters involved in fitting the data. In addition to using these attributes for full characterization of the material of the films, XPS, and ARXPS in particular, has been developed as a metrology tool for controlling thickness, composition, and depth distribution to adequate precision, with sufficient speed and robustness (instrument uptime and non-expert operation) to meet industry requirements for, example, gate oxide material, in the most advanced transistor production (at the 32 nm node).

Acknowledgements

The authors would like to acknowledge contributions from several people, without which this review would not have been possible. Discussion with Peter Cumpson has been particularly enlightening concerning the fundamental limitations in ARXPS depth profiling. The calculations and measurements reported by Cedric Powell, and by Martin Seah, and coworkers, and their discussions of them have been most valuable. Long involvement by one of us (CRB) with Alain Diebold has led to an appreciation of how metrology is used in the semiconductor industry. John Wolstenholme of Thermo Fisher Scientific has contributed in many ways (discussions, data analysis, and presentations at conferences) over a number of years. Finally we appreciate the support provided by Yuri Uritsky at Applied Materials.

References

[1] C.R. Brundle, M.W. Roberts, Proc. R. Soc. London:+ Ser. A 331 (1972) 383.
[2] C.R. Brundle, J. Vac. Sci. Technol. 11 (1974) 212.
[3] K.S. Kim, W.E. Baitinger, J.W. Amy, N. Winograd, J. Electron Spectrosc. Relat. Phenom. 5 (1976) 351;
T.J. Chuang, C.R. Brundle, K. Wandelt, Thin Solid Films 53 (1978) 19.
[4] A. Joshi, L.E. Davis, P.W. Palmberg, in: A.W. Czanderna (Ed.), Methods of Surface Analysis, Elsevier, Amsterdam, 1975 (Chapter 5).
[5] J.H. Scofield, J. Electron Spectrosc. Relat. Phenom. 8 (1976) 129.
[6] www.revera.com.
[7] www.thermo.com.
[8] W. Johnson, C. Yarling, Appendix 23, in: Y. Strausser (Ed.), Characterization in Silicon Processing, Butterworth-Heinemann, Boston, 1993.
[9] C. Hayzelden, Gate dielectric metrology, in: A.C. Diebold (Ed.), Handbook of Silicon Semiconductor Metrology, Marcel Dekker, Inc., 2001.
[10] D.A. Muller, T. Sorsch, S. Moccia, F.H. Bauman, G. Timp, Nature 399 (1999) 758.
[11] D. Matsushita, D. Muraoka, K. Nakasaki, Y. Kato, Microelectron. Eng. 80 (2005) 424;
L. da Siva Zambom, Thin Solid Films 515 (2006) 596.
[12] S. Guha, V. Narayanan, Annu. Rev. Mater. Res. 39 (2009) 161.
[13] http://download.intel.com/pressroom/kits/32nm/westmere/Intel_32nm_overview.pdf.
[14] E. Bersch, M. Di, S. Consiglio, R.D. Clark, G.J. Leusink, A. Diebold, AIP Conf. Proc. 1173 (2009) 55.
[15] S. Walther, R. Liebert, J. Vac. Sci Technol. B 24 (2006) 482.
[16] G. Saheli, G. Conti, Y. Uritsky, M.A. Foad, C.R. Brundle, P. Mack, D. Kouzminov, M. Werner, J.A. van den Berg, J. Vac. Sci. Technol. B26 (2008) 298.
[17] J.F. Watts, J. Wolstenholme, An Introduction to Surface Analysis by XPS and AES, John Wiley and Sons, Chichester, UK, 2003.
[18] J. Wolstenholme, Thermo Fisher Scientific, Private Communication.
[19] C.J. Powell, A. Jablonski, J. Vac. Sci. Technol. A 19 (2001) 2604.
[20] Thermo Application Notes, AN 31016.
[21] P. Mack, R. White, J. Wolstenholme, A. Wright, Advanced Semiconductor Manufacturing Conference and Workshop, ASMC, 2003; and Thermo Application Notes, AN 31025.
[22] M. Seah, et al., Surf. Interface Anal. 33 (2002) 960.
[23] M. Bigi, D. Caputo, Comparison Between Different Noncontact Techniques to Characterize Thin Nitrided Oxides, Issue 27 (November), Future Fab International, 2008.
[24] Data supplied by Charles Evans and Associates/Physical Electronics on the prototype of what is now the ReVera RVX 1000 instrument.
[25] C. Wang, Applied Materials, unpublished data.
[26] S. Tougaard, J. Electon Spectrosc., this volume.
[27] C.J. Powell, J. Electron Spectrosc., this volume.
[28] W.S.M. Werner, J. Electron Spectrosc., this volume.
[29] M.P. Seah, J.H. Qiu, P.J. Cumpson, J.E. Castle, Surf. Interface Anal. 21 (1994) 336.
[30] W. Nieveen, B.W. Schuler, G. Goodman, Appl. Surf. Sci. 231 (2004) 556 (also in AEG Application Notes, AN 400, AN 415 and AN 416).
[31] P.J. Cumpson, Angle-Resolved XPS: the current status and future prospects for angle-resolved XPS of nano and subnao films, in: presented at the 47th IUVSTA Workshop, in March 2007 and summarized in Surf. Interface Anal., vol. 41, 2009, p. 839.
[32] G.C. Smith, A.K. Livesey, Surf. Interface Anal. 18 (1992) 175.
[33] J.P. Chang, et al., J. Appl. Phys. 87 (2000) 4449;
R. Opila, J. Eng, Prog. Surf. Sci. 69 (2002) 125.
[34] Thermo Application Notes, AN 31065.
[35] M. Douglas, et al., J. Electrochem. Soc. 147 (2000) 1893.
[36] J. Wolstenholme, Semiconductor Manufacturing, April 2006 (p. 52). Also presented at the 47th IUVSTA Workshop, as in Ref [31].
[37] T. Conard, Private Communications and to be Published.