

Effects of creep and thermal drift on modulus measurement using depth-sensing indentation

G. Feng and A.H.W. Ngan

Department of Mechanical Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, Peoples' Republic of China

(Received 16 January 2001; accepted 7 January 2002)

In modulus measurement by depth-sensing indentation, previous considerations assume elastic recovery to be the sole process during unloading, but in reality creep and thermal drift may also occur, causing serious errors in the measured modulus. In this work, the problem of indentation on a linear viscoelastic half-space is solved using the correspondence principle between elasticity and linear viscoelasticity. The correction term due to creep in the apparent contact compliance is found to be equal to the ratio of the indenter displacement rate at the end of the load hold to the unloading rate. A condition for nullifying the effect of thermal drift on modulus measurement is also proposed. With this condition satisfied, the effect of thermal drift on the calculated modulus is negligible irrespective of the magnitude of the drift rate.

I. INTRODUCTION

Depth-sensing indentation is rapidly becoming a standard method for measuring the elastic modulus of materials.^{1–6} In the Oliver–Pharr scheme,⁴ the elastic modulus of the specimen is estimated from the unloading segment of the load–displacement curve in which the material is assumed to undergo purely elastic recovery. However, even for metals at room temperature, creep effects may be significant at the peak load. Examples to illustrate this are shown in Fig. 1, which shows the displacement–time curves of aluminium and Ni₃Al samples during load hold at the peak load. It can be seen that the indenter displacement continues to increase in both cases, even after a long hold of 10 min, and that the displacement rate appears to settle to a steady value of approximately 0.02 nm/s in Al and 0.009 nm/s in Ni₃Al. Such a significant creep effect at the peak load may influence the subsequent unloading behavior, especially when the unloading rate is slow. In the extreme case of creep dominating elastic recovery at the onset of unload, the load–displacement curve may even exhibit a “nose.” An example of this is shown in Fig. 2 for aluminium. Figure 2(a) shows load schedules (i) to (iii) for three similar indentation experiments on the same aluminium sample. Load schedule (i) has a very rapid unloading rate, while (ii) and (iii) have a common but slower unloading rate. Schedules (i) and (ii) have a short load hold before unload, while (iii) has a longer load hold. The unloading curves for the three schedules are shown in Fig. 2(b). A conspicuous nose can be seen in the unloading curve for (ii). It can be seen that the nose disappears when the unloading rate is increased [c.f. (i)] or when the

load hold before unload is lengthened [c.f. (iii)]. When a nose appears, the elastic modulus cannot be calculated accurately using the Oliver–Pharr scheme, since the apparent contact stiffness is now negative. Even when a nose does not occur, the presence of creep may lead to serious errors in the estimation of the modulus, unless one corrects for its effect.

In this paper, we present a simple scheme by which we can correct for the creep effects during modulus measurement. We also derive a condition by which the effects of thermal drift on the measured modulus can be nullified.

II. CORRECTION FORMULA FOR CREEP

A. Theory

In the Oliver–Pharr scheme, the reduced modulus E_r is calculated from the contact stiffness S at the onset of unload as

$$E_r = \frac{\sqrt{\pi}}{2} \frac{S}{\sqrt{A_c}}, \quad (1)$$

where A_c is the contact area at full load. The contact area A_c is calculated from the contact depth h_c by assuming a shape function of the indenter, i.e.,

$$A_c = f(h_c), \quad (2)$$
$$h_c = h_{\max} - \epsilon \frac{P_{\max}}{S},$$

where ϵ is a constant depending on the indenter geometry ($\epsilon = 0.75$ for the Berkovich tip). Thus, as can be seen from Eqs. (1) and (2), an accurate estimate for E_r hinges

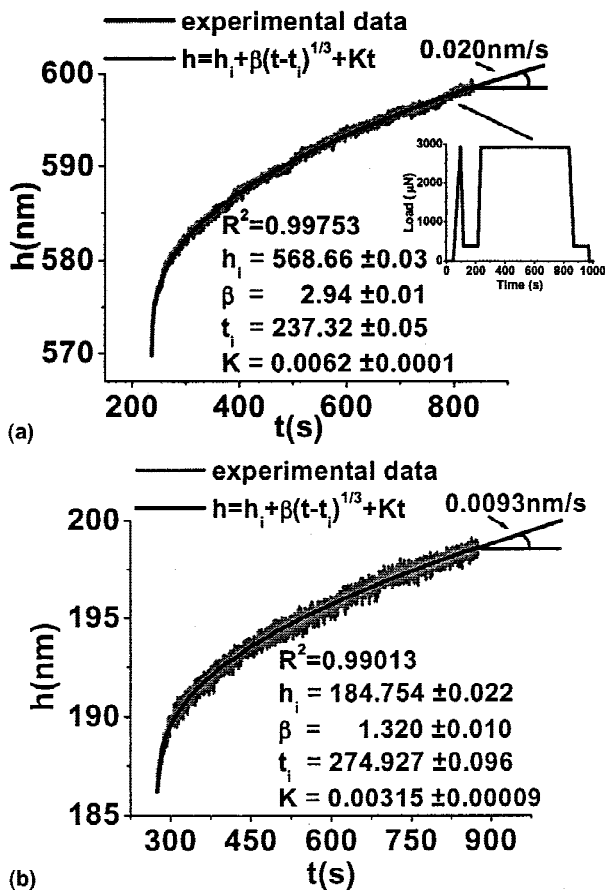


FIG. 1. Displacement–time curve of the load–hold process after correction of thermal drift in (a) Al, load = 2918.1 ± 1.5 μN and (b) Ni₃Al(111), load = 4964.6 ± 2.1 μN. The inset in (a) shows the load schedule.

on the accurate determination of the elastic contact stiffness S at the onset of unloading. When the specimen undergoes purely elastic recovery during unloading, the contact stiffness to be used in Eqs. (1) and (2) would indeed be the observed (or apparent) contact stiffness S_u at the onset of unload. However, when creep occurs alongside elastic deformation, we will show below that, for a load schedule consisting of a load hold followed by unloading, the S to be used in Eqs. (1) and (2) will no longer be the same as the apparent S_u but is related to it by

$$\frac{1}{S} = \frac{1}{S_u} + \frac{\dot{h}_h}{|\dot{P}|}, \quad (3)$$

where the second term above is the correction due to creep and thermal drift. Here, \dot{h}_h is the indenter displacement recorded at the end of the load hold, and \dot{P} the unload rate at the onset of unload.

To deal with creep during unload, one has to work within a viscoelastic framework, although the problem of pure creep during indentation has been solved.^{7,8} In doing so, it may be tempting to assume that the overall

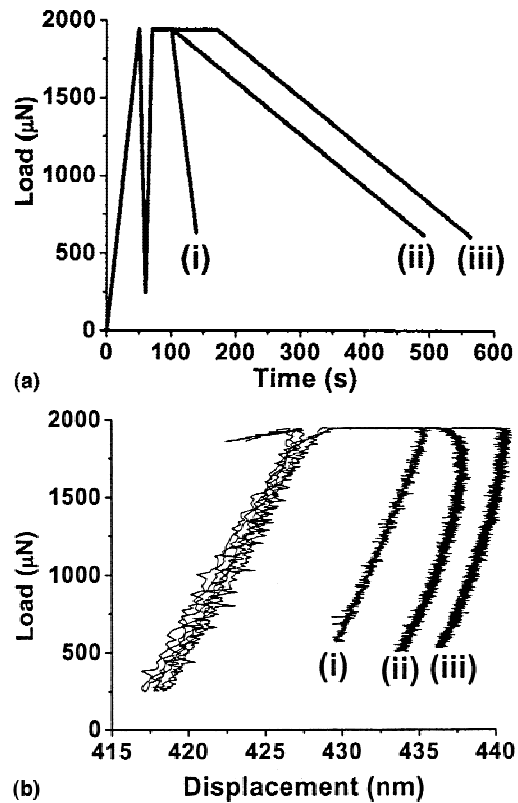


FIG. 2. The effect of creep on the unloading curve after correction for thermal drift in Al.

indenter displacement h is the sum of elastic and creep components and to apply standard spring–dashpot models to such a system. We however observe that, even in the linear viscoelastic situation, h does not superpose linearly in this way. The reason is that h does not only mark the magnitude of the displacement field but also marks the size of the contact zone so that, unlike the case of simple tension where the contact area with the load is fixed, doubling the material’s compliance, for example, will not double h proportionately in indentation. Furthermore, during a transient stage, the creep component of the displacement field will not be self-similar, and hence individual consideration of it will become difficult. Therefore, we believe it is necessary to work from first principles, and this is the aim of this section.

Suppose that the material to be indented is linear viscoelastic with stress–strain behavior represented by a Maxwell spring–dashpot model as shown in Fig. 3. Since, as far as modulus measurement is concerned, we are mostly interested in the load hold and unloading segments of the load schedule, the power-law behavior of plasticity is ignored. In practice, time-independent plasticity effects can be eliminated in subsequent considerations if the indent shape is introduced beforehand by a quick load-ramp and unload cycle as shown in the load schedules in Figs. 1 and 2. The assumption of linearity in

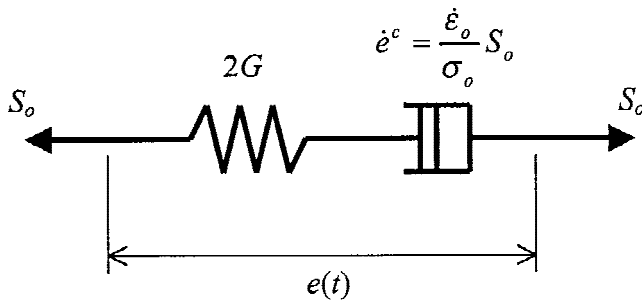


FIG. 3. Maxwell model for linear viscoelasticity. S_o is a deviatoric stress and e the resultant deviatoric strain.

the viscoelasticity allows an analytical solution to be found. Power-law creep may well occur in practice, but as we will see later, the good agreement with experiments serves as the main justification for adopting the linear approximation here. The solution can be derived using the correspondence principle between elasticity and linear viscoelasticity suggested by Radok,⁹ which states that the solution to every purely linear elastic problem, the Laplace transform of the solution to a corresponding linear viscoelastic problem under identical boundary conditions and zero initial conditions can be obtained by simply substituting the elastic constants in the purely elastic solution by the Laplace transform of the time differential operators in the viscoelastic problem. However, as discussed by Lee and Radok,¹⁰ the substitution procedure in Radok's correspondence principle does not always lead to valid operators in the stress relations, and for the case of indentation, these authors have shown that the correspondence principle gives valid viscoelastic solution only when the contact area does not decrease.¹⁰ Ting¹¹ reformulated the viscoelastic indentation problem and provided solutions for general load schedules including unload. Ting's solution agrees with that by Lee and Radok¹⁰ for the case of monotonically increasing contact area. In what follows, we reproduce the derivation of the viscoelastic indentation solution following the spirit of Lee and Radok's and Ting's work but cast it in a different form to suit our present purpose.

The stress-strain relationship expressed in terms of the deviatoric stress $S_{ij} = \sigma_{ij} - \delta_{ij}\sigma_{kk}/3$ and deviatoric strain $e_{ij} = \epsilon_{ij} - \delta_{ij}\epsilon_{kk}/3$ for a linear elastic material and the Maxwell viscoelastic material illustrated in Fig. 3 are, respectively,

$$e_{ij} = \frac{1}{2G} S_{ij} \quad , \quad \sigma_{ii} = 3B\epsilon_{ii} \quad \text{for elasticity, and}$$

$$\dot{e}_{ij} = \frac{1}{2G} \dot{S}_{ij} + \frac{3\dot{\epsilon}_o}{2\sigma_o} S_{ij} \quad , \quad \sigma_{ii} = 3B\epsilon_{ii} \quad \text{for viscoelasticity,}$$

or, in terms of Laplace transforms,

$$e_{ij}^* = \frac{1}{2G} S_{ij}^* \quad , \quad \sigma_{ii}^* = 3B\epsilon_{ii}^* \quad \text{for elasticity, and} \quad (4)$$

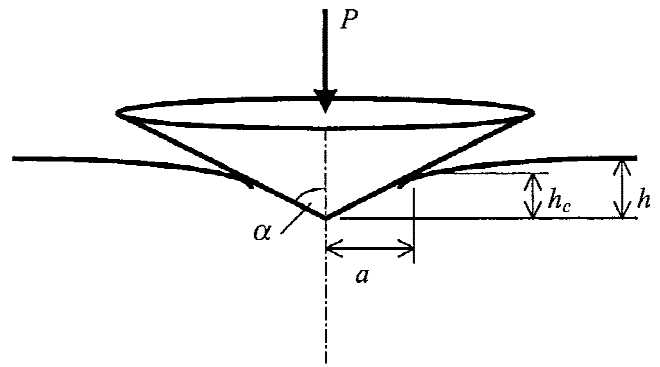


FIG. 4. Conical indentation into a half-space.

$$e_{ij}^* = \left(\frac{1}{2G} + \frac{3\dot{\epsilon}_o}{2\sigma_o} \cdot \frac{1}{s} \right) S_{ij}^* \quad , \quad \sigma_{ii}^* = 3B\epsilon_{ii}^* \quad \text{for viscoelasticity.} \quad (5)$$

Here G and B are the shear and bulk modulus respectively, $\dot{\epsilon}_o$ and σ_o are normalizing material constants, and $()^*$ is the Laplace transform of a time-dependent quantity $()$.

Thus, by Radok's correspondence principle, the elastic constants in the purely elastic problem are to be replaced by the following in the Laplace transform of the viscoelastic problem:

$$\frac{1}{G} \rightarrow \frac{1}{G} + \frac{3\dot{\epsilon}_o}{\sigma_o} \cdot \frac{1}{s} \quad ; \quad B \rightarrow B \quad . \quad (6)$$

The problem of indenting on a purely elastic solid has been solved by Sneddon.¹² In particular, we are interested in conical indentation as shown in Fig. 4, which shows a conical indenter with half-angle α being pressed into an elastic half space by a load $P(t)$ which varies with time t , making a depression with a contact circle of instantaneous radius a . The Sneddon solution yields

$$\frac{1}{E_r} P(t) = \frac{\pi}{2} \cot\alpha a^2(t) \quad . \quad (7)$$

To obtain the corresponding relation in the viscoelastic case, the reduced modulus E_r in Eq. (7) must be transformed. Since $E_r = 4G(3B+G)/(3B+4G)$, as a result of Eq. (6), E_r is to be transformed by

$$\frac{1}{E_r} \rightarrow \frac{1}{E_r} + \frac{3\dot{\epsilon}_o}{4\sigma_o} \cdot \frac{1}{s} + \frac{E^2\dot{\epsilon}_o}{36B^2\sigma_o} \cdot \frac{1}{(s + E\dot{\epsilon}_o/\sigma_o)} \quad , \quad (8)$$

where E is the Young's modulus. Therefore the Laplace transform of the viscoelastic version of Eq. (7) is

$$\left\{ \frac{1}{E_r} + \frac{3\dot{\epsilon}_o}{4\sigma_o} \cdot \frac{1}{s} + \frac{E^2\dot{\epsilon}_o}{36B^2(\sigma_o s + E\dot{\epsilon}_o)} \right\} P^* = \frac{\pi}{2} \cot\alpha (a^2)^* \quad . \quad (9)$$

It can be shown that Eq. (9) agrees with the formulation given at Ting¹¹ for the case of nondecreasing contact area. Inverse transforming Eq. (9) leads to

$$\begin{aligned} \frac{P(t)}{E_r} + \frac{3\dot{\epsilon}_0}{4\sigma_0} \int_0^t P(t') dt' \\ + \frac{E^2 \dot{\epsilon}_0}{36B^2 \sigma_0} \int_0^t \exp\left[-\frac{E\dot{\epsilon}_0}{\sigma_0} (t-t')\right] P(t') dt' \\ = \frac{\pi}{2} a^2 \cot \alpha \quad , \end{aligned}$$

or, upon differentiating with respect to t ,

$$\begin{aligned} \frac{\dot{P}(t)}{E_r} + \frac{3\dot{\epsilon}_0}{4\sigma_0} P(t) \\ + \frac{E^2 \dot{\epsilon}_0}{36B^2 \sigma_0} \int_0^t \exp\left[-\frac{E\dot{\epsilon}_0}{\sigma_0} (t-t')\right] \dot{P}(t') dt' \quad (10) \\ = 2a\dot{h} \quad , \end{aligned}$$

if $P(0) = 0$. In deriving Eq. (10), the purely elastic result¹² of $2a = \pi h \tan \alpha$ is used, and this is legitimate as this relation does not contain any elastic constant and so it applies equally well to the viscoelastic situation.

Next, we consider a general load schedule as shown in Fig. 5 consisting of a brief hold at a maximum load P_h immediately before unload at time t_h . We impose the condition that the load schedule up to the onset of unload does not decrease the contact size. The formal treatment by Ting¹¹ shows that the solution at maximum contact size is continuous. Thus the above results can be applied up to the onset of unload, and the length of the hold and the loading schedules, as well as the details of the

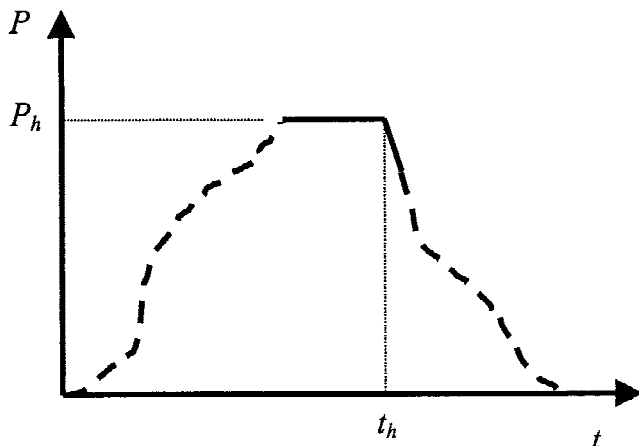


FIG. 5. General load schedule consisting of a load hold followed by unloading.

subsequent unload schedule do not matter. During the load hold, $\dot{P} = 0$, so that the creep rate towards the end of the load hold, i.e., at $t = t_h^-$, from Eq. (10) is

$$\begin{aligned} 2a(t_h^-) \dot{h}_h^c = \frac{3\dot{\epsilon}_0}{4\sigma_0} P(t_h^-) \\ + \frac{E^2 \dot{\epsilon}_0}{36B^2 \sigma_0} \int_0^{t_h^-} \exp\left[-\frac{E\dot{\epsilon}_0}{\sigma_0} (t_h^- - t')\right] \dot{P}(t') dt' \quad . \end{aligned} \quad (11)$$

At the onset of the unloading segment, $t = t_h^+$ with $\dot{P} < 0$, and Eq. (10) yields

$$\begin{aligned} 2a(t_h^+) \dot{h}_u = \frac{\dot{P}}{E_r} + \frac{3\dot{\epsilon}_0}{4\sigma_0} P(t_h^+) \\ + \frac{E^2 \dot{\epsilon}_0}{36B^2 \sigma_0} \int_0^{t_h^+} \exp\left[-\frac{E\dot{\epsilon}_0}{\sigma_0} (t_h^+ - t')\right] \dot{P}(t') dt' \quad . \end{aligned} \quad (12)$$

Since a and P are continuous at the cross-over point between the load hold and the unload, Eqs. (11) and (12) may be combined to give the jump in displacement rate across $t = t_h$ as

$$\dot{h}_u - \dot{h}_h^c = \frac{\dot{P}}{2aE_r} \quad . \quad (13)$$

By noting that \dot{h}_u/\dot{P} is the contact compliance $dh/dP|_u$ at the onset of unload, Eq. (13) may be rearranged to give

$$\left. \frac{dh}{dP} \right|_u = \frac{1}{2E_r a} + \frac{\dot{h}_h^c}{\dot{P}} \quad , \quad (14)$$

which is essentially Eq. (3).

In reality, thermal drift may also occur alongside creep and elastic recovery. Assuming a constant drift rate \dot{h}^t at around the cross-over point between the load hold and unload, the overall indenter displacement rate at the onset of unload from Eq. (13) should now be

$$\dot{h}_u = \frac{\dot{P}}{2E_r a} + \dot{h}_h^c + \dot{h}^t = \frac{\dot{P}}{2E_r a} + \dot{h}_h \quad ,$$

where $\dot{h}_h = \dot{h}_h^c + \dot{h}^t$ is the total displacement rate recorded at the end of the load hold. Hence, the apparent contact compliance at the onset of unload is

$$\left. \frac{dh}{dP} \right|_u = \frac{1}{2E_r a} + \frac{\dot{h}_h}{\dot{P}} \quad . \quad (15)$$

In other words, to correct for both creep and thermal drift effects in modulus measurement, one needs only to note the total displacement rate \dot{h}_h at the end of the load hold. A separate measurement of the thermal drift rate, however, is still required to calculate the contact area A_c

accurately from Eq. (2). From Eq. (15), a creep factor C to measure the importance of creep over elasticity can be defined as

$$C = \frac{\dot{h}_h^c S}{|\dot{P}|} \quad (16)$$

Evidently, a large C is caused by a slow unloading rate and/or a short load hold so that the creep displacement rate at the end of the load hold is large.

B. Experimental details

The experiments shown earlier in Figs. 1 and 2 and later in this section were carried out as follows. Three materials were used, namely, a single crystal of Ni₃Al, a single crystal of copper, and polycrystalline Al. And the composition of the Ni₃Al single crystal was 75 at.% Ni, 16.7 at.% Al, 8.0, at.% Cr, and 0.3 at.% B. Prior to nanoindentation, the crystal was homogenized at 1250 °C for 120 h. The copper single crystal was 99.99% pure and was annealed for 5 h at 800 °C. The polycrystalline Al was in the as-cast state with grain size approximately 1 to 2 μm. For both Ni₃Al and Cu, (111) surfaces for indentation were cut by a spark machine followed by grinding and electropolishing. The Al surfaces were also electropolished. Indentation experiments were carried out at room temperature using a Hysitron/Thermomicroscopes nanoindenter/atomic force microscope (AFM) setup with a Berkovich tip.

The load schedules employed were similar to the inset diagrams in Figs. 1 and 2(a). Typically, the load was ramped up quickly to the peak value and followed by rapid unload. This was to initiate the indent shape so that time-independent plasticity can be made negligible in subsequent stages. The load was then ramped up again to the peak value for a load hold, followed by the final unload from which the contact stiffness S for modulus calculation was to be measured. Low-load holds at 15% of the peak load were placed toward the end of the final unload and sometimes between the cycles to measure the thermal drift rate. This was to ensure that the thermal drift rate did not change much during the entire experiment. Results gathered under circumstances when the thermal drift rate changed a lot during the course of the experiment were discarded.

C. Results

To calculate the reduced modulus using Eqs. (1–3), the contact stiffness S_u at the onset of unload has to be measured from the unloading portion of the P - h curve. When the creep effect is small, the unloading curve can be fitted accurately with a power law according to Oliver and Pharr³

$$P = A(h - h_f)^m \quad (17a)$$

where A , h_f , and m are empirical parameters, and m is usually about 2. However, when creep is significant, a nose in the P - h curve is about to occur and in this case Eq. (17a) cannot give a good fit to the onset portion of the unload curve. Figure 6(a) shows an example of this in Cu. In this case, the creep factor defined in Eq. (16) is 60%, and so creep is significant. In situations like this, the following equation was found to give more accurate fit to the onset portion of the unload curve:

$$h = h_o + A_1 P^m + A_2 P^n \quad (17b)$$

where h_o , A_1 , A_2 , m , and n are fitting constants, and m is about 0.5. Figure 6(b) shows that Eq. (17b) gives a much better fit to the onset portion for the same data as in Fig. 6(a). Note that in this case, the value of S_u fitted by Eq. (17b) is 3454 μN/nm, but if Eq. (17a) is used, S_u will be underestimated to be only 706 μN/nm.

Figure 7 shows the modulus data measured from three metallic materials plotted against the creep factor defined in Eq. (16). In Eq. (16), \dot{h}_h^c is the creep part of \dot{h}_h and is obtained by subtracting the thermal drift rate from

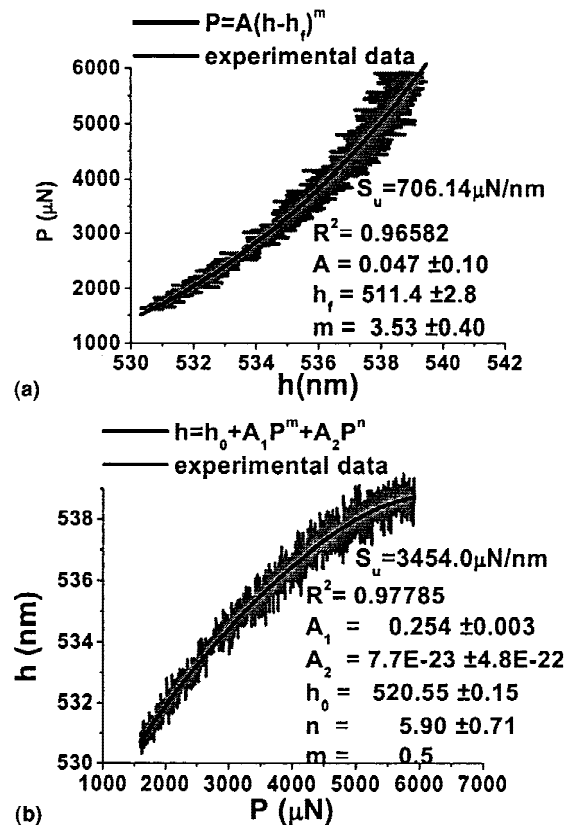


FIG. 6. Fitting of unloading curve for Cu with large creep effects using (a) Eq. (17a) and (b) Eq. (17b). In (a) the fit is poor near the maximum load. Unloading rate = 12.47 μN/s creep factor $C = 60.3\%$.

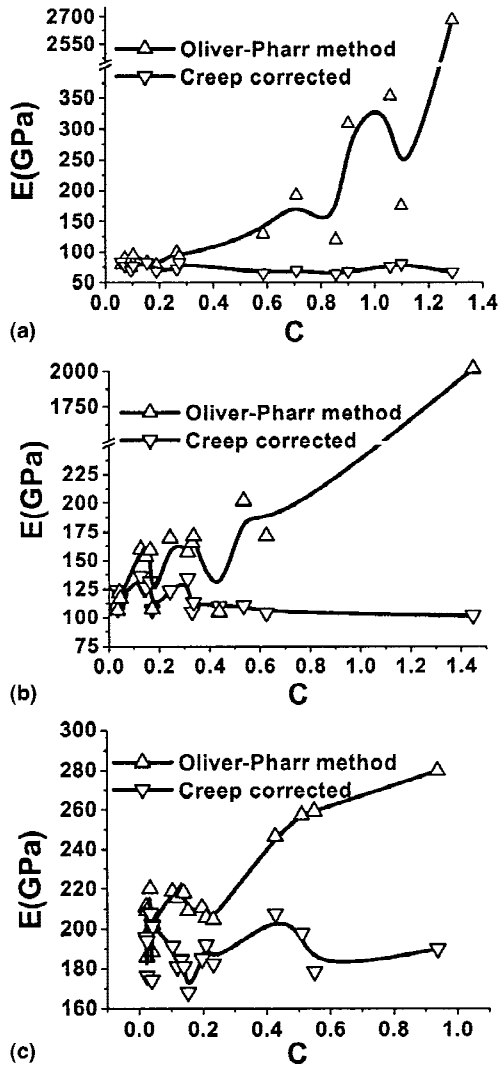


FIG. 7. The modulus-creep factor curve for (a) Al, (b) Cu(111), and (c) Ni₃Al(111).

the latter. The \dot{h}_h is obtained by fitting the load-hold displacement–time curve like those shown in Fig. 1 by the following empirical law:

$$h(t) = h_i + \beta(t - t_i)^{1/3} + K_t \quad ,$$

where h_i , β , t_i , and K are fitting constants. This relation is found to produce very good fits to most of our results. The modulus results shown in Fig. 7, whether corrected for creep or not, have all been corrected for thermal drift and machine compliance. It can be seen from Fig. 7 that for all three materials studied, the modulus calculated without considering creep increases as the creep factor C increases. The apparent modulus for Al and Cu is dramatically large (over 2000 GPa) when the creep factor is larger than about 1.4. In Fig. 7 are also shown the modulus values calculated from Eqs. (1) and (2) using the S values corrected by Eq. (3). It can be seen that the value of the calculated modulus tends towards a

constant value after the effect of creep is corrected for by Eq. (3). The corrected modulus is 72.3 ± 7.5 GPa for Al, 116.9 ± 11.1 GPa for Cu, and 189.9 ± 11.9 GPa for Ni₃Al. The theoretical values of the reduced modulus calculated using the method of Vlassak and Nix¹³ are 74.8 GPa for polycrystalline Al, 125.9 GPa for Cu(111) and 201.9 GPa for Ni₃Al(111), and these all fall within the range of the measured values.

III. NULL EFFECT OF THERMAL DRIFT

A. Theory

In this section, we will look at a condition by which the effects of thermal drift on modulus measurement will become negligible. Such a condition is obtained by analyzing the various errors leading to the calculation of modulus. First, let us assume for simplicity the indenter shape function in Eq. (2) to be parabolic, i.e., $A_c \propto h_c^2$. This assumption is valid when the indenter shape is pyramidal and when the penetration is deep enough that the blunted portion of the indenter tip has negligible effects. Under such a condition, from Eq. (3), we have

$$\frac{\Delta E_r}{E_r} = \frac{\Delta S}{S} - \frac{\Delta h_c}{h_c} \quad , \quad (18)$$

where $\Delta E_r/E_r$, $\Delta S/S$, and $\Delta h_c/h_c$ are the fractional errors of the corresponding quantities. Furthermore, from Eq. (3),

$$\frac{\Delta S}{S} = \frac{-\Delta(1/S)}{1/S} = S \left(\frac{\dot{h}_h}{|\dot{P}|} \right) = S \frac{(\dot{h}_h^c + \dot{h}_h^t)}{|\dot{P}|} \quad , \quad (19)$$

when machine errors are negligible. According to Eq. (2), if the thermal drift rate is assumed to be a constant \dot{h}^t during the entire course of the experiment, we have

$$\begin{aligned} \Delta h_c &= \dot{h}^t t_h + \epsilon P_{\max} \frac{\dot{h}_h^c + \dot{h}^t}{|\dot{P}|} \\ &= \dot{h}^t (t_h + \epsilon t_u) + \epsilon P_{\max} \frac{\dot{h}_h^c}{|\dot{P}|} \quad . \end{aligned} \quad (20)$$

Here, t_h is the duration of the test from the start to the onset of the unload (see Fig. 5), and t_u is the duration of the unload period alone. From Eqs. (18), (19), and (20), we obtain

$$\frac{\Delta E_r}{E_r} = \frac{\dot{h}_h^c}{|\dot{P}|} \left(S - \epsilon \frac{P_{\max}}{h_c} \right) + \frac{\dot{h}^t}{|\dot{P}|} \left(S - \frac{t_h + \epsilon t_u}{h_c} |\dot{P}| \right) \quad ,$$

but in general, t_h is much larger than t_{tr} , and, for ductile materials, S is much larger than $\epsilon P_{max}/h_c$, so that this can be simplified into

$$\begin{aligned} \frac{\Delta E_r}{E_r} &\approx \frac{\dot{h}_h^c}{|\dot{P}|} S + \frac{\dot{h}^t}{|\dot{P}|} \left(S - \frac{t_h}{h_c} |\dot{P}| \right) \\ &= C + \frac{\dot{h}^t S}{|\dot{P}|} \left(1 - \frac{t_h |\dot{P}|}{h_c S} \right), \end{aligned} \quad (21)$$

where C is the creep factor defined in Eq. (16). Therefore, to minimize the error of modulus due to creep and thermal drift, we can use the following two conditions:

$$C < 10\%, \text{ or } |\dot{P}| > 10\dot{h}_h^c S, \quad \text{and} \quad (22)$$

$$t_h \approx \frac{S}{|\dot{P}|} h_c. \quad (23)$$

The first condition serves to keep the creep effect small, while the second condition minimizes the thermal drift effect.

B. Comparison with experiments

To illustrate the validity of Eq. (23), experiments on Al and Ni₃Al(111) were performed using the load schedules shown in Fig. 8. The aim of the experiments was to investigate the influence of thermal drift on the measured modulus in the light of Eq. (21). Therefore, it is necessary to vary the factor $t_h |\dot{P}|/h_c S$ while, if possible, all other parameters are kept constant. It is therefore desirable to vary $|\dot{P}|$ on the same indent, and so the load schedules of Fig. 8 were used, which consisted of firstly a cycle of rapid load ramp and unload to initiate the indent plasticity, and then a series of ramp-hold-unload cycles with different unloading rates. As before, low-load holds at 15% of the peak load were placed in between cycles and at the end to monitor the actual thermal drift in the course of the experiment. Before the experiment, the machine was allowed to stabilize in a closed room without ventilation for several hours and in this way, the actual thermal drift rate was found to be quite steady during the subsequent experiment. The thermal drift (= drift rate × time) was subtracted from the recorded displacement data to obtain a load–displacement curve free from thermal drift effects. This is termed the “standard” load–displacement curve for the purpose of later reference. Then artificial drift rates were superimposed onto the displacement data in the standard curve to stimulate the effects of repeating the experiment at different thermal drift rates. The reason for doing this instead of actually repeating the experiment at different drift rates is that, with our setup, it is very difficult to control the drift rate at an arbitrary constant value during the entire duration of the experiment.

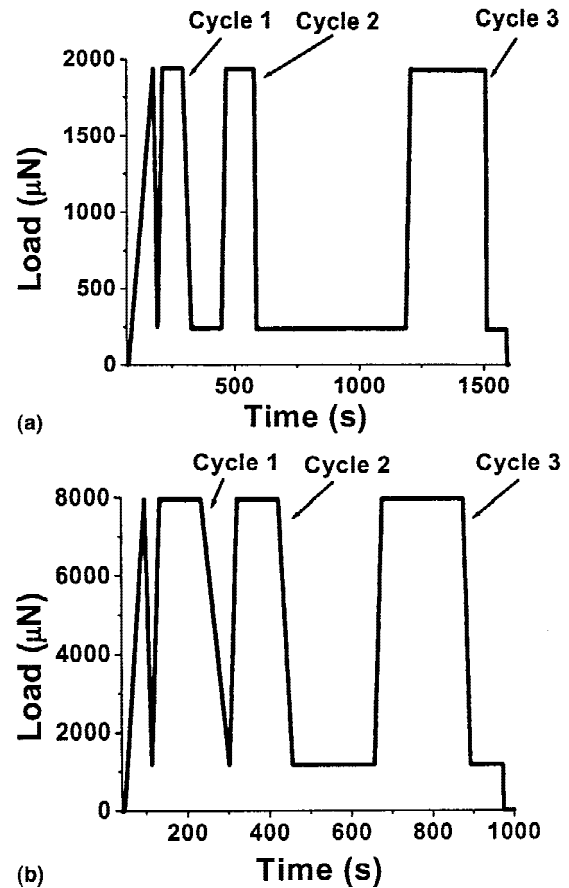


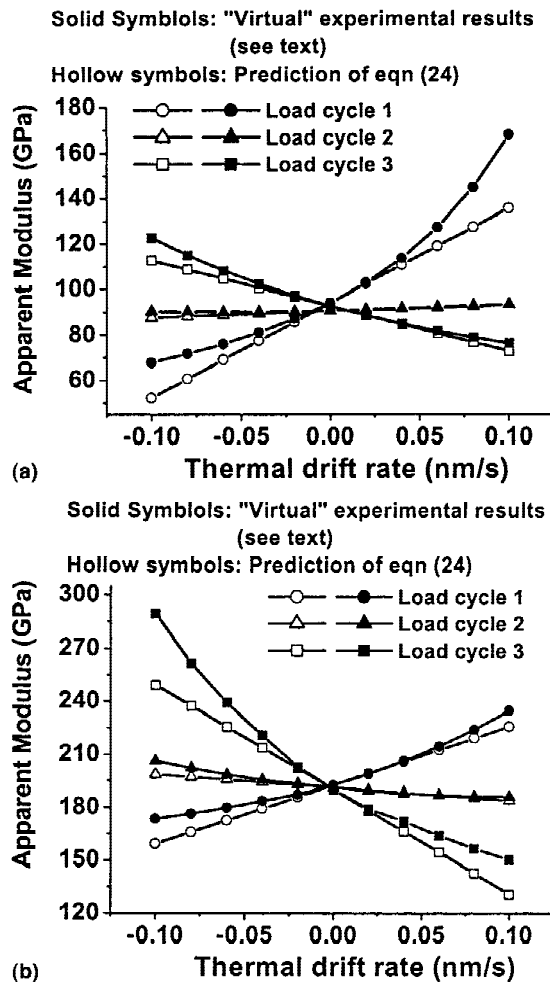
FIG. 8. Load schedule for the investigation of thermal drift effects in Ni₃Al(111).

From the results obtained, the C factors calculated for the six labeled cycles in Fig. 8 were reasonably small, implying that creep is insignificant. The key experimental data for these cycles are listed in Table I. The moduli calculated from the standard load–displacement curves using the Oliver–Pharr scheme are shown as $E_r(\dot{h}^t = 0)$ in Fig. 9. These are reasonably consistent with one another and with the theoretical values calculated using the method by Vlassak and Nix.¹³ Also shown in Fig. 9 are the apparent moduli calculated using the Oliver–Pharr scheme after thermal drifts corresponding to an artificial \dot{h}^t were superimposed onto the standard load–displacement curves for different cycles. As discussed above, these represent “virtual” experimental results at various values of \dot{h}^t . The apparent modulus that would have been measured under a general thermal drift rate of \dot{h}^t and with negligible creep may also be calculated from Eq. (21) as

$$\begin{aligned} E_r(\dot{h}^t) &= E_r(\dot{h}^t = 0) \\ &\times \left[1 + \frac{\dot{h}^t}{|\dot{P}|} \left(S - \frac{t_h}{h_c} |\dot{P}| \right) \right], \end{aligned} \quad (24)$$

Table I. Experimental data for the load cycles in Fig. 8.

	C (%)	t_h (s)	$ \dot{P} $ ($\mu\text{N/s}$)	S ($\mu\text{N/nm}$)	h_c (nm)	$1 - \frac{t_h \dot{P} }{S h_c}$
Al cycle 1	22.14	293	50	267.1	463.4	0.881639
Al cycle 2	5.55	577	170	261.2	491.5	0.235939
Al cycle 3	1.92	1,507	400	286.1	540.1	-2.90105
Ni ₃ Al cycle 1	8.72	233.5	100	283.2	239.8	0.656169
Ni ₃ Al cycle 2	3.05	421.5	200	278.8	245.3	-0.23264
Ni ₃ Al cycle 3	1.98	875.5	400	271.6	241.1	-4.34797

FIG. 9. Apparent reduced modulus versus thermal drift rate for (a) Al and (b) Ni₃Al(111).

where $E_r(\dot{h}^t = 0)$ is the modulus calculated from the standard load–displacement curve using the Oliver–Pharr scheme. The variations of the apparent modulus with \dot{h}^t obtained this way for different cycles are also shown in Fig. 9. The abscissa scale selected in Fig. 9 is ± 0.1 nm/s and this well covers the usual thermal drifts encountered in our machine. It can be seen that the prediction of Eq. (24) agrees reasonably well with the “virtual” experimental results. More importantly, the apparent modulus for cycle 2 for both Al and Ni₃Al stays more or less constant

at the value $E_r(\dot{h}^t = 0)$ when \dot{h}^t varies within ± 0.1 nm/s, while that for cycles 1 and 3 changes remarkably with \dot{h}^t . This implies that the calculated modulus for cycle 2 would be least sensitive to thermal drift. The reason is that, as shown in the table, the t_h and \dot{P} selected for cycle 2 are such that the factor $(1 - t_h |\dot{P}| / h_c S)$ is closest to zero amongst the three cycles for both Al and Ni₃Al. This helps to verify the condition for minimal effect of thermal drift as given in Eq. (23).

IV. DISCUSSION

It is interesting to note from Fig. 7 that indentation creep effect is significant at room temperature even for a moderately high melting metal like Cu or Ni₃Al. We attribute this to the high local values of the hydrostatic stress, which makes dislocation climb very easy. After the indenter is rapidly pressed into the specimen, the instant plasticity occurs in such a way that the shear stresses due to the applied load are more or less balanced by the intrinsic Peierls stress plus resistance due to work hardening. The hydrostatic component of the stress state, however, may exceed well beyond the elastic limit, although it attenuates in the far field. The very first layer of material in contact with the indenter may be subject to hydrostatic stresses on the order of the measured hardness, and this is typically a few times the yield stress. Under such high hydrostatic stresses, dislocation climb may take place very effectively, even though the test temperature is a relatively small fraction of the melting temperature. As an illustration, the measured hardness of Cu in the nanoindentation range may exceed 1 GPa. When multiplied to the atomic volume, such a high value of hydrostatic stress will reduce the activation energy for dislocation climb by a fraction of an eV, which is similar in magnitude to the activation energy for pipe diffusion in Cu. The result is that the local work-hardening dislocation structure may relax at a significant speed with time, causing the hardness to fall.

In deriving Eq. (14), the assumption of linear viscoelasticity was used. The question therefore is whether the result will be modified a lot if the viscoelasticity is power-law instead of linear. This is difficult to estimate without a full solution to the power-law problem, which can only be obtained by numerical means. However, the

main idea behind Eq. (14) is that the influence of creep on the contact stiffness S increases with \dot{h}_h^c and decreases with $|\dot{P}|$. This is in agreement with the experimental observation in Fig. 2. Equation (3) or (15) in fact indicates that the overall contact stiffness during unload is larger than what elasticity alone would produce. Hence, if the creep contribution is ignored, the modulus calculated from Eq. (1) would be an overestimate, and indeed, if the creep factor C in Eq. (16) is not small compared with unity, the overestimation could be very severe. The results in Fig. 7 confirm this. When C is large enough, the apparent contact stiffness will become negative, i.e., a “nose” will occur in the unloading segment in the load-displacement curve. The power-law fit of the P - h curve in the Oliver–Pharr scheme obviously cannot conform to a negative slope at the onset of the unload, but this has a fortuitous “correction” effect on the measured modulus since the erroneous negative contact stiffness is now replaced by a positive value. This correction, however, is not very effective at large values of C as Fig. 7 shows.

We established in Eq. (23) a condition such that, if satisfied, the measured modulus would be relatively free from thermal drift effects, and hence correction of thermal drifts on the displacement data would become unnecessary. This condition works on the premise that the modulus is proportional to S and inversely proportional to h_c , and that the effects of thermal drift on S and h_c are of the same sign as shown in Eqs. (19) and (20). The significance, however, is that if Eq. (23) is satisfied, the influence of thermal drift on S and h_c would cancel one another irrespective of the magnitude of the drift rate as long as it is small. Thus the effect of thermal drift on modulus measurement can be nullified without knowledge of the drift rate or the need to do any correction. Therefore, if it is certain that creep effects can be ignored (e.g., when a very rapid unloading rate is used), then there is no need to record the drift rate if the hold time is selected according to Eq. (23). If the effect of creep is uncertain, however, then it would be necessary to measure the thermal drift rate. This is because to check whether Eq. (22) is satisfied or to calculate C if it is not satisfied, one would still need to know \dot{h}_h^c , which can only be obtained by subtracting the thermal drift component from the overall observable \dot{h}_h .

In terms of practicalities, Eq. (23) suggests that for a fixed material, a short t_h will have to be used if a large $|\dot{P}|$ is used, but if a short t_h is used, \dot{h}_h^c may be too

large to fulfill Eq. (22). There may, therefore, be a compatibility problem between the two conditions. Fortunately, the t_h in Eq. (23) is also proportional to the contact depth, which can be increased by using a larger peak load. In any case, this potential problem of incompatibility did not arise in our experiments on Al and Ni₃Al. Finally, to use Eq. (23) to select t_h , one would need to know the contact depth and the unloading contact stiffness beforehand. A trial run is therefore required to estimate the values of these two parameters.

V. CONCLUSIONS

We have shown that, in a linear Maxwell solid, the effect of creep on the apparent contact compliance at the onset of unload can be worked out in a simple way from the indenter displacement rate during the end of the load hold and the unloading speed. Incorporating this correction term into the elastic modulus calculation greatly improves the accuracy of the estimated modulus, especially when the creep effects are large. A condition for null effect of thermal drift has also been identified so that, if it is satisfied, thermal drift, irrespective of its magnitude, has minimal effects on the measured modulus.

REFERENCES

1. J.B. Pethica, R. Hutchings, and W.C. Oliver, *Philos. Mag. A* **48**, 593 (1983).
2. M.F. Doerner and W.D. Nix, *J. Mater. Res.* **1**, 601 (1986).
3. G.M. Pharr, W.C. Oliver, and F.R. Brotzen, *J. Mater. Res.* **7**, 613 (1992).
4. W.C. Oliver and G.M. Pharr, *J. Mater. Res.* **7**, 1564 (1992).
5. R.Y. Lo and D.B. Bogy, *J. Mater. Res.* **14**, 2276 (1999).
6. J.C. Hay, A. Bolshakov, and G.M. Pharr, *J. Mater. Res.* **14**, 2296 (1999).
7. R. Hill, *Proc. R. Soc. Lond. A* **436**, 617 (1992).
8. A.F. Bower, N.A. Fleck, A. Needleman, and N. Ogbonna, *Proc. R. Soc. Lond. A* **441**, 97 (1993).
9. J.R.M. Radok, *Q. App. Math.* **15**, 198 (1957).
10. E.H. Lee and J.R.M. Radok, *J. App. Mech.* **27**, 438 (1960).
11. T.C.T. Ting, *J. App. Mech.* **33**, 845 (1966).
12. I.N. Sneddon, *Int. J. Engng. Sci.* **3**, 47 (1965).
13. J.J. Vlassak and W.D. Nix, *J. Mech. Phys. Solids.* **42**, 1223 (1994).