Available online at www.sciencedirect.com
ScienceDirect
Wear 265 (2008) 729–734
WEAR
www.elsevier.com/locate/wear

Inclusion of "interaction" in the Greenwood and Williamson contact theory

M. Ciavarella [a,*], J.A. Greenwood [b], M. Paggi [c]

[a] *CEMEC-PoliBA – Centre of Excellence in Computational Mechanics, Politecnico di Bari, V.le Japigia 182, 70125 Bari, Italy*
[b] *Engineering Department, Cambridge University, Trumpington Street, Cambridge CB2 1PZ, UK*
[c] *Department of Structural and Geotechnical Engineering, Politecnico di Torino, C.so Duca degli Abruzzi 24, 10129 Torino, Italy*

Received 9 October 2006; received in revised form 17 December 2007; accepted 9 January 2008
Available online 19 March 2008

Abstract

Recent direct implementation of asperity theories is reinterpreted here to formulate an improved version of the Greenwood and Williamson (GW) theory with the inclusion of interaction between asperities. This is achieved by treating the contact pressures as uniformly distributed over the apparent contact area and the resulting deformation as uniform. The correction is equivalent to an increase of the effective separation of the mean planes by a quantity proportional to the nominal pressure, resulting in a reduction of the "real" area of contact and of total load for a given separation. However, the area–load relationship is unchanged. The correction effectively depends on the ratio between the nominal pressure and the elastic modulus multiplied by the ratio between the size of the nominal contact area and standard deviation of the asperity heights. For contacts much larger than the size of roughness, uniform interaction effects would be dominant at relatively modest pressures (particularly for soft materials). This also means that the effect of interaction is unlimited. However, the only significant change is in the prediction of gas-tightness, it is harder to seal a large area than a small one. The modification of the theory has a significant effect on stiffness and conductance. Indeed, a parallel is drawn between this correction and the "clustering" terms of resistance in the Holm–Greenwood formulae for a cluster of circular spots. Finally, numerical contact simulations using Weierstrass–Mandelbrot (WM) surfaces show a general agreement with the improved theory but also significant scatter for low load levels. Taking into account the effect of asperity interaction, the improved GW theory is now able to predict the numerically obtained contact response for intermediate load levels.

Keywords: Greenwood–Williamson theory; Contact mechanics; Roughness; Fractals; Contact conductance

1. Introduction

The Greenwood and Williamson theory [1] of contact between rough planes (we denote it as GW in the sequel), is a standard theory for predicting the load–displacement behaviour of rough surfaces in contact given the height distribution of the contacting asperities, and is one classical explanation of many linear or nearly linear laws in tribology. The theory is even used as an inverse technique to determine the asperity distribution directly from topographic measurement and subsequent data reduction [2], for example to follow how the asperity height distribution can evolve with loading. In the GW theory, each asperity contact is assumed to behave independently. Recently, Ciavarella et al. [3] compared a direct calculation of the contact of a simulated rough surface against a plane (based on a "discrete" interpretation of the asperity model, using first-order interaction terms) with the matching GW calculation, and found significant differences between the results at intermediate loads, clearly due to the inclusion of interaction between asperities in the direct calculation. Specifically, the direct calculation assumed, as in the GW theory, that a force P_1 on an asperity lowers the height of that asperity by the Hertzian deformation $(((3/4)P_1)/E^*R^{1/2})^{2/3}$, where E^* and R denote the composite Young's modulus and the asperity radius of curvature, respectively, but also that it reduces the height of any other asperity by an amount $P_1/\pi E^* r$, where r is the distance between

* Corresponding author. Tel.: +39 080 5962811; fax: +39 080 5962777.
E-mail address: mciava@poliba.it (M. Ciavarella).

0043-1648/$ – see front matter © 2008 Published by Elsevier B.V.
doi:10.1016/j.wear.2008.01.019

the two. As a result, their load–separation and area–separation curves lay well below the corresponding GW curves at intermediate loads, although their area–load relationship agreed reasonably well with GW predictions. The following simplified theory reproduces this behaviour, but also raises rather deeper questions.

2. The GW theory and improved version

In the GW theory [1], when the separation between the mean plane of the rough surface and the contacting rigid plane is d, an asperity of height z will be compressed by $(z-d)$, and a force $P_1 = (4/3)E^* R^{1/2}(z-d)^{3/2}$ will be produced. Adding these contact forces together by statistical averaging as in the GW theory, or by direct addition in a numerical simulation, gives the total force acting on the surfaces. Dividing this total force by the nominal area of contact, A, then gives the nominal pressure, p_{nom}.

At light loads, when there are relatively few asperity contacts, the inclusion of interaction between asperities has very little effect, and may be ignored. For heavier loads, we assume that asperity contacts can be treated as uniformly distributed over the apparent contact area and that the resulting deformation can be treated as uniform. According to Timoshenko and Goodier [4], the average deformation over a compact area A due to a uniform pressure p_{nom} acting over that area is $mp_{\text{nom}}\sqrt{A}/E^*$, where the factor m equals 0.96 for a circular area, 0.95 for a square, and decreases gently for rectangles of increasing aspect ratio. Note, incidentally, that, for a Hertzian pressure distribution, the factor is $m = (9/16)\sqrt{\pi} = 0.997$. Here, it will be accurate enough to take $m = 1$. Thus, the height of each asperity is in effect reduced from z to $(z - (p_{\text{nom}}\sqrt{A}/E^*))$, and the corresponding force becomes $P_1 = (4/3)E^* R^{1/2}(z - p_{\text{nom}}\sqrt{A}/E^* - d)^{3/2}$.

Greenwood and Williamson [1] introduced the functions $F_n(h) \equiv (1/\sqrt{2\pi})\int_h^\infty (s-h)^n \exp(-0.5s^2)\mathrm{d}s$, where $n = 0$, 1/2, 1, 3/2. Then for a Gaussian distribution of asperity heights, the load when the mean planes are separated by d becomes $W = (4/3)NE^* R^{1/2}\sigma^{3/2}F_{3/2}(d/\sigma)$, where N is the number of asperities – not the number of contacts, which is $n = NF_0(d/\sigma)$ – and σ is the standard deviation of the asperity heights. If the asperity density is $\eta = N/A$, then the nominal pressure is

$$p_{\text{nom}} = \frac{4}{3}\eta E^* R^{1/2}\sigma^{3/2}F_{3/2}\left(\frac{d}{\sigma}\right). \tag{1}$$

Accordingly, the effect of asperity interaction is to modify (1) to

$$p_{\text{nom}} = \frac{4}{3}\eta E^* R^{1/2}\sigma^{3/2}F_{3/2}\left(\frac{d}{\sigma} + \frac{p_{\text{nom}}\sqrt{A}}{E^*\sigma}\right) \tag{2}$$

Introducing the non-dimensional nominal pressure

$$\hat{p}_{\text{nom}} = \frac{p_{\text{nom}}}{\eta E^* R^{1/2}\sigma^{3/2}}, \tag{3}$$

this becomes

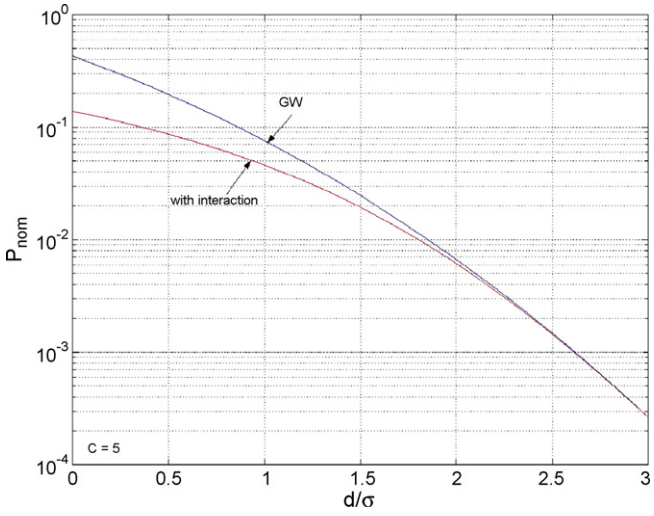

Fig. 1. The load vs. separation curve with and without interaction.

$$\hat{p}_{\text{nom}} = \frac{4}{3}F_{3/2}\left(\frac{d}{\sigma} + \hat{p}_{\text{nom}}\eta\sqrt{AR\sigma}\right) \equiv \frac{4}{3}F_{3/2}\left(\frac{d}{\sigma} + C\hat{p}_{\text{nom}}\right) \tag{4}$$

where the parameter $C = \eta\sqrt{AR\sigma} = \sqrt{N}\sqrt{\eta R\sigma}$ is independent of load and separation.

In this form, the solution must be found iteratively.[1] It is more convenient, however, to regard the effect of asperity interaction as an increase in the effective separation of the mean planes to $d_1 = d + p_{\text{nom}}\sqrt{A}/E^*$. The load and hence the nominal pressure (and area, conductance and number of contacts) are found as functions of d_1: only then the true mean plane separation is calculated from $d/\sigma = d_1/\sigma - C\hat{p}_{\text{nom}}$ and the true load–separation curve plotted. Many researchers have found that values of the product $\eta R\sigma$ are generally close to 0.05; so the constant C can be taken as $\sqrt{N/20}$. Fig. 1 shows the result of including asperity interaction in this way, assuming a compact area of contact containing 500 asperities ($C = 5$).

However, it will be clear that the effect of interaction in this model is when the variation with respect to separation is considered. A study of real contact area against load will produce identical curves whether the individual quantities are found as functions of d or d_1. The only practically significant change will be when the effectiveness as a seal for gas-tightness matters. Here it is clear that it is the effective separation d_1 which is a measure of the leakage, while the nominal separation d measures the load. In short, the effect of interaction is to make sealing very much more difficult to accomplish.

The above theory covers only the way in which asperities fail to behave independently. A full analysis of contact must also take into account the fact that individual contact areas can merge as the load increases, so that the number of contact areas

[1] However, a very good approximation could be found by the first iteration only, i.e. using the original GW theory to estimate the mean pressure.

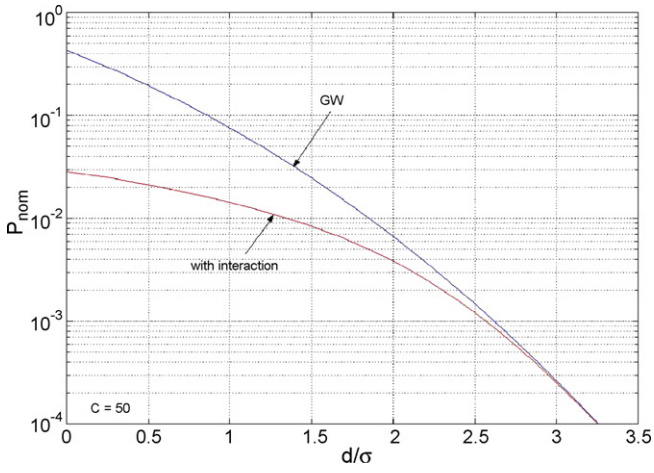

Fig. 2. Same as Fig. 1 but with 50,000 asperities rather than 500.

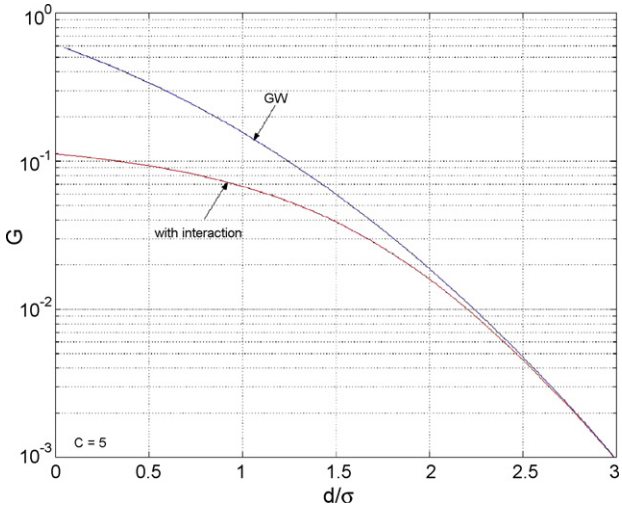

Fig. 3. The GW stiffness $\partial W/\partial d_1$ and the correct value allowing for the effect of interaction, $\partial W/\partial d$.

may rise to a maximum but may then fall. Certainly this occurs when contacts deform plastically, where the ultimate state may be a single contact area. The elastic behaviour is still an open question.

A worrying implication of the above theory is what happens when the nominal area of contact becomes large, and the number of asperities correspondingly large. Suppose there were 50,000 asperities rather than 500? The inclusion of interaction is now no longer a small correction but makes an enormous difference (see Fig. 2).

Certainly this makes clear that there can be no universal graph showing the effect of interaction: the effect is unlimited. Fortunately, as we have seen, the only significant change is in the prediction of gas-tightness, and here the result, that it is harder to seal a large area than a small one, is hardly revolutionary!

Ciavarella et al. [3,5,6] also calculated the conductance (per unit area) between the surfaces, by using the theorem due to Barber [7], that the conductance may be found from the stiffness by replacing the contact modulus, E^*, by the electrical or thermal conductance, ρ^{-1} or K. We note that GW did not believe this to be the actual electrical conductance, since it ignored the effect of oxide films, preferring to believe that only plastically deforming contacts would conduct. It is clear that it would not give the thermal conductance either, neglecting oxide films but also the contribution of thermal radiation. The GW stiffness is $\partial W/\partial d_1$, while the correct value accounting for interaction will be $\partial W/\partial d$. The two predictions are shown in Fig. 3, and it is seen that the effect of differentiating is to increase considerably the importance of interaction.

3. Further results on conductance

Holm [8] and Greenwood [9] developed analytical expressions for the constriction resistance (inverse of conductance) due to a cluster of nearly uniformly distributed circular contact areas of mean radius $\bar{a}$. The specific constriction resistance, R/ρ, is given by the parallel resistance of the spots, incremented by an interaction term, which is equivalent to the resistance of a single spot of radius α of the cluster, the Holm radius α:

$$\frac{R}{\rho} = \frac{1}{2n\bar{a}} + \frac{1}{2\alpha} \tag{5}$$

Naturally, in a real contact, both the size and the distribution of the spots vary with load. Also, the spot sizes are not constant nor spacing is uniform. A more general and correct solution is then Greenwood's original formula

$$\frac{R}{\rho} = \frac{1}{2\sum a} + \frac{1}{\pi}\left(\frac{\sum\sum_{i,j,i\neq j}(a_i a_j)/s_{ij}}{\left(\sum a_i\right)^2}\right) \tag{6}$$

When using the statistical GW theory, only an estimate can be made on the average radius, and hence we find no advantage in using this formula. On the other hand, from a numerical point of view, already with 500 asperities we found prohibitive computational costs in using Eq. (6) with respect to a more trivial differentiation of the load–displacement relationship.

It is interesting to show that the modification of the original GW theory, in terms of conductance, corresponds to the introduction of the interaction term in the Holm–Greenwood formulae. This is immediate in the case of an exponential distribution of asperity heights $(1/\sigma)\exp(-z/\sigma)$ (where σ is the analogue of the RMS height in the exponential case), in which Eq. (2) becomes

$$p_{\text{nom}} = \eta\pi^{1/2}E^*R^{1/2}\sigma^{3/2}\exp\left[-\frac{d}{\sigma} + \frac{p_{\text{nom}}\sqrt{A}}{E^*\sigma}\right], \tag{7}$$

Then

$$\ln(p_{\text{nom}}) + \frac{p_{\text{nom}}\sqrt{A}}{E^*\sigma} = \ln(\eta\pi^{1/2}E^*R^{1/2}\sigma^{3/2}) - \frac{d}{\sigma} \tag{8}$$

Recalling that the stiffness is $S \equiv \partial(Ap_{\text{nom}})/\partial d$, then the compliance $1/S$ can be derived by determining d from Eq. (8) and

differentiating it with respect to the load:

$$\frac{1}{S} = \frac{\sigma}{p_{\text{nom}} A} + \frac{1}{E^* \sqrt{A}} \tag{9}$$

According to Barber's theorem [7], we obtain the conductance from the stiffness by replacing E^* by the conductivity K. Here, noting that the first term does not contain E^*, we use the theorem in the form 'multiply $(1/S)$ by $E^*/K \equiv E^* \rho$ to get the constriction resistance':

$$\frac{R}{\rho} = \frac{E^* \sigma}{p_{\text{nom}} A} + \frac{1}{\sqrt{A}} \tag{10}$$

The first term here is exactly the first term of the Greenwood–Holm formula, while the second term, $\left(1/\sqrt{A}\right)$, closely resembles the 'Holm radius' term $1/2\alpha$ in Eq. (5). Hence, Eq. (5) is obtained rather closely by this completely different route.

4. Comparison with numerical simulations

In order to test the results of the "improved" GW model and to show further aspects of the interaction concerning scatter, realistic 3D fractal surfaces are generated using a modified two-variable Weierstrass–Mandelbrot (WM) function [10,11], which can be written as

$$z(x, y) = B \sum_{m=1}^{M} \sum_{n=0}^{n_{\max}} \gamma^{(D-3)n} \left\{ \cos \phi_{m,n} - \cos \left[\frac{2\pi \gamma^n (x^2+y^2)^{1/2}}{L} \times \cos \left(\tan^{-1} \frac{y}{x} - \frac{\pi m}{M} \right) + \phi_{m,n} \right] \right\} \tag{11}$$

where

$$B = L \left(\frac{G}{L} \right)^{D-2} \left(\frac{\ln \gamma}{M} \right)^{1/2} \tag{12}$$

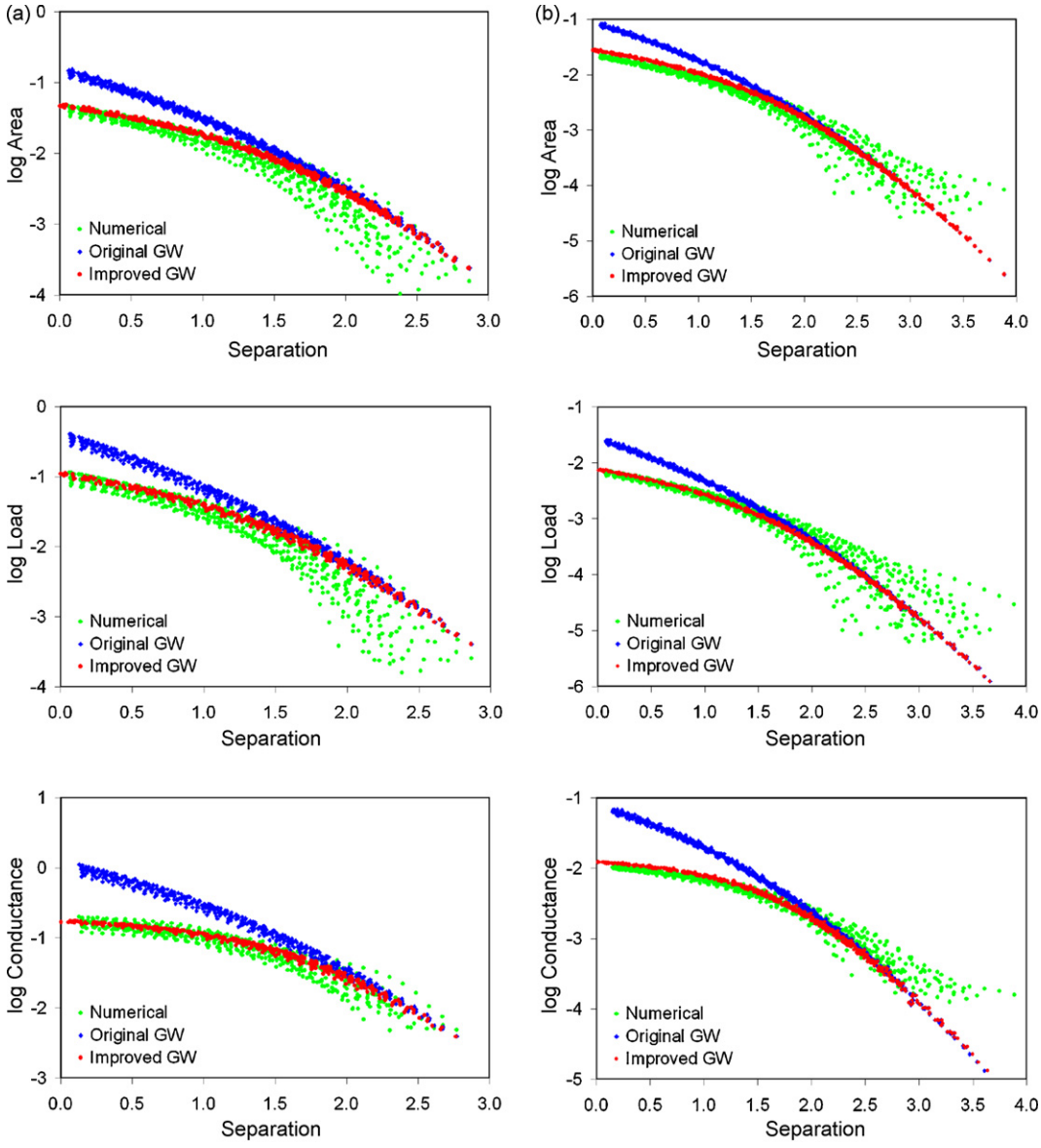

Fig. 4. Area, load and conductance as functions of the separation for WM surfaces with $\gamma = 1.5$, $n_{\max} = 8$ and $M = 10$. (a) $D = 2.05$ and (b) $D = 2.95$.

and L is the sample length, G some measure of amplitude roughness, D the fractal dimension ($2 < D < 3$), γ a scaling parameter, M the number of superposed ridges used to construct the surface profile, n is a frequency index (with n_{max} representing the upper limit of n), and ϕ is the random phase among the sinusoids. The parameter γ controls the density of frequencies in the surface. In the numerical examples we shall use $\gamma = 1.5$, $n_{max} = 8$, $M = 10$ and $D = 2.05$, or 2.95, these being limit values close to the limits 2 and 3.

For the direct simulation, we shall use the "discrete" numerical implementation of the GW theory [3], also taking into account of interaction effects, at least to the first-order, as explained in the following. In fact, we start by detecting asperities in the GW sense, so that the entire surface is described by few asperities which in turn can be described by their location height and radius of curvature. We use then the Hertz theory for each individual asperity, adding the elastic displacements caused by the other asperities to their self-displacements. We simplify non-circular asperities according to the Greenwood suggestion [12] of using the geometric mean of the principal radii instead of the arithmetic mean. A non-linear system of equations is put together by writing the condition of displacement compatibility at each asperity centre. Hence, starting from half of the "bearing area" overlap contact radius between each asperity and the undeformed half-plane, for each iteration and for each asperity a correction in the contact radius is made, proportional to the difference between the height of the summit and the corresponding elastic displacement of the half-plane. No matrix inversion is required in the iterative scheme. Once the distribution of contact radii has been determined, the discrete distribution of pressure can be obtained. This numerical approach is very efficient and, from the computational point of view, is less expensive than FEM simulations [13] or numerical methods based on the active set strategy [14,15], since only a limited number of variables for each asperity in contact is required.

As a summary, in this section we present a comparison among the following predictions:

(1) "*Numerical*", for the results obtained using the algorithm so far described as the theory, keeping in mind that this includes interaction effects to the first-order. Here, the scatter in the results is due to the actual discrete distribution of asperities.
(2) "*Original GW*", as the original GW theory, i.e. each surface asperity set is fitted to a Gaussian distribution of heights, and we take the mean values for the asperity radius as needed in the GW theory—here the scatter in the results is only due to the fact that RMS asperity heights and mean radius varies from one realization to another.
(3) "*Improved GW*", as the improved GW theory proposed in this paper.

We repeat the simulations for 25 realizations of nominally identical surfaces, and this permits also to assess the repeatability and the expected scatter in the results. The contact results will be presented in non-dimensional form. Hence, the results in the graphs can be read as *Area*, *Load*, *Separation* and *Conductance*, defined as follows:

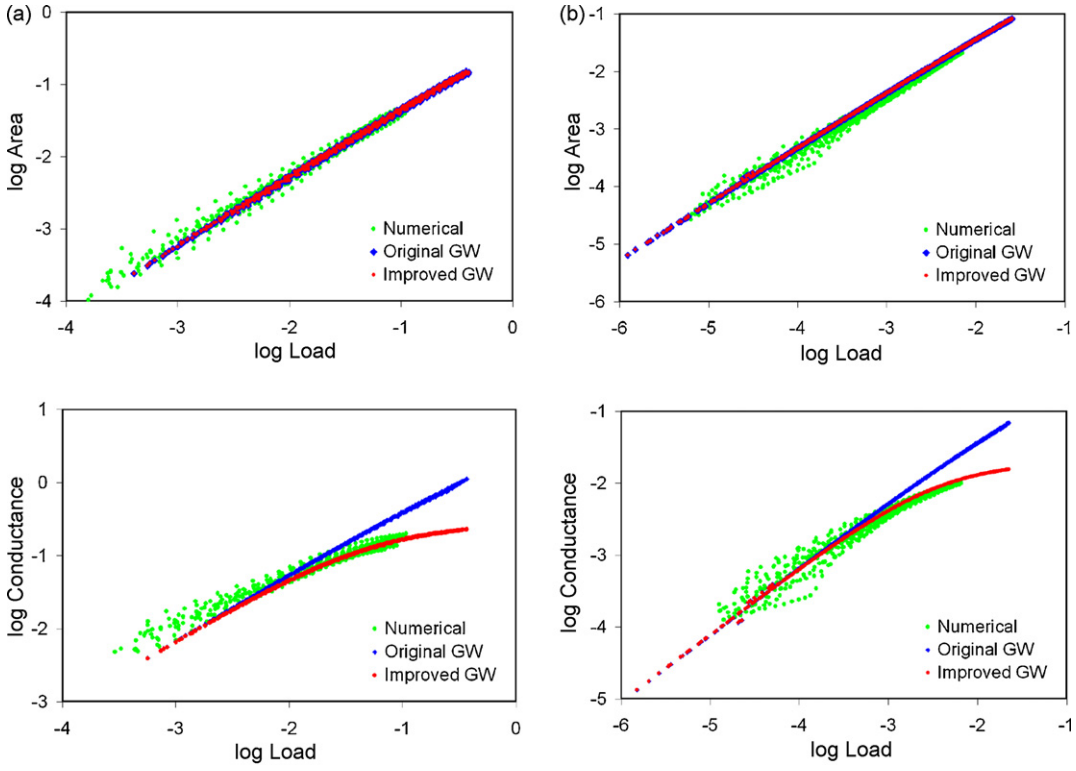

Fig. 5. Area–load and conductance–load relationships for WM surfaces with $\gamma = 1.5$, $n_{max} = 8$ and $M = 10$. (a) $D = 2.05$ and (b) $D = 2.95$.

(1) The *Area* is the ratio between the real contact area and the nominal one, i.e. $Area = A/L^2$, where L denotes the lateral size of the sample.
(2) The *Load* corresponds to the applied normal force divided by the nominal contact area and the composite Young's modulus, i.e. $Load = W/(E^* L^2)$.
(3) The *Separation* is computed as the ratio between the mean plane separation, d, and the standard deviation of the asperity heights, σ, i.e. $Separation = d/\sigma$.
(4) The *Conductance* is defined as: $Conductance = C\rho\sigma/L^2 = -2(\partial Load/\partial Separation)$.

The comparisons are shown in Fig. 4 for $D = 2.05$ (Fig. 4a) and $D = 2.95$ (Fig. 4b), where it is evident that the improved theory correctly captures the deviation from the original GW predictions at low separations (high loads), where the scatter is also much smaller than at high separations. This can be expected because the interaction effect is independent of the exact location of asperities and hence of the random aspect of the realized surfaces. Moreover, in good agreement with the numerical results, the effect of interaction is emphasized in the conductance versus separation diagram due to the differentiation of the load with respect to d instead of d_1.

The same contact results are plotted in terms of load in Fig. 5, where it is confirmed that the area–load relationship is almost perfectly linear (except for the scatter at low loads). Similarly, the conductance is not too far from linear at low loads. However, the effect of interaction is to reduce the conductance at intermediate load levels with respect to the original GW predictions.

Obviously, these results are obtained with a significant distance from the full load limit (where asperity theories and models do not work properly) and a correct theory would show at larger loads that the area–load would deviate from the linearity. Notice however that the deviation from linearity for the conductance occurs already with relatively modest loads, this being the effect of the differentiation.

5. Conclusion

An improved version of the original GW theory has been proposed, permitting to include the effect of interaction in the simplest manner, by considering the effect of the mean pressure as a uniform displacement of the surface. This has been shown to fit relatively well the numerical results obtained using a discrete version of the GW theory recently proposed by Ciavarella et al. [3]. In particular, interaction effects are important for the prediction of gas-tightness, and to some extent of elastic stiffness and conductance of the interface. Finally, they tend to reduce the effect of scatter at low separations.

Acknowledgment

Thanks to G. Valenza for providing numerical data in the plots.

References

[1] J.A. Greenwood, J.B.P. Williamson, The contact of nominally flat surfaces, Proc. Roy. Soc. Lond. A 295 (1966) 300–319.
[2] R.E. Jones, D.A. Zeigler, A method for determining the asperity distribution of contacting rough surfaces, J. Tribol. ASME 127 (2005) 24–29.
[3] M. Ciavarella, V. Delfine, G. Demelio, A "re-vitalized" Greenwood & Williamson model of elastic contact between fractal surfaces, J. Mech. Phys. Solids 54 (2006) 2569–2591.
[4] S. Timoshenko, J.N. Goodier, Theory of elasticity, McGraw-Hill, New York, 1951.
[5] M. Ciavarella, S. Dibello, G. Demelio, Conductance of rough random surfaces, Int. J. Solids Struct. 45 (2008) 879–893.
[6] M. Ciavarella, G. Murolo, G. Demelio, Elastic contact stiffness and contact resistance for the Weierstrass profile, J. Mech. Phys. Solids 52 (2004) 1247–1265.
[7] J.R. Barber, Bounds on the electrical resistance between contacting elastic rough bodies, Proc. Roy. Soc. Lond. A 459 (2003) 53–66.
[8] R. Holm, Wiss. Veroff. Siemens-Werken, 7(2) (1929) 217–258.
[9] J.A. Greenwood, Constriction resistance and the area of real contact, Br. J. Appl. Phys. 17 (1966) 1621–1632.
[10] M. Ausloos, D.H. Berman, A multivariate Weierstrass–Mandelbrot function, Proc. Roy. Soc. Lond. A 400 (1985) 331–350.
[11] M. Ciavarella, G. Demelio, J.R. Barber, Y.H. Jang, Linear elastic contact of the Weierstrass profile, Proc. Roy. Soc. Lond. A 456 (1994) 387–405.
[12] J.A. Greenwood, A simplified elliptic model of rough surface contact, Wear 261 (2006) 191–200.
[13] S. Hyun, L. Pei, J.F. Molinari, M. Robbins, Finite-element analysis of contact between elastic self-affine surfaces, Phys. Rev. E 70 (2004) 026117.
[14] M. Borri-Brunetto, A. Carpinteri, B. Chiaia, Scaling phenomena due to fractal contact in concrete and rock fractures, Int. J. Fract. 95 (1999) 221–238.
[15] M. Borri-Brunetto, B. Chiaia, M. Paggi, Multiscale models for contact mechanics of rough surfaces, in: R. Buzio, U. Valbusa (Eds.), Advances in Contact Mechanics: Implications for Materials Science, Engineering & Biology, Research Signpost, Trivandrum (India), 2006, pp. 1–41 (Chapter 1).