Half a century of the Greenwood & Williamson paper

J A Greenwood Cambridge

Emeritus Reader in Tribology

The GW theory really is 51 years old!

Proceedings of the

2nd International Symposium on Electric Contact Phenomena

GRAZ AUSTRIA

MAY 4,5,6 1964

Edited by P. A. KLAUDY

Technische Hochschule GRAZ

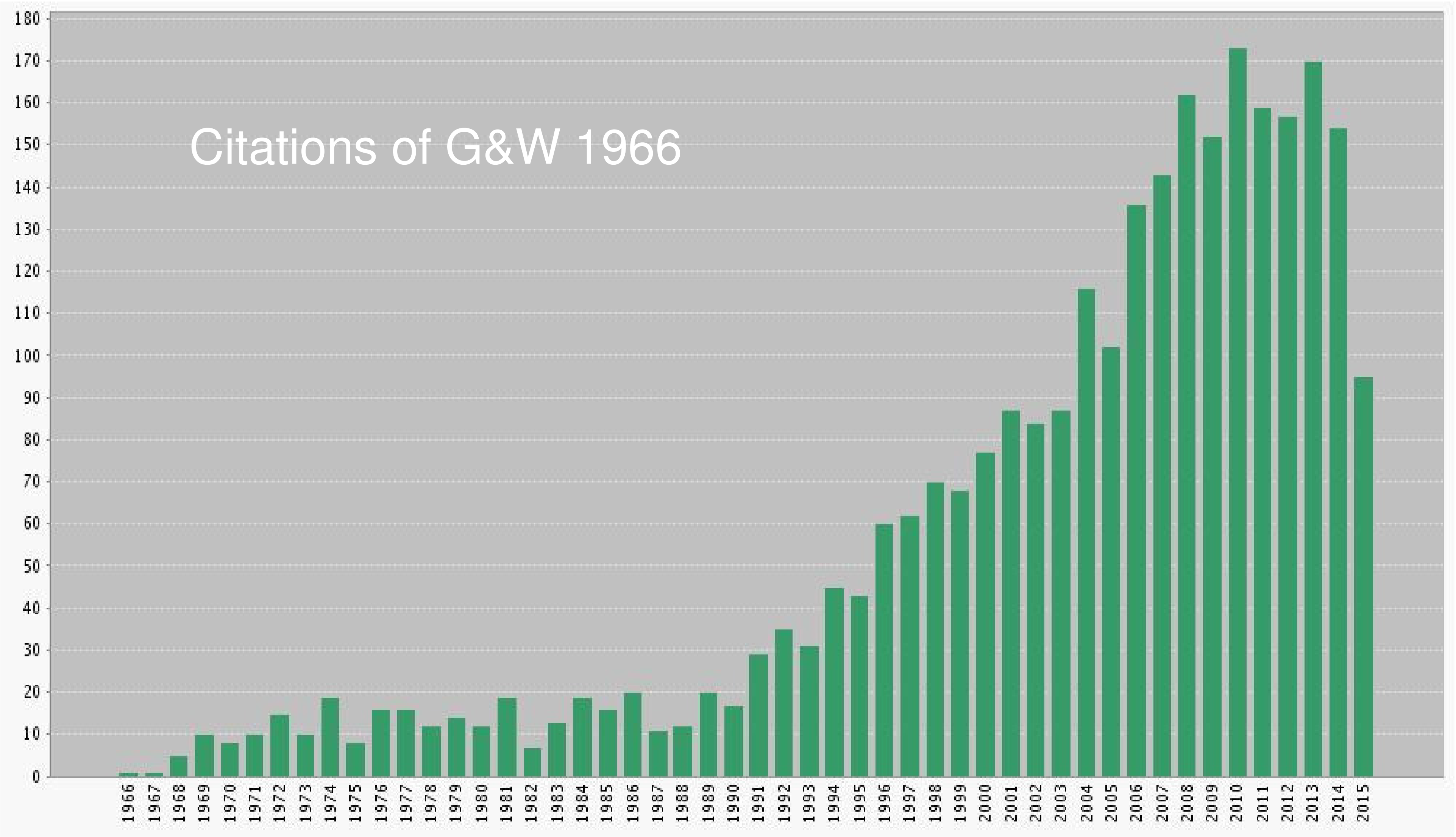
Lehrstuhl für Grundlagen der Elektrotechnik

Research Report No. 15
July 24, 1964

BURNDY
RESEARCH
DIVISION

The Contact of Nominally Flat Surfaces

J. A. Greenwood
J. B. P. Williamson



But how many of them have read it ?

Of course it all began with Archard……..
and this was long before the concept of a fractal *was invented.*

And perhaps we took over the idea of multiple Hertzian contacts from him

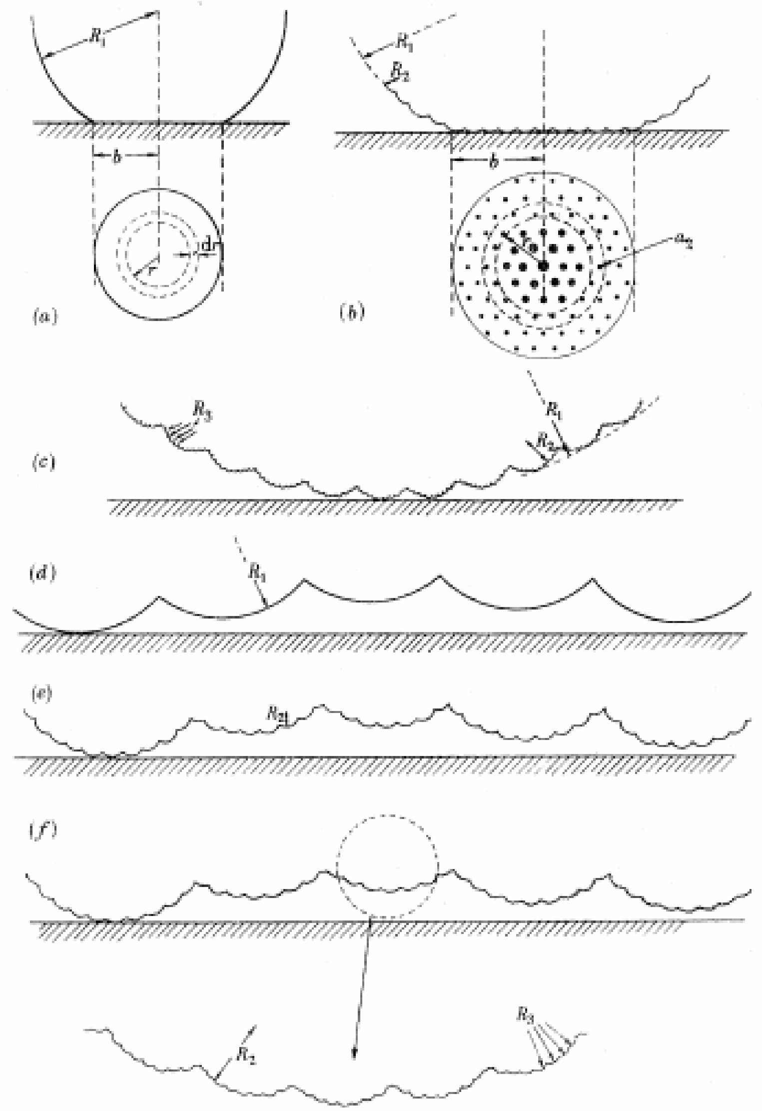


FIGURE 1. Models used in the theory. In (*a*) and (*b*) the surfaces are shown under a load, in (*c*)–(*f*) under zero load. The deduced relations between A and W for these models are (*a*) $A \propto W^{\frac{2}{3}}$, (*b*) $A \propto W^{\frac{8}{9}}$, (*c*) $A \propto W^{\frac{26}{27}}$, (*d*) $A \propto W^{\frac{4}{5}}$, (*e*) $A \propto W^{\frac{14}{15}}$, (*f*) $A \propto W^{\frac{44}{45}}$.

The same year (1958) F F Ling published his contact analysis, accompanied by some good experimental load ***v*** approach data.

So what did he do wrong, ..
or what did we do right, so that we, not Ling, become a standard reference?

Ling assumed, naturally, that the asperities deformed plastically. And he offered too many asperity shapes [*and an implausible fracture mechanism*], and too many possible height distributions ***without measuring any***

But what really mattered was his choosing the point of first contact as his datum, and seeking a power law relation between load and approach.

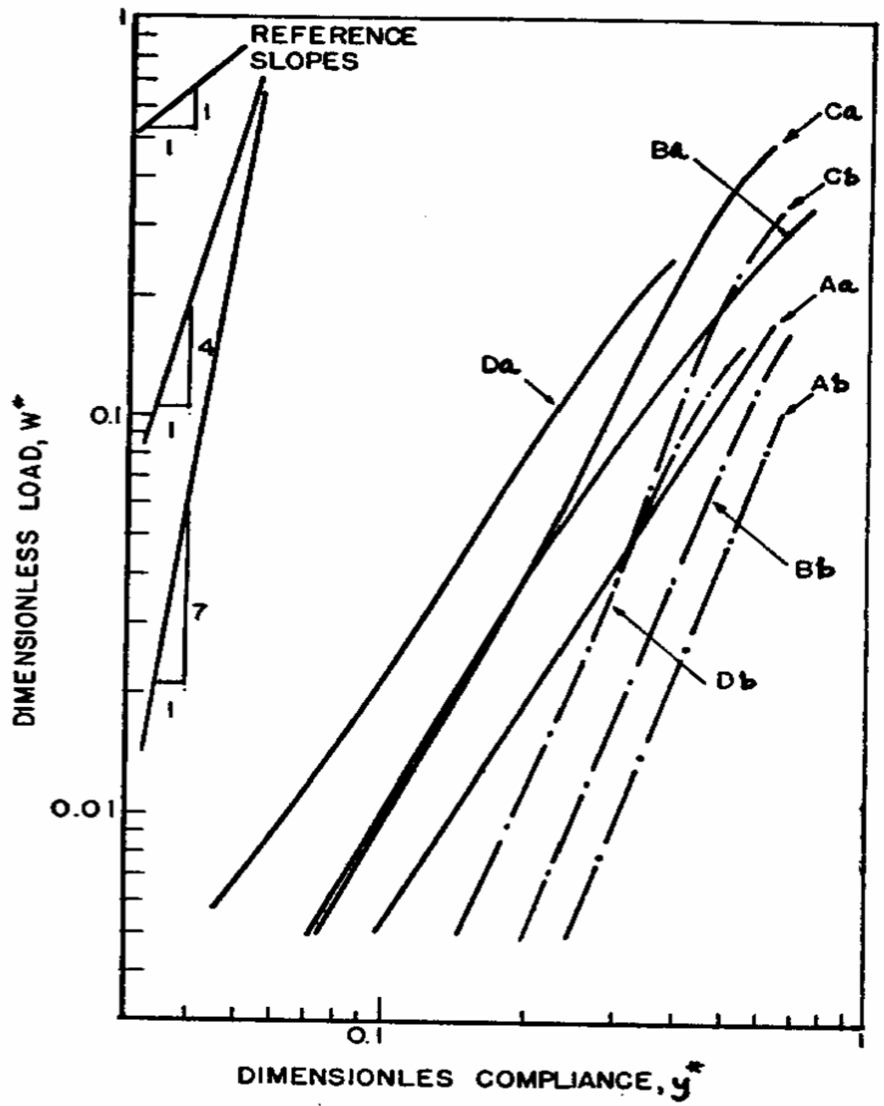


Fig. 1. Log-log plot of dimensionless load *vs* dimensionless compliance of eight selected models. Aa—compression by a rigid flat of a set of rigid, perfectly-plastic wedges or hemispheres uniformly distributed through a depth of unity from the surface. Ba—same distributed linearly. Ca—same distributed normally. Da—same having a Poisson distribution. Ab, Bb, Cb, and Db—same as Aa, Ba, Ca, and Da respectively except cones are compressed.

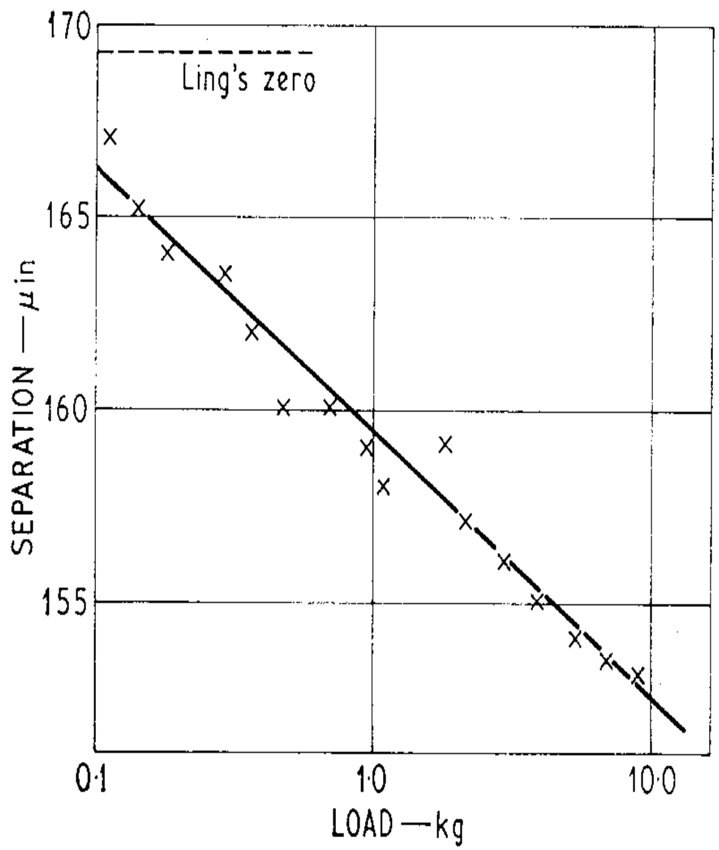


Fig. 7. Replot of Ling's experimental results (Fig. 9 of (15)) Ling's original criticism of single-rough-surface models was based on the discrepancy between experimental and theoretical curves on a log–log plot. On a log–linear plot, the discrepancy does not appear and it seems that it arises from an invalid extrapolation to zero load to find the origin for the log–log plot

Proc Instn Mech Engrs 1970–71

Ling plotted his data on a log-log plot $W(\delta)$ ..but this just gave a curve, with a slope increasing from 2 to 8

The point of first contact is an unreliable datum.
Even with a large population it will be erratic, and its neighbours can be ***anywhere***

In 1963 SKF didn't have a computer in the laboratory either!

Lubricant Films in Rolling Contact of Rough Surfaces

T. E. TALLIAN, Y. P. CHIU, D. F. HUTTENLOCHER, J. A.KAMENSHINE, L. B. SIBLEY, & N. E. SINDLINGER

Surface microgeometry of the rolling tracks on the balls is statistically analyzed by processing electrical analogs of surface profiles through on-line computing equipment.

The output of the surface tracing instrument was fed into an FM magnetic tape recorder according to diagram A, Fig. 10. The low-frequency band pass filter inserted between the surface tracing instrument and the tape recorder was set to a pass band of calculated width. It will be seen that this finite band width is necessary for usable results.

A. RECORDING

TRANSDUCER ⇒ CARRIER AMPLIFIER & DEMODULATOR ⇒ 0.02 – 20 CPS BAND PASS FILTER ⇒ AMPLIFIER ⇒ F M TAPE RECORDER

The tape-recorded electrical analog of the surface profile was then processed through the circuitry shown in block diagram B of Fig. 10. The arrangement has a common input consisting of the playback system of the FM tape recorder, an amplifier, and a variable band pass filter which, again, was set as explained later

-- -- -- -- -- -- -- -- -- -- --

Output Channel 1 is a level discriminator, operating in conjunction with a 400 channel memory, being swept by an internal clock at a predetermined rate. *[To give a frequency distribution of dwell times at a chosen level]*

-- -- -- -- -- -- -- -- -- -- --

Output channel 2 comprises circuitry to obtain an amplitude histogram of the signal by way of periodic sampling triggered by a pulse generator. Each momentary amplitude sampled is converted in the amplitude time converter to a proportional time interval. The time intervals are used to accumulate counts in the memory unit as described for Channel 1, giving the amplitude distribution. Print-out is performed on command.

They were the first *tribologists* to think of Gaussian height distributions and to use signal theory to understand contacts

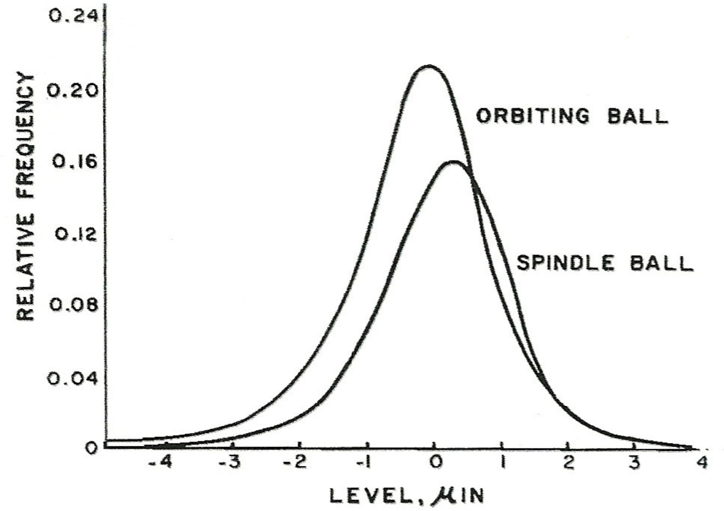


FIG. 11A. Roughness amplitude distribution; frequency diagram on linear scale.

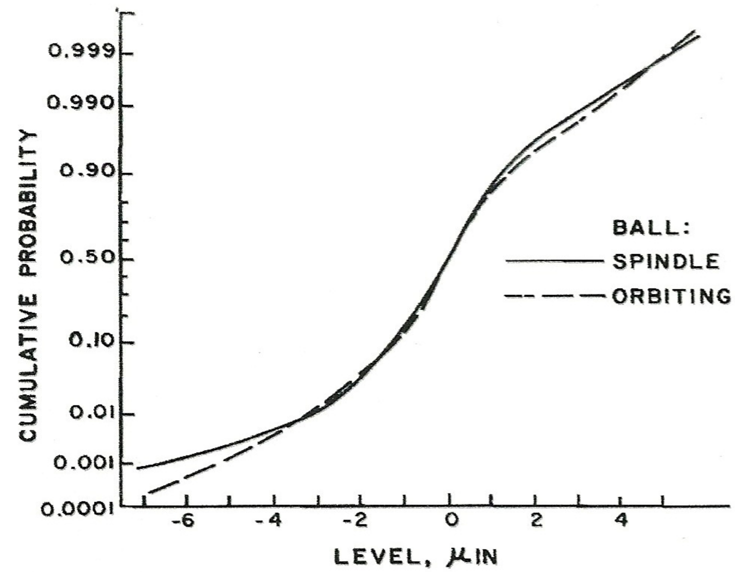


FIG. 11B. Roughness amplitude distribution; cumulative distribution on Gaussian scale.

So why G&W 1966 ?

We actually measured surface roughness…but so did Abbott & Firestone in 1932, (and invented the bearing area curve): Bickell (1963) published work showing heights were gaussian ...by drawing lines on the pen recorder output and counting …Tallian's group found the height distribution by sorting the signal into a 400 channel memory….so was it ***feeding it into a computer*** that made the difference?. Or was it the (obvious) next step; using the computer to locate peaks, so we could plot their heights and curvatures …and link up with Archard's ideas?

Perhaps we just got the timing right: for the metrologists (Reason at Rank Taylor-Hobson; Sharman at the National Engineering Laboratory) had also begun to feed their signals into a computer.

But perhaps we got the statistical theory right, by focussing on means and standard deviations, and having nothing to do with extreme values? Or even by firmly avoiding the term *normal* distribution, and using the magic password ***Gaussian***?

Worn surfaces do not have Gaussian height distributions: and random field theory is inapplicable! But the higher peaks may well behave as Gaussian….

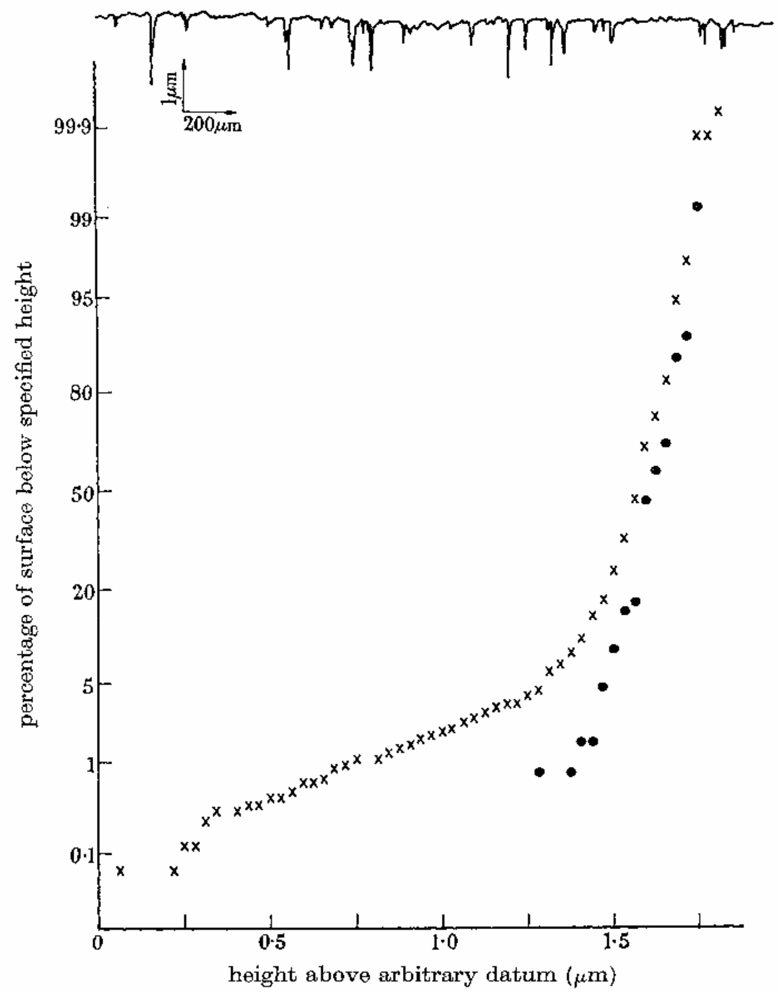
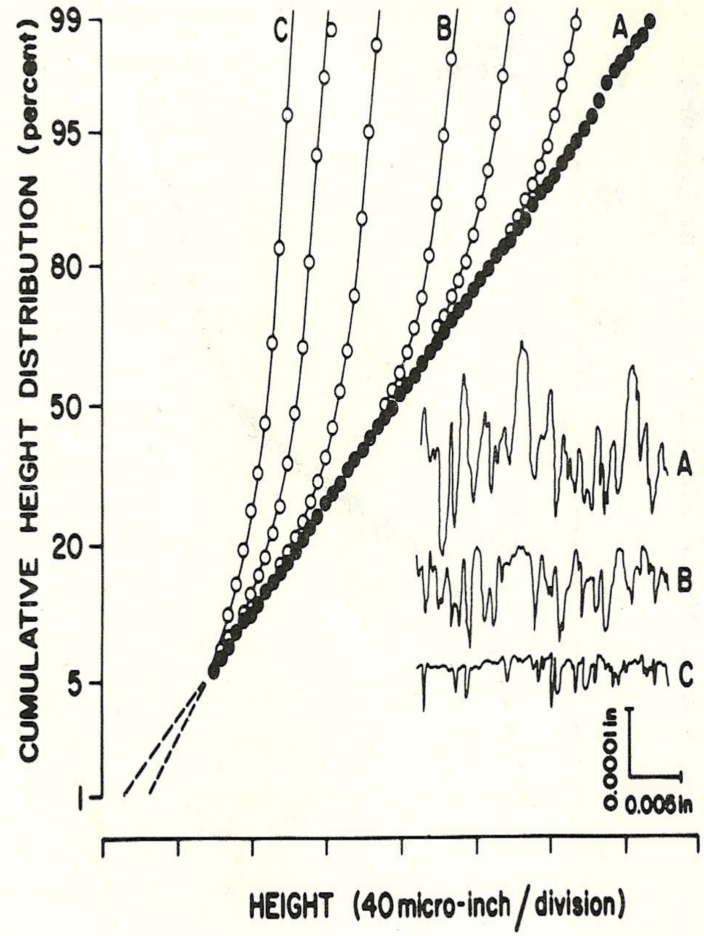


FIGURE 6. Cumulative height distribution of mild steel specimen. Distribution of all heights, ×. Distribution of peaks, ●. This specimen was abraded on 400 grade carborundum paper, then slid against a copper block flooded with oleic acid, at approximately 10 Kg, 130 cm/s for 30 s. Although the distribution is at first sight highly non-Gaussian, in fact nearly 90 % of the surface is approximately Gaussian; the surface, with an actual standard deviation of 1·3μm, would behave in contact as if Gaussian with a standard deviation of half this. The profile of the same surface is shown in the upper diagram: the vertical magnification is 200 times the horizontal magnification.

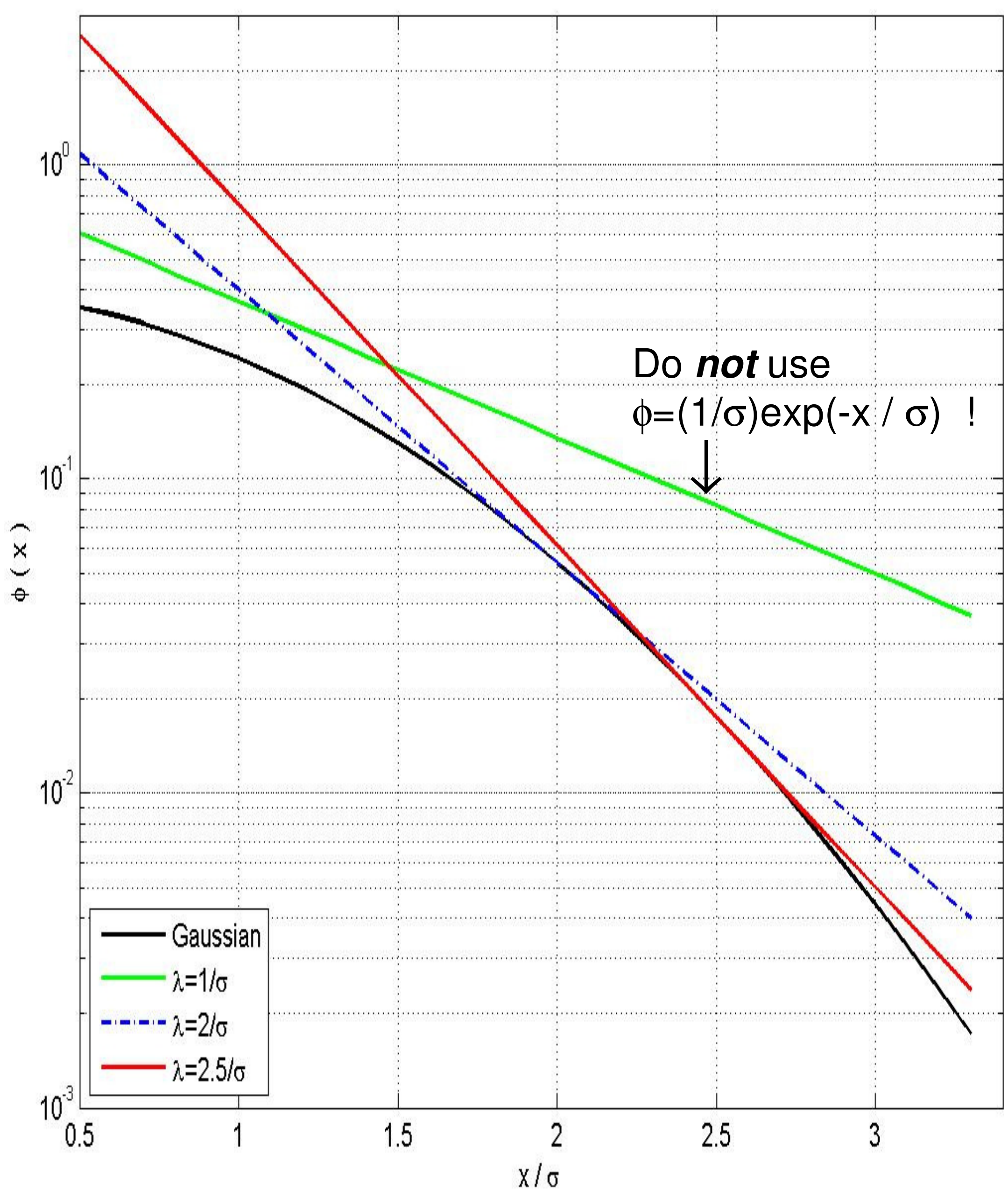
Probability paper is a more informative way of studying wear *than measuring skewness or kurtosis*



CUMULATIVE HEIGHT DISTRIBUTIONS SHOWING THE EFFECT OF WEAR ON THE INITIALLY GAUSSIAN HEIGHT DISTRIBUTION (●) OF A BEAD-BLASTED SURFACE. The six non-Gaussian distributions (○) represent, from right to left, progressive stages in the wearing process. The curves are related to each other by setting the heights of the fifth percentiles equal. The experimental points are omitted wherever

Approximating a Gaussian by an exponential makes simple analysis possible: but do use the best exponential !

And since the skewness is completely different, it should stop all investigations of the effect of skewness on contact behaviour !

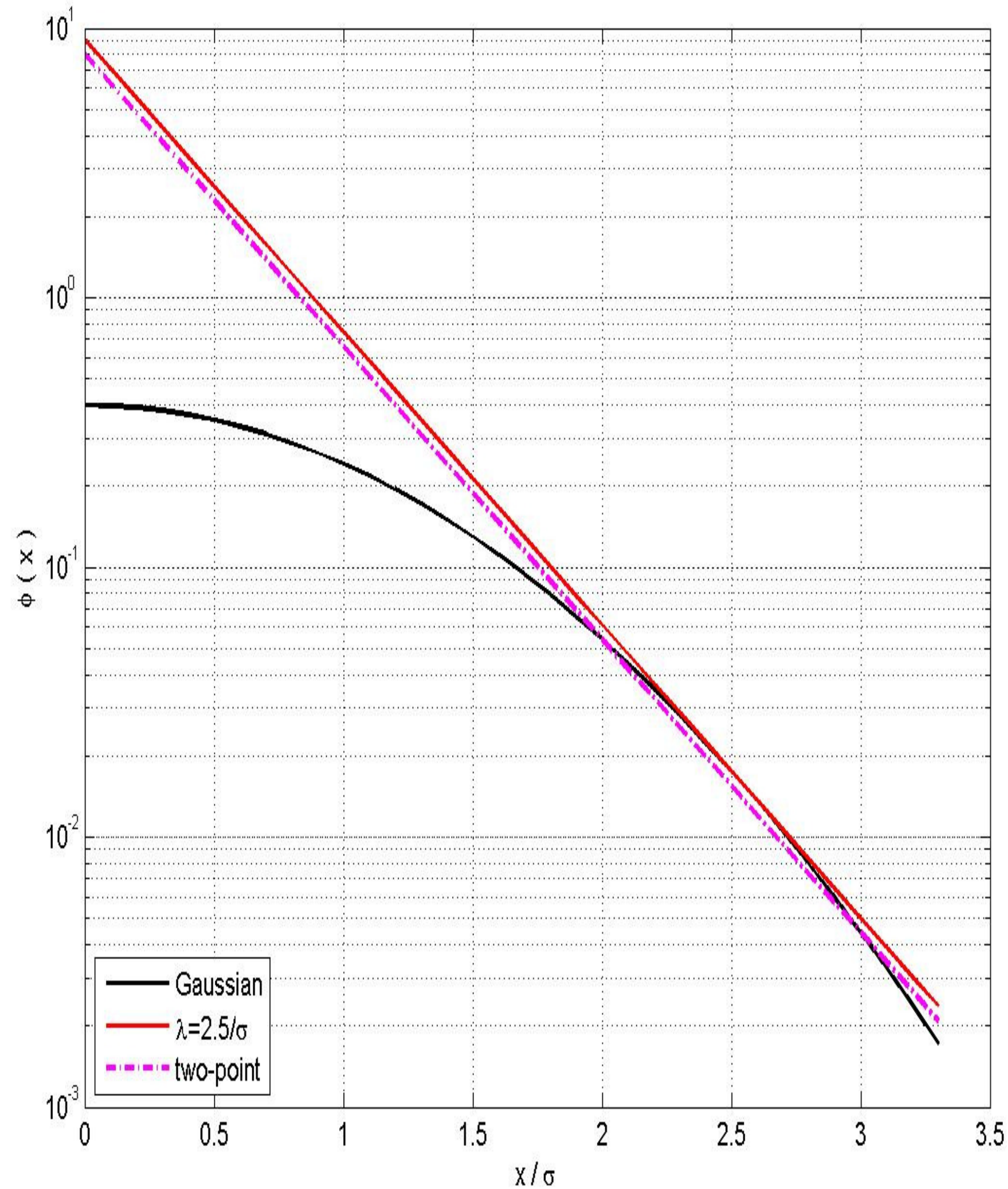


To get the best approximation, fit at two points $x(1)$, $x(2)$: then use

$\Phi(x)=\lambda \exp(-\lambda x)$

with

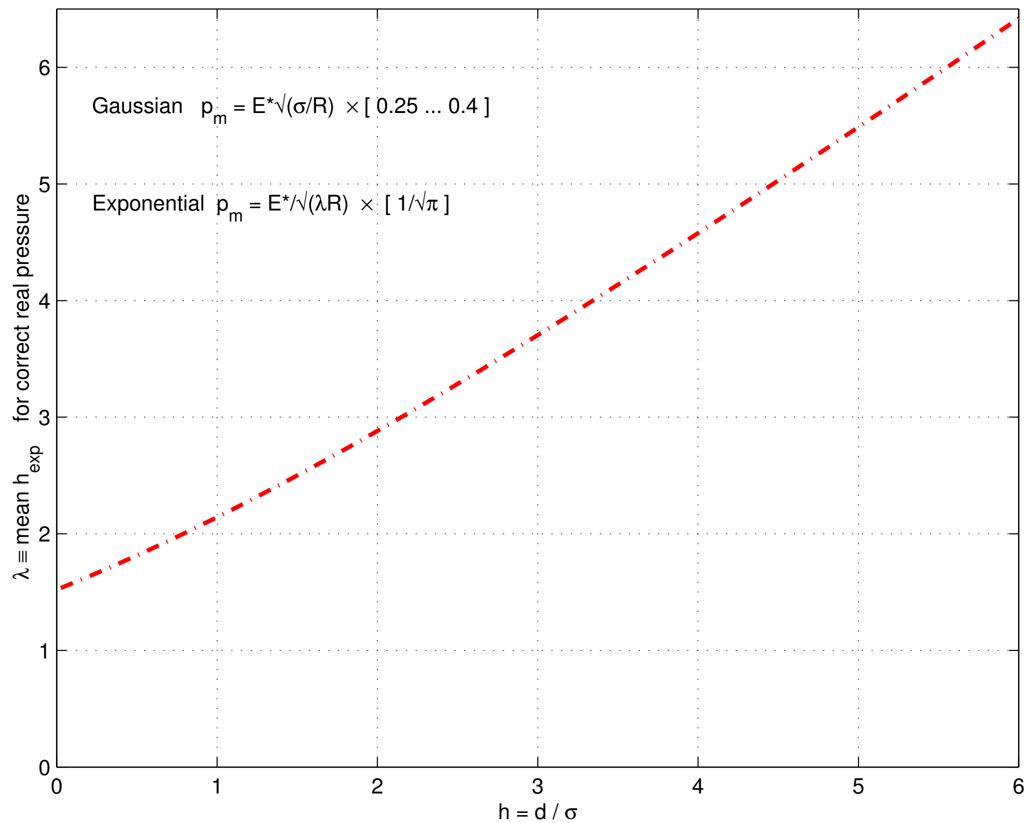
$\lambda=(1/2)[x(1)+x(2)]$



To match the mean real contact pressure at a particular height

choose $\lambda = (h_1+h_2)/2$ so that $1/\sqrt{(\pi\lambda\sigma)} = [\ 0.25 \ldots 0.4\ ]$

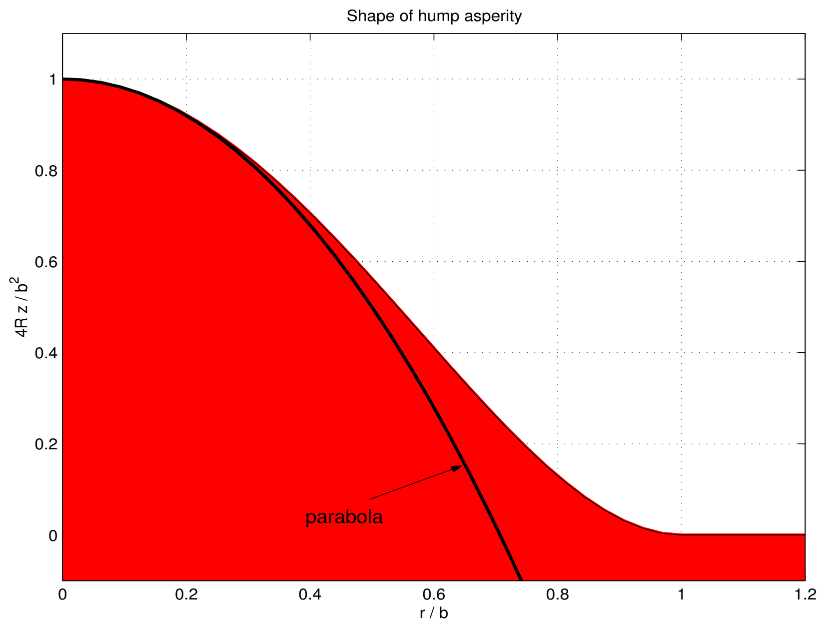
Never choose $\lambda = 1 / \sigma$ **!**



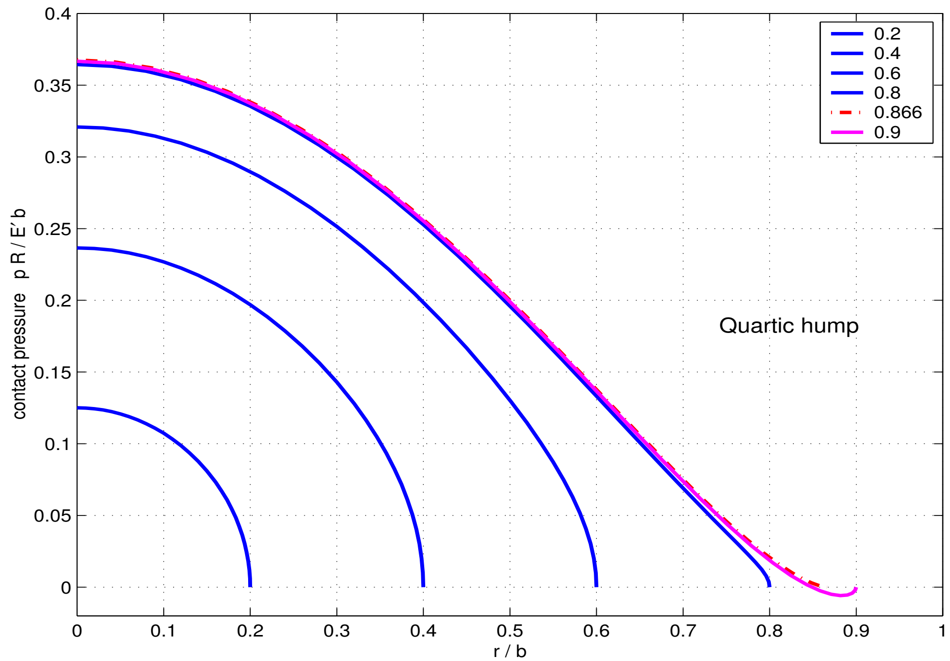
So we can get the "same" answers with an exponential as with a Gaussian. The skewest distribution possible gives the same answers as one with zero skewness.   So??

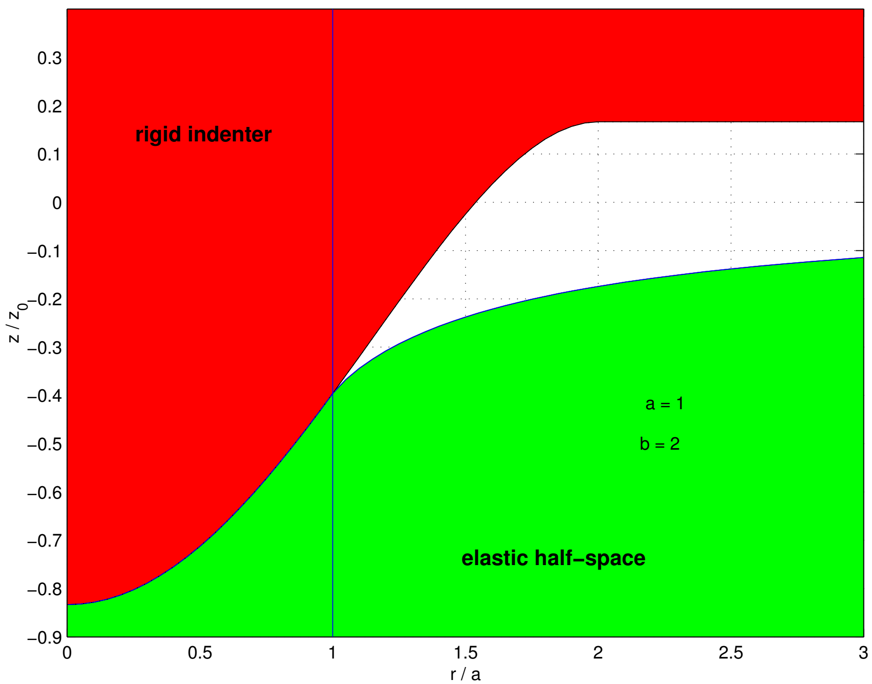
An obvious and necessary extension of GW was to replace Hertz by a load-approach law continuing up to the fully plastic limit. But not by analysing the deformation of a *hemisphere*. If you don't like a parabola, much better is the *"quartic hump"*

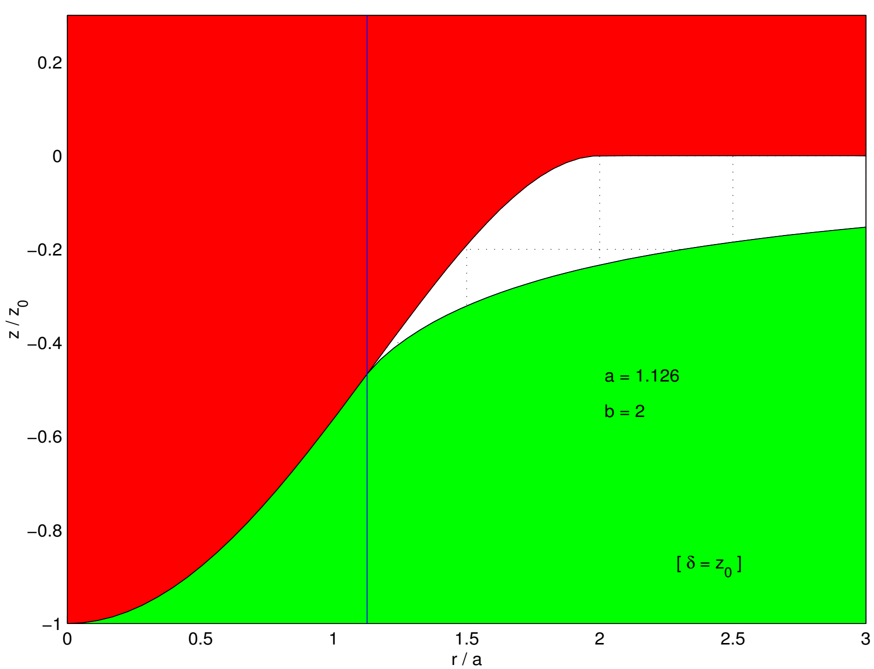
$$z = (1 - r^2 / b^2)^2$$

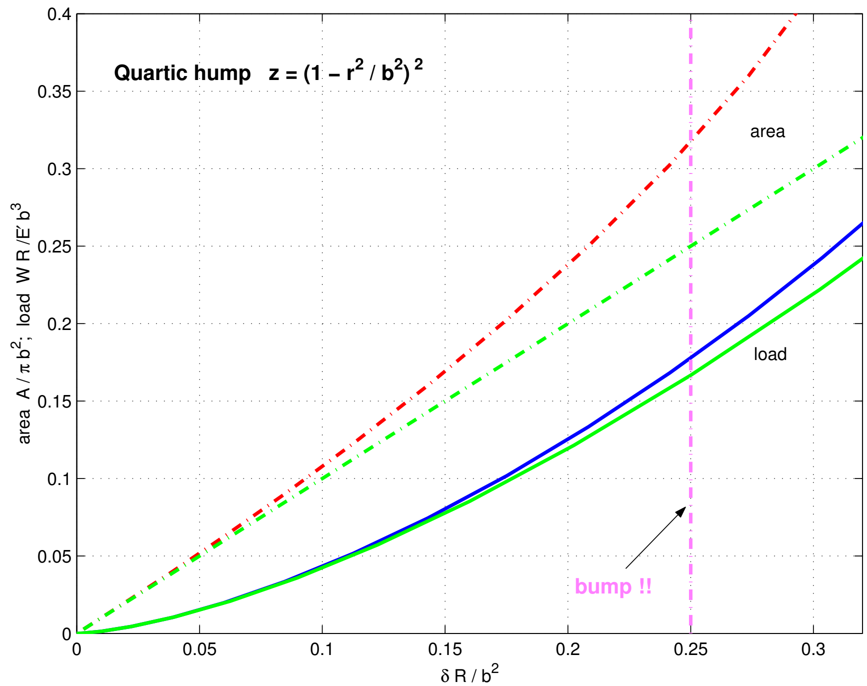


For a / b small, get Hertz pressures: but at $a / b = (\sqrt{3}) / 2$, pressures tend smoothly to zero; and for $a / b > \sqrt{3} / 2$ contact requires tension!







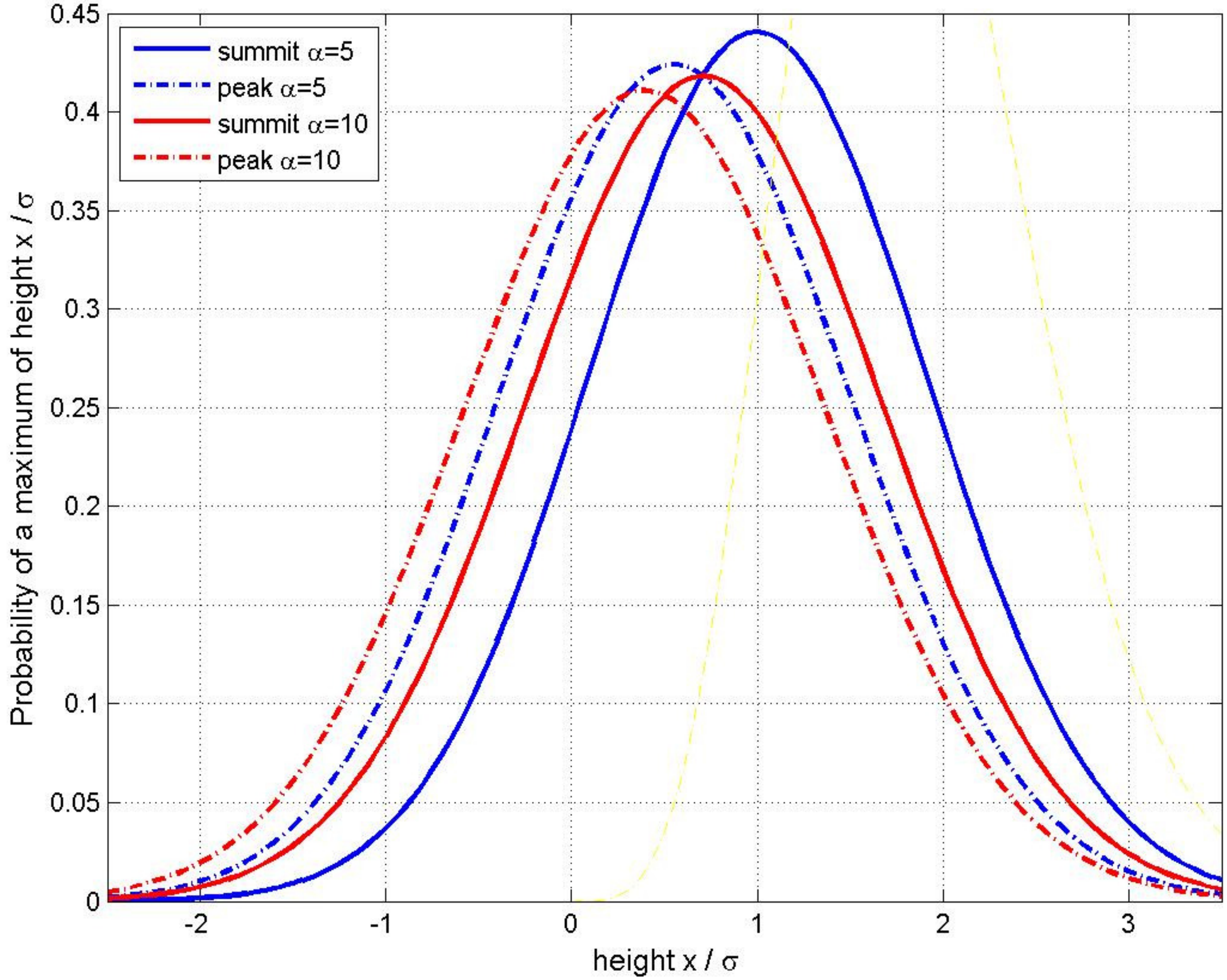


At $R\delta / b^2 = 0.25$, $[\delta = z_0]$ the base planes come into contact.

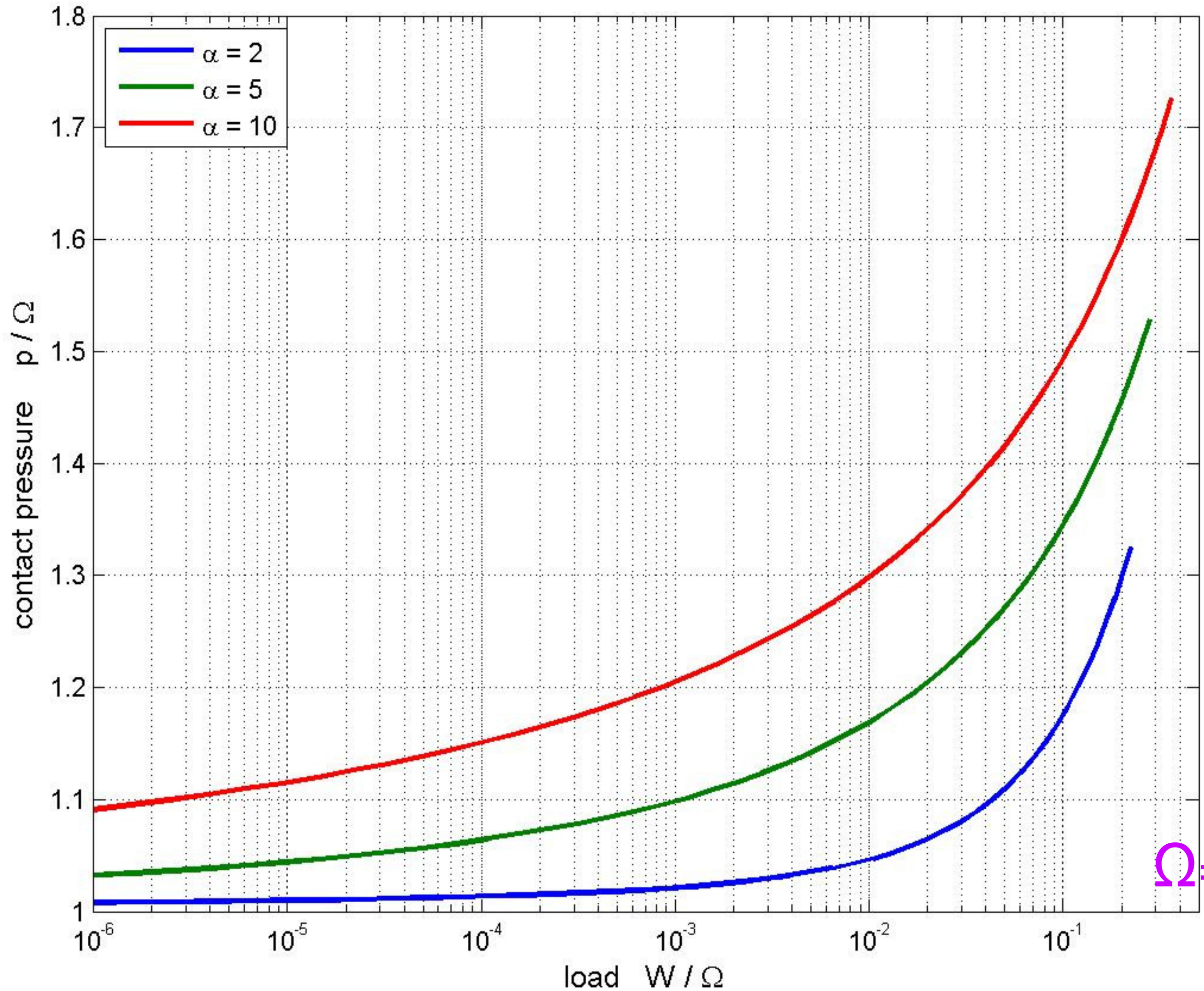
So tensile stresses never in question

Nayak's Gaussian random field theory brought out the different properties of peaks and summits

$\alpha = m_0 m_4 / m_2^2$

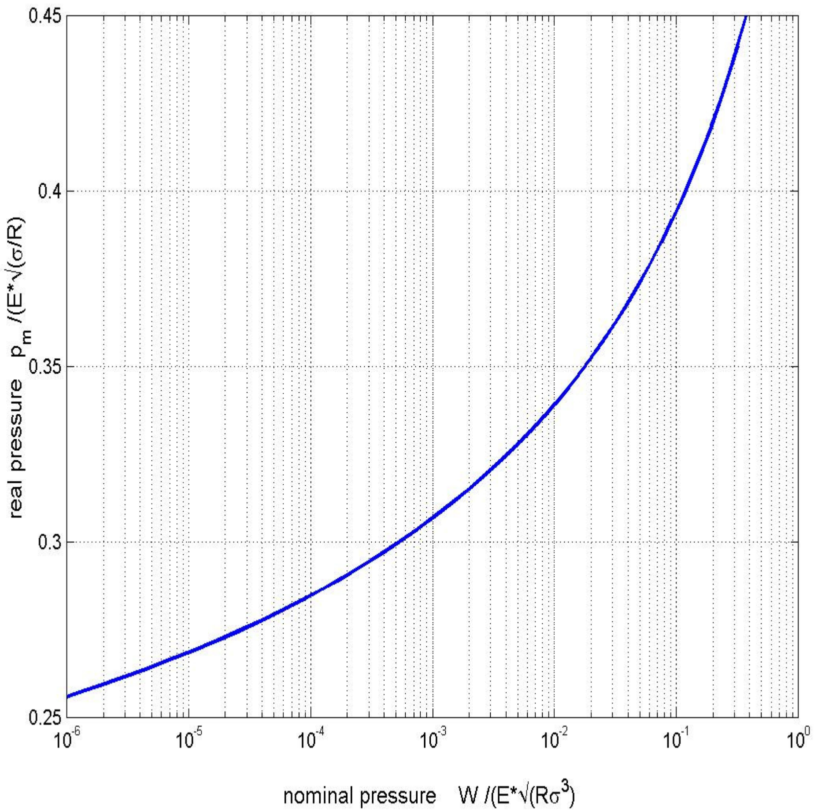


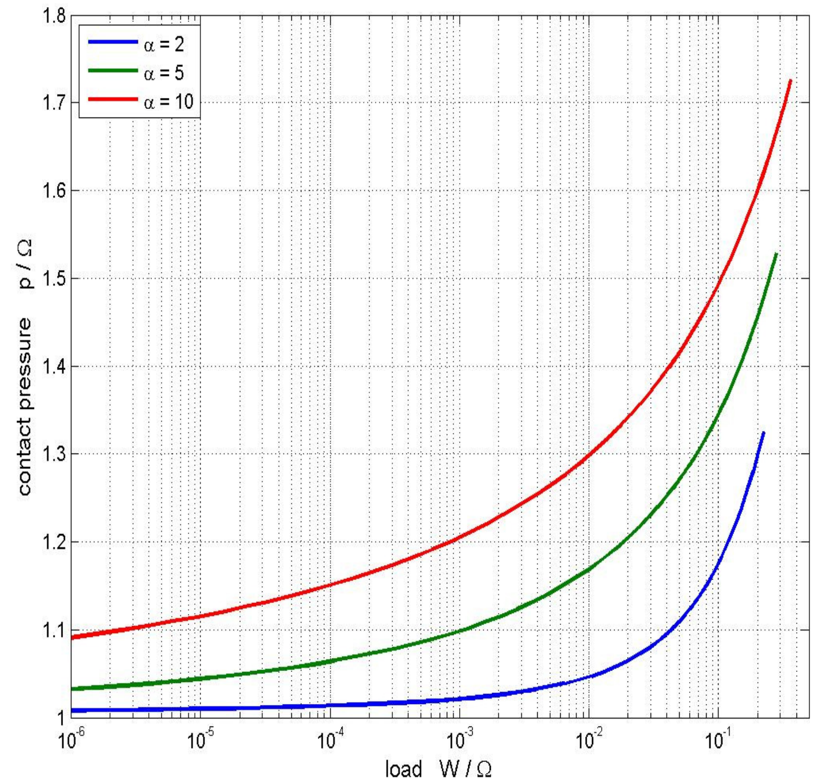
Bush, Gibson & Thomas used Nayak's summit and summit curvature distributions to do a full analysis of elastic contact of a Gaussian surface



$\Omega = E'\sqrt{(m_2/\pi)}$

The GW theory does not give proportionality between load and area, while the BGT theory does… ??



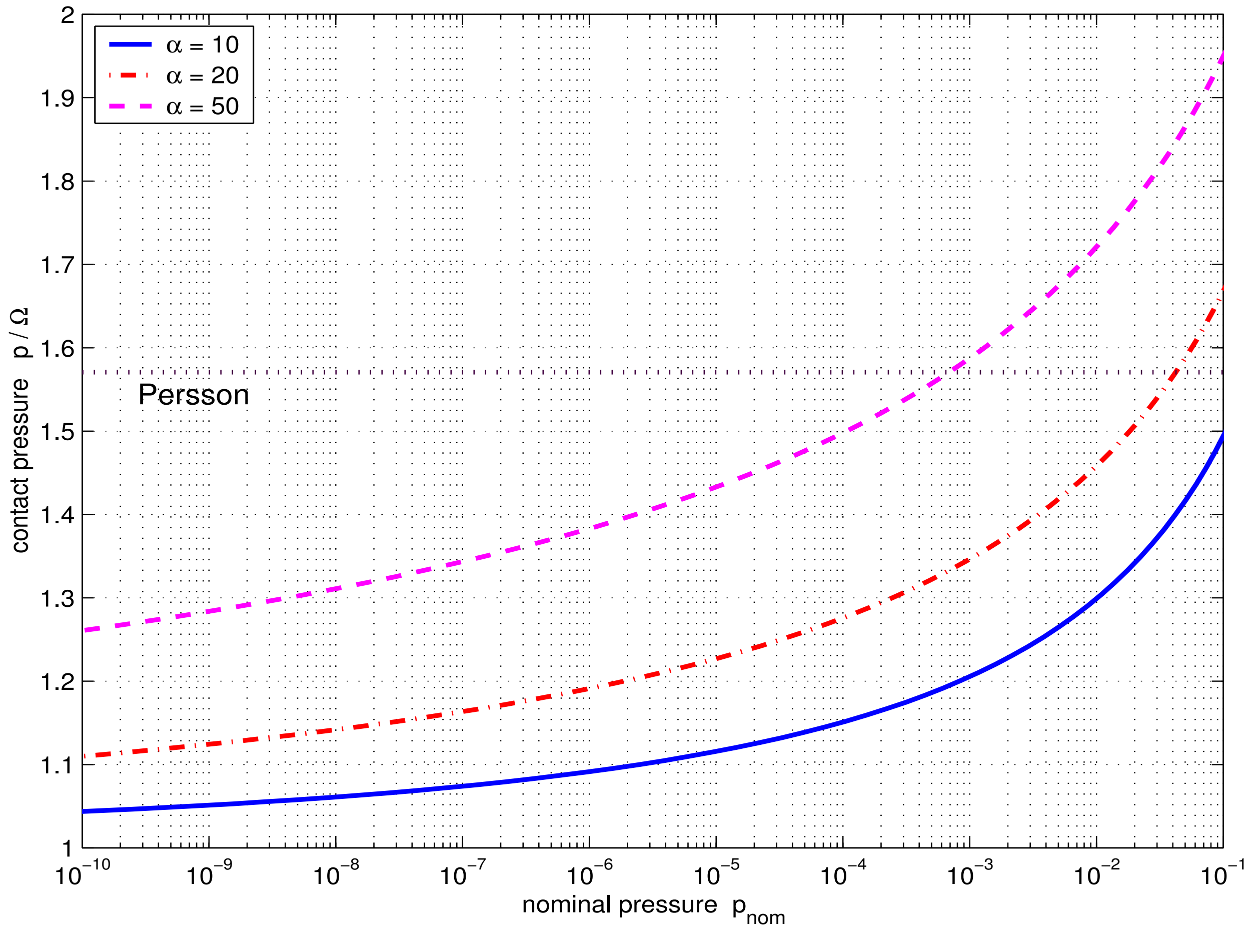


The vaunted asymptotic proportionality between contact area and load only occurs at impractically large separations and negligible loads

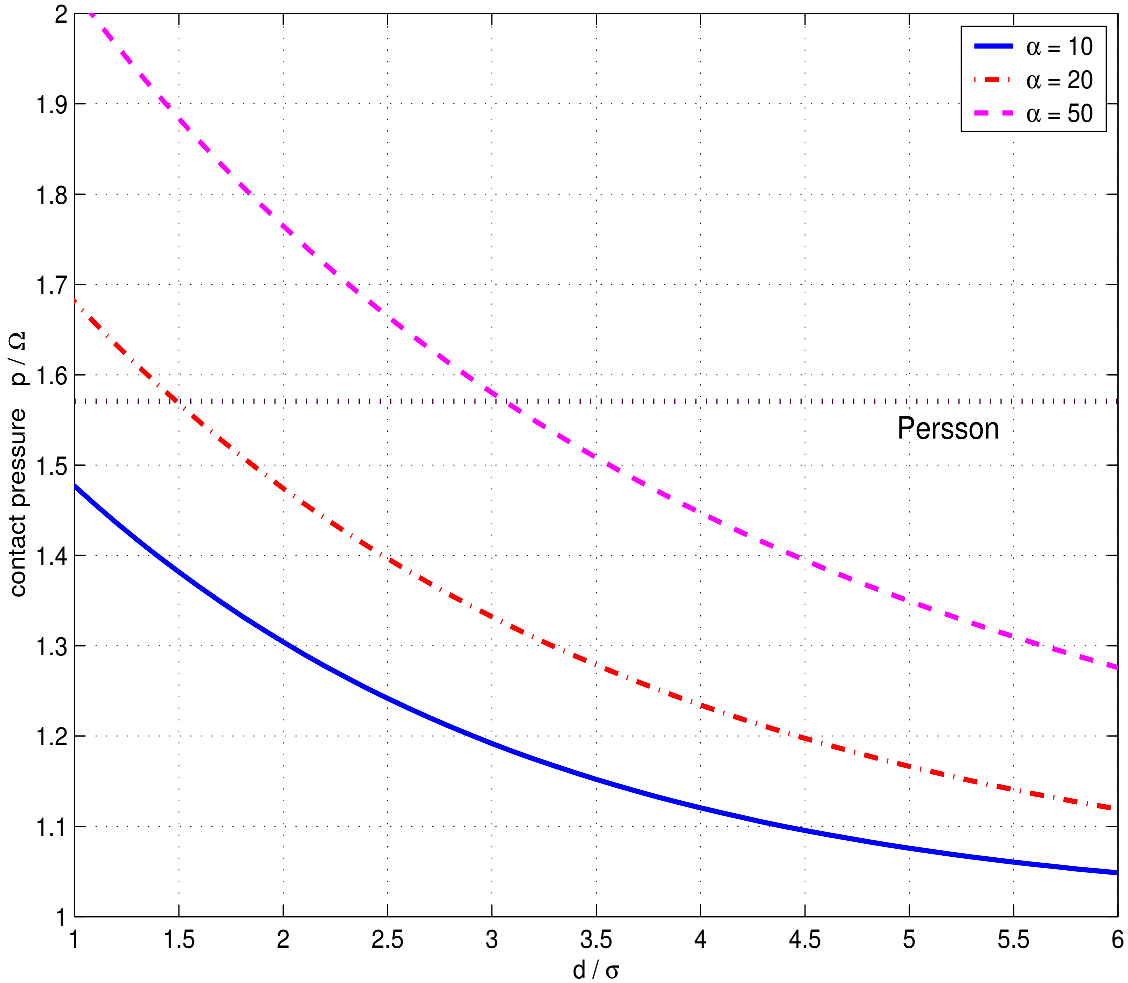
Numerical solutions often compare the values of the reciprocal contact pressure Ω / p_r (not a quantity in everyday use!) with theory.

I find the contact pressure p_r / Ω easier $[\, \Omega \equiv E^* \sqrt{(m_2/\pi)}\,]$.

Usually said that BGT theory is $p_r / \Omega = 1$: while Persson gives $p_r / \Omega = \pi / 2$



But where do we get nominal pressures of 10^{-10} ?

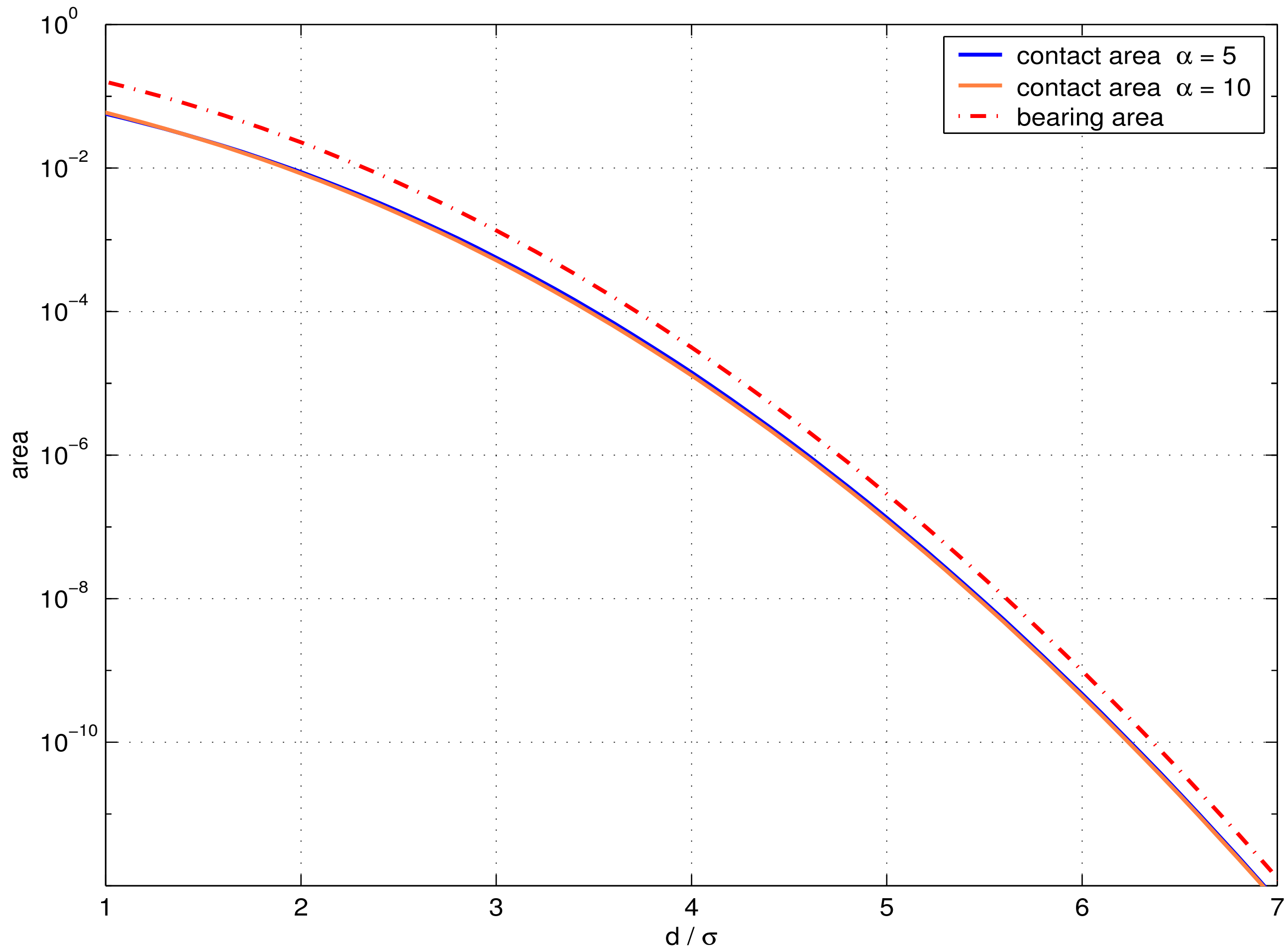


At $d / \sigma = 6$!

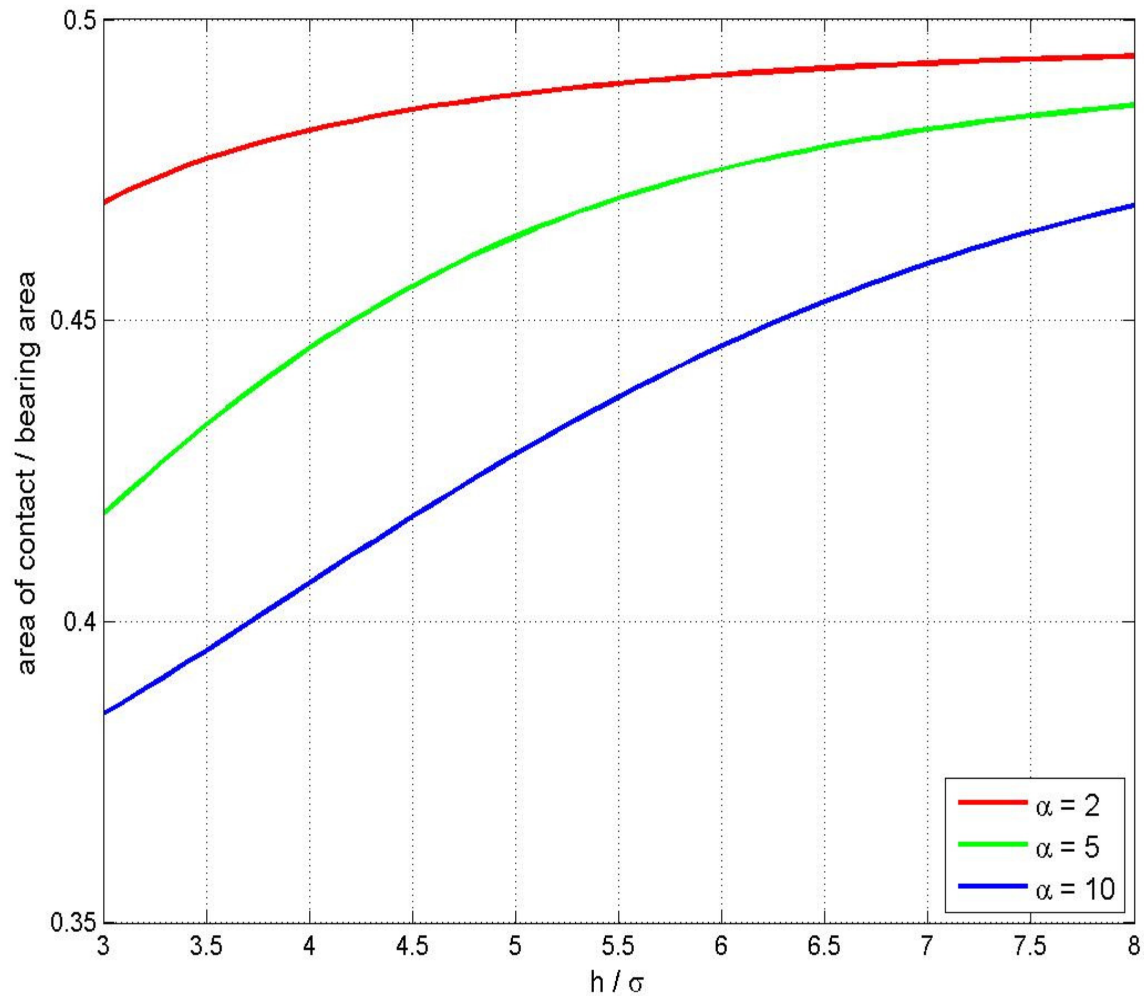
It is hard to believe that any of the numerical solutions used to study the real area of contact found many contact areas at heights above 4σ:
so quoting the BGT asymptotic value is absurd.

An astonishing prediction of the BGT analysis is the close correspondence between the contact area and the bearing area

How can two completely different quantities turn out to be so related?



It's not really very close: until you remember that both quantities vary over a range of 10^4



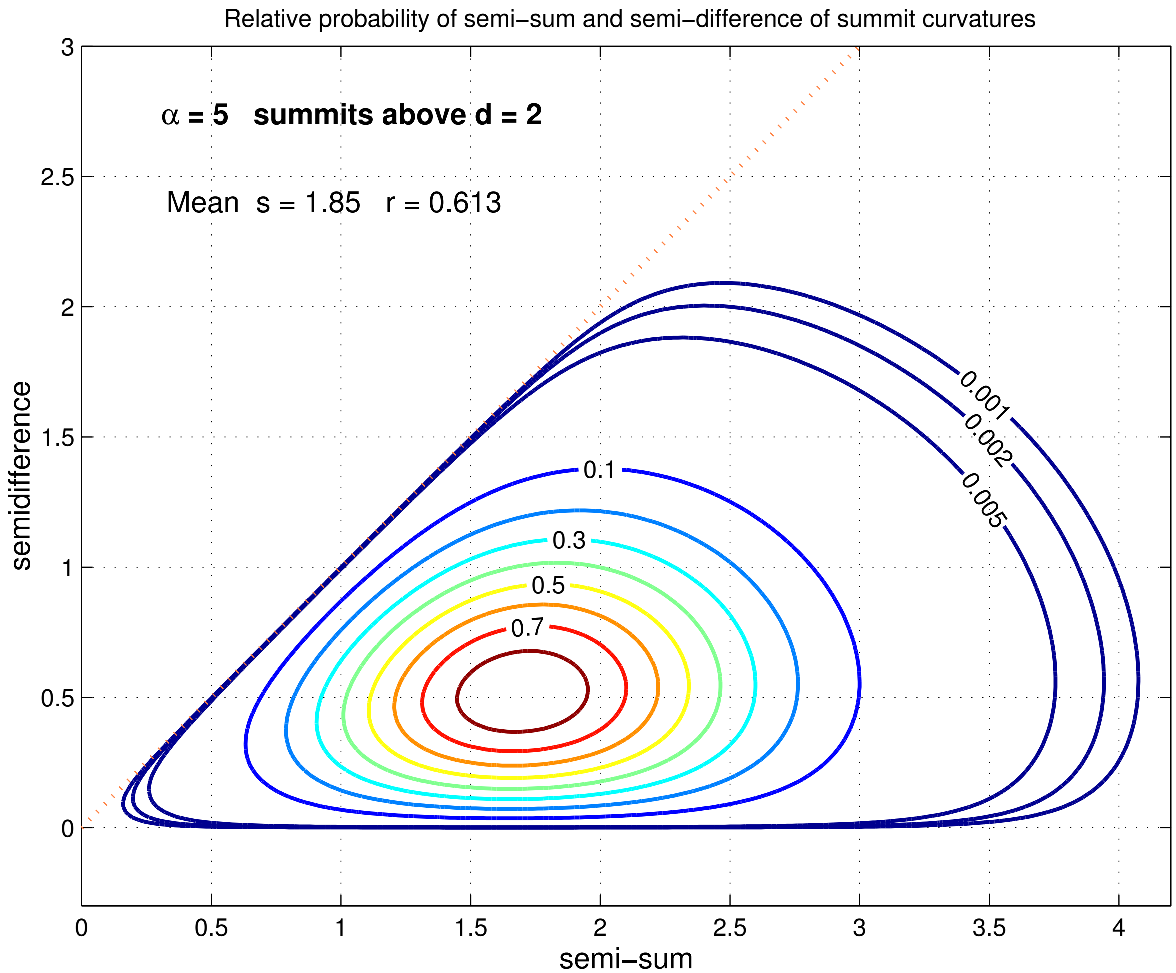
To understand the proportionality, we need to examine what Nayak's analysis says about the shape of asperities

For Nayak himself ducks the issue, contenting himself with finding the mean summit curvature and showing that (and how) it increases with the summit height

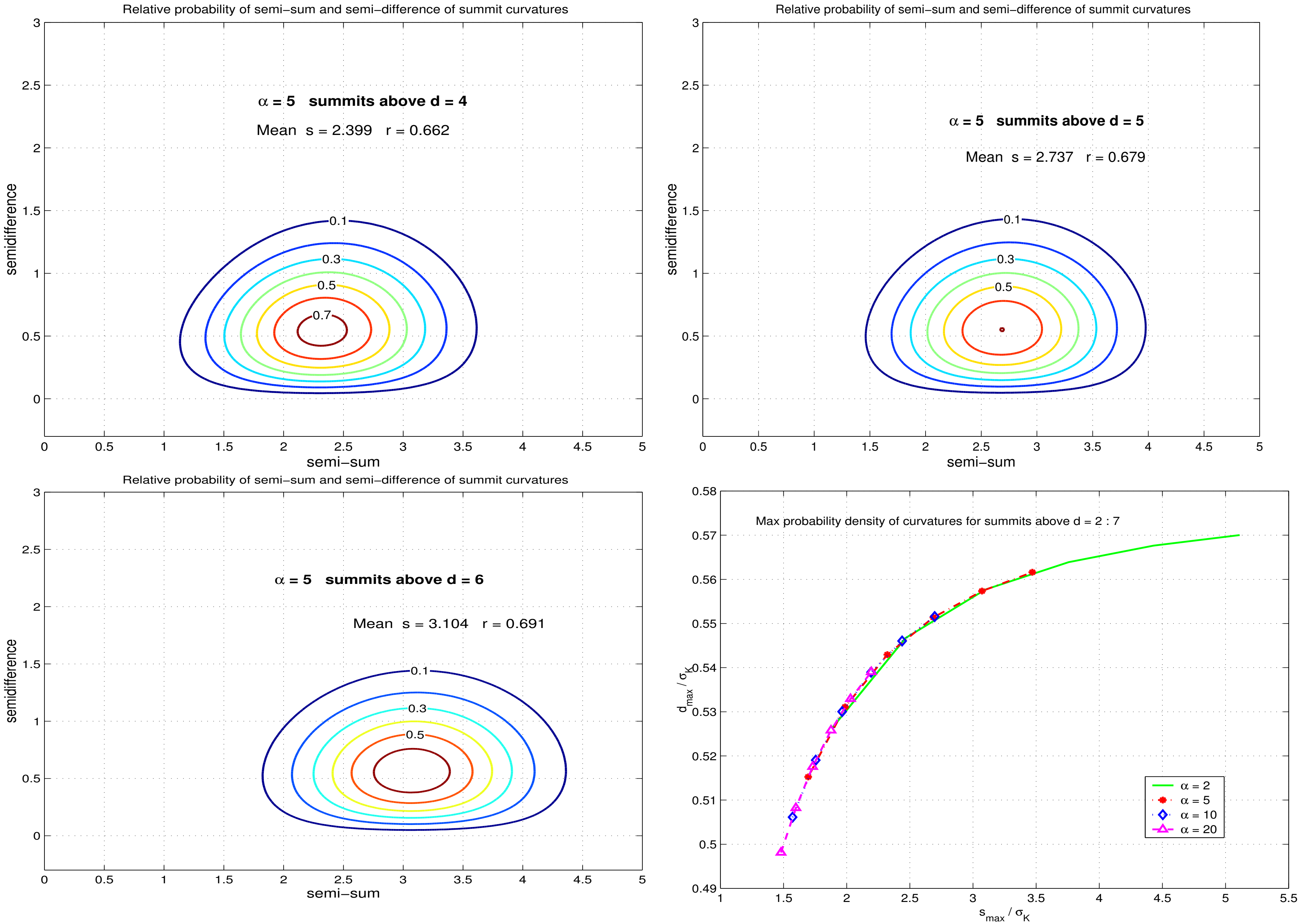
Need we take into account the ellipticity of the summits?

Hertz theory for circular contacts, using $1/R=\sqrt{(\kappa_1\kappa_2)}$ is accurate to 0.1% for $\kappa_1/\kappa_2 < 2$. {and to 2% for $\kappa_1/\kappa_2 < 5$}

Well, what shape are the summits?



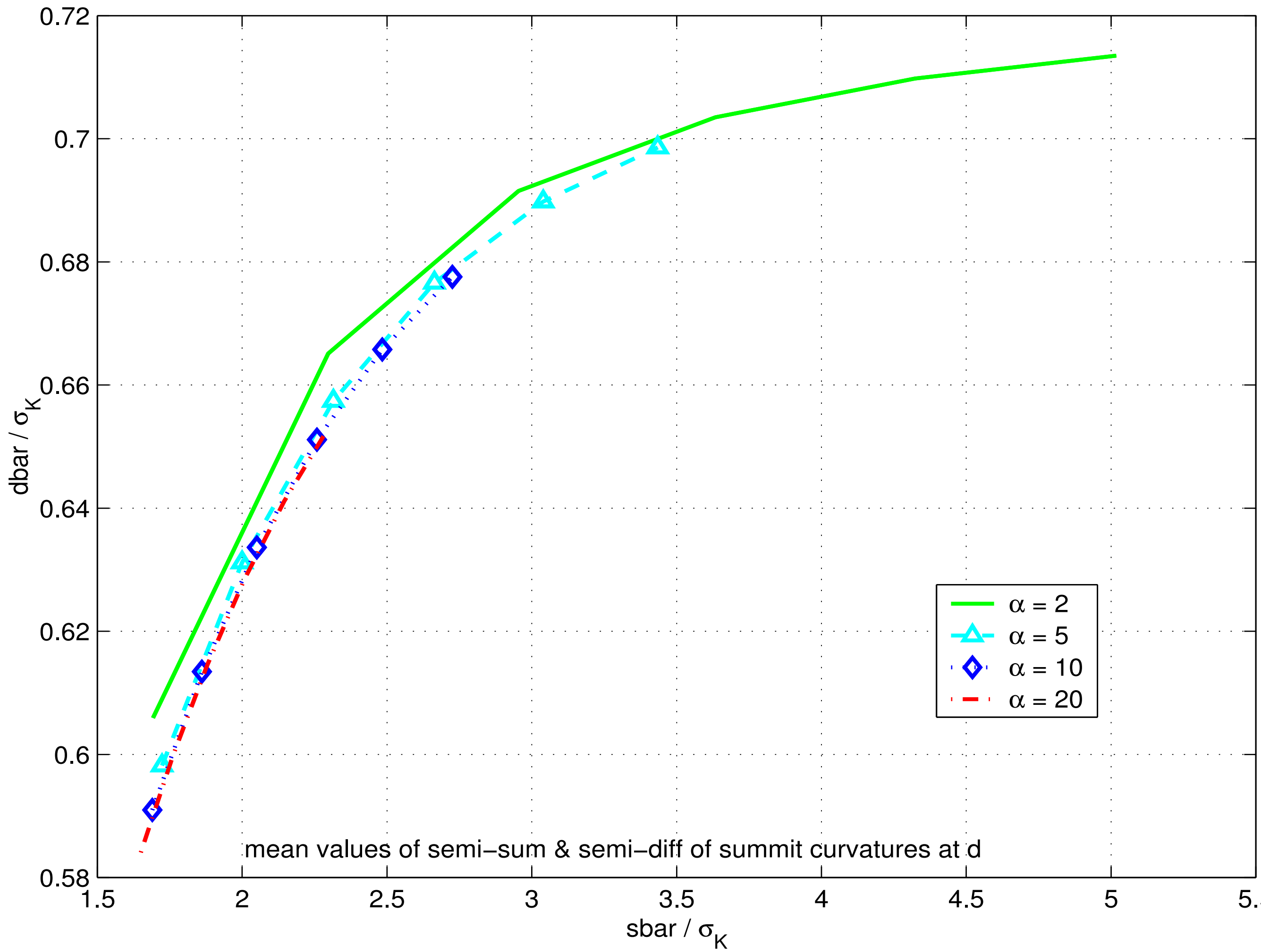
There are ***no*** circular contacts **!**



Relative probability of semi−sum and semi−difference of summit curvatures
α = 5 summits above d = 4
Mean s = 2.399 r = 0.662
semidifference
semi−sum
Relative probability of semi−sum and semi−difference of summit curvatures
α = 5 summits above d = 5
Mean s = 2.737 r = 0.679
semidifference
semi−sum
Relative probability of semi−sum and semi−difference of summit curvatures
α = 5 summits above d = 6
Mean s = 3.104 r = 0.691
semidifference
semi−sum
Max probability density of curvatures for summits above d = 2 : 7
d_max / σ_K
s_max / σ_K
α = 2
α = 5
α = 10
α = 20

Summits get more circular: but is κ_1/κ_2=2 adequately circular? Very definitely so:

Hertz theory for circular contacts using $\sqrt{(\kappa_1\kappa_2)}$ is valid

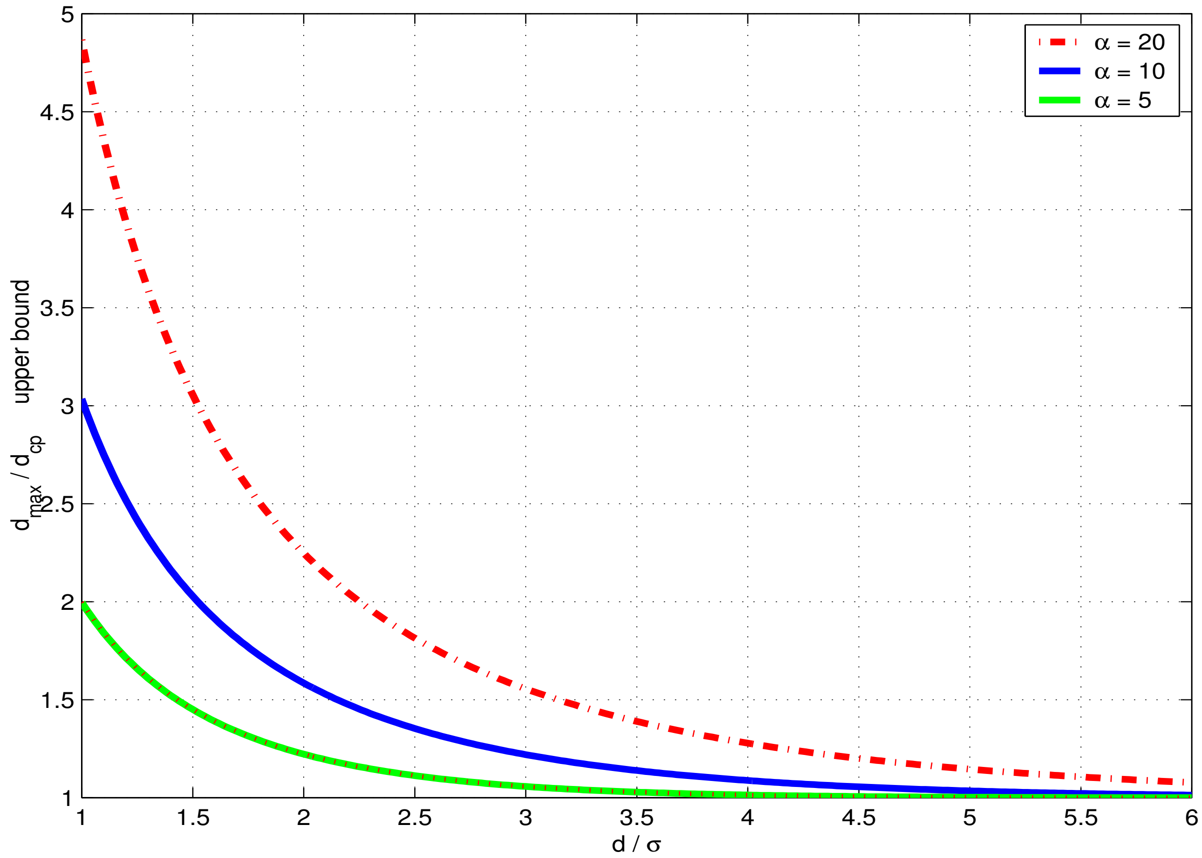


But of course Nayak's analysis tells us only about the summits of the asperities (as does GW peak-counting scheme); a group of circular summits may very quickly become a non-circular contact patch.

So we need to ask, are the summits isolated?

Nayak has the answer: he gives the number of contour areas at a given height: so we compare that number with the number of summits above that height

If each contour area contains only a single summit, the summits are isolated

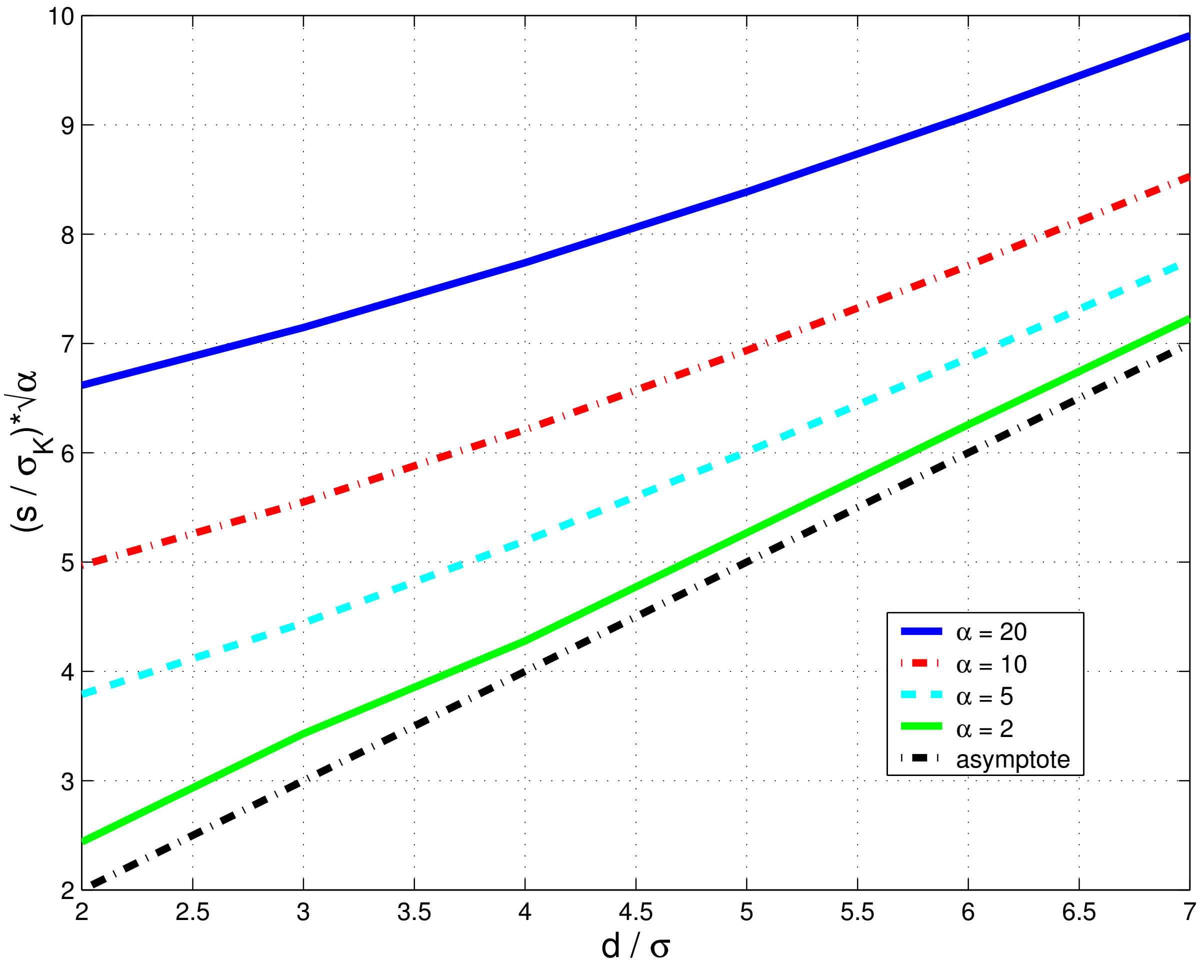


So yes, high summits are indeed isolated paraboloidal asperities

And the contact area will indeed be just half of the bearing area

… if only we could find the load so easily !

An interesting attempt to make GW as good as BGT is to include the dependence of summit curvature on summit height. If the curvature is *directly* proportional to the height, we get area proportional to load

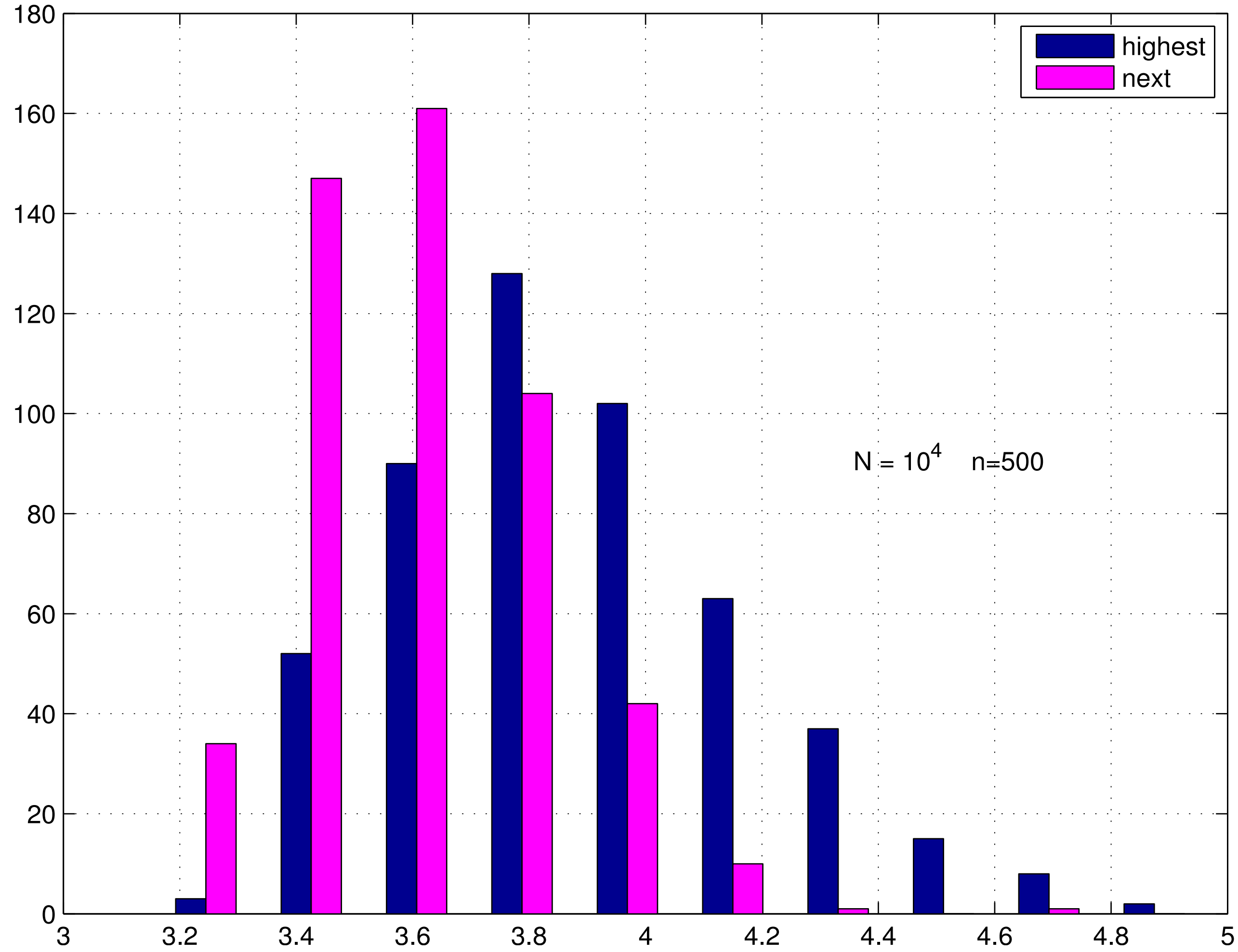


The mean summit curvature does tend asymptotically to $h/\sqrt{\alpha}$; but except for α=2, too slowly to be of use

So, sadly, it was too optimistic

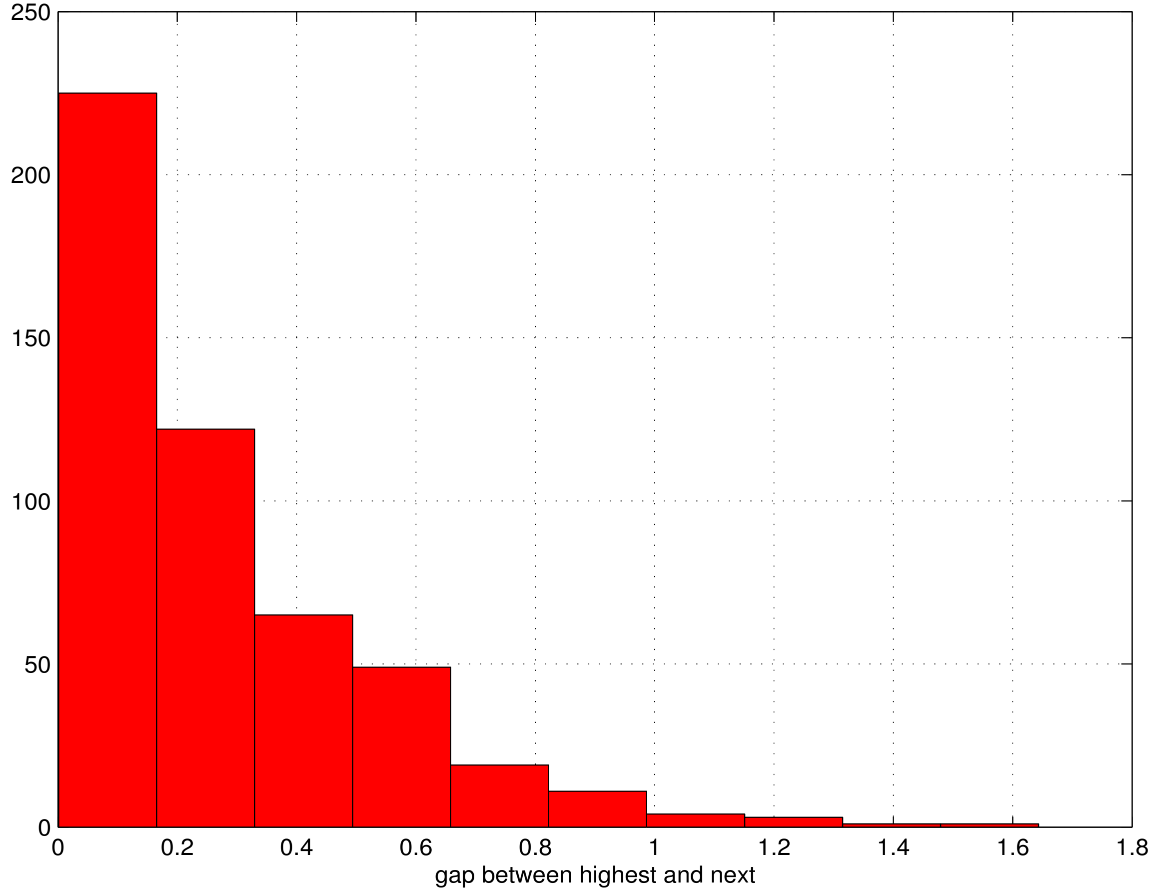
Carbone 2009 "A slightly corrected GW…

Earlier I argued that Ling was misguided to base his analysis on the point of first contact: but recently I was persuaded by O'Shea to study it. So I generated 10^4 Gaussian heights, 500 times, and investigated where contact will occur:

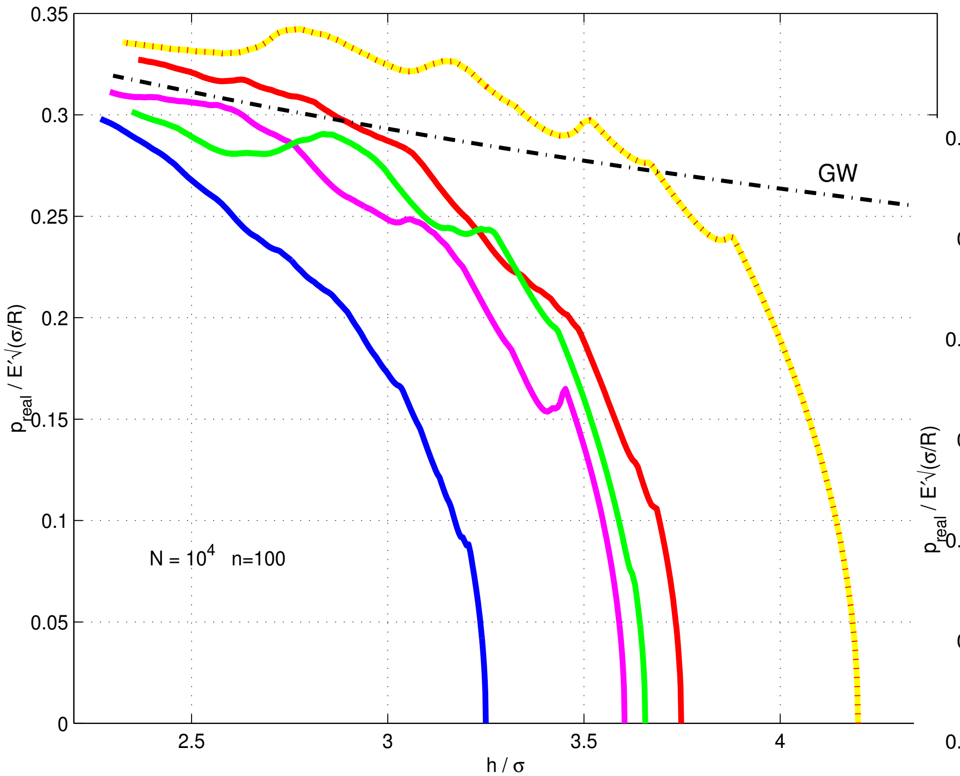


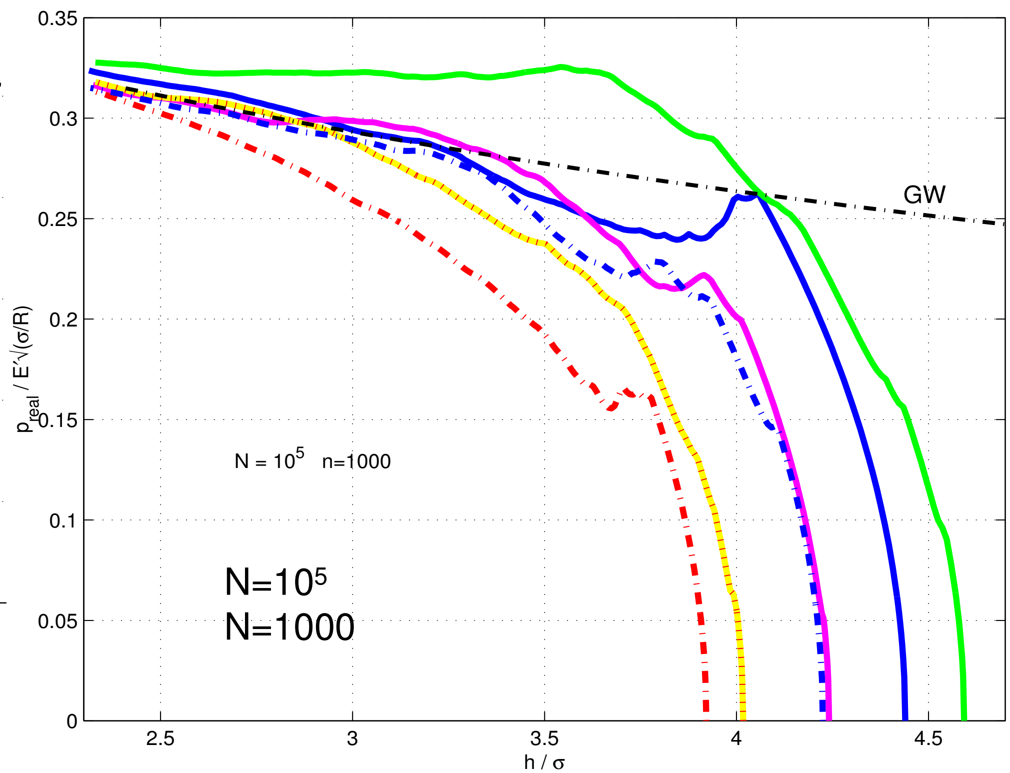
And particularly, how big the gap between the first and second contacts will be:

often a whole rms height σ.

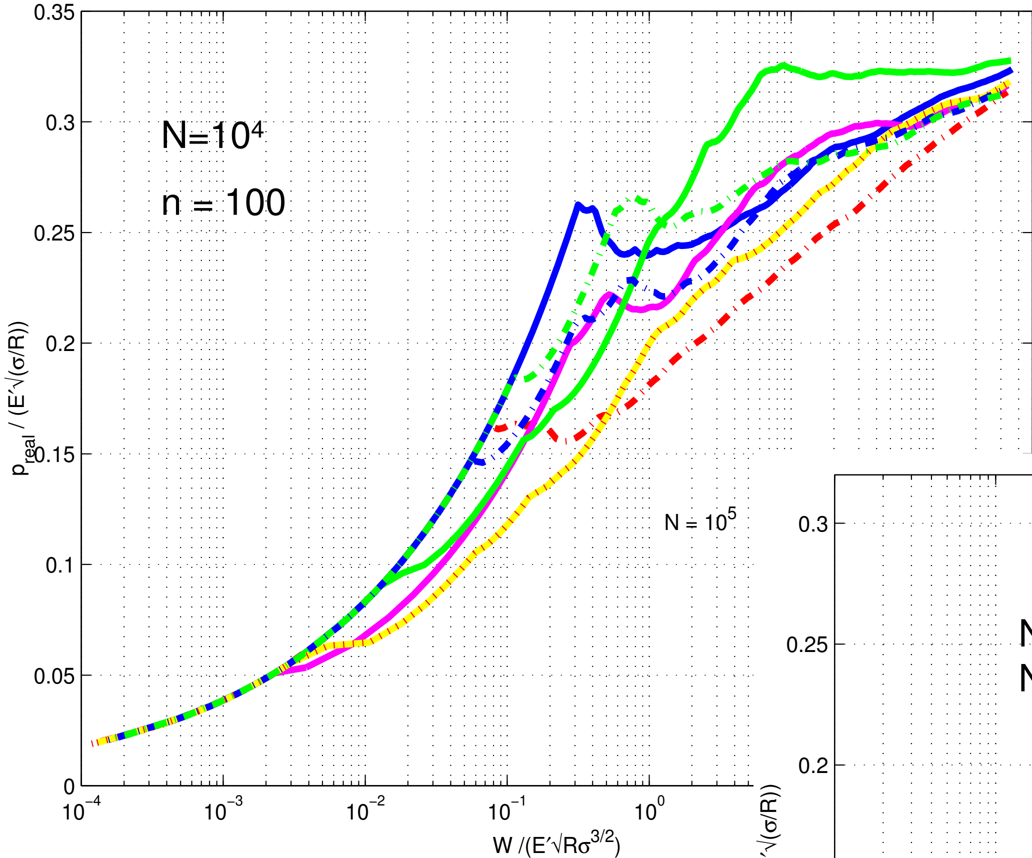


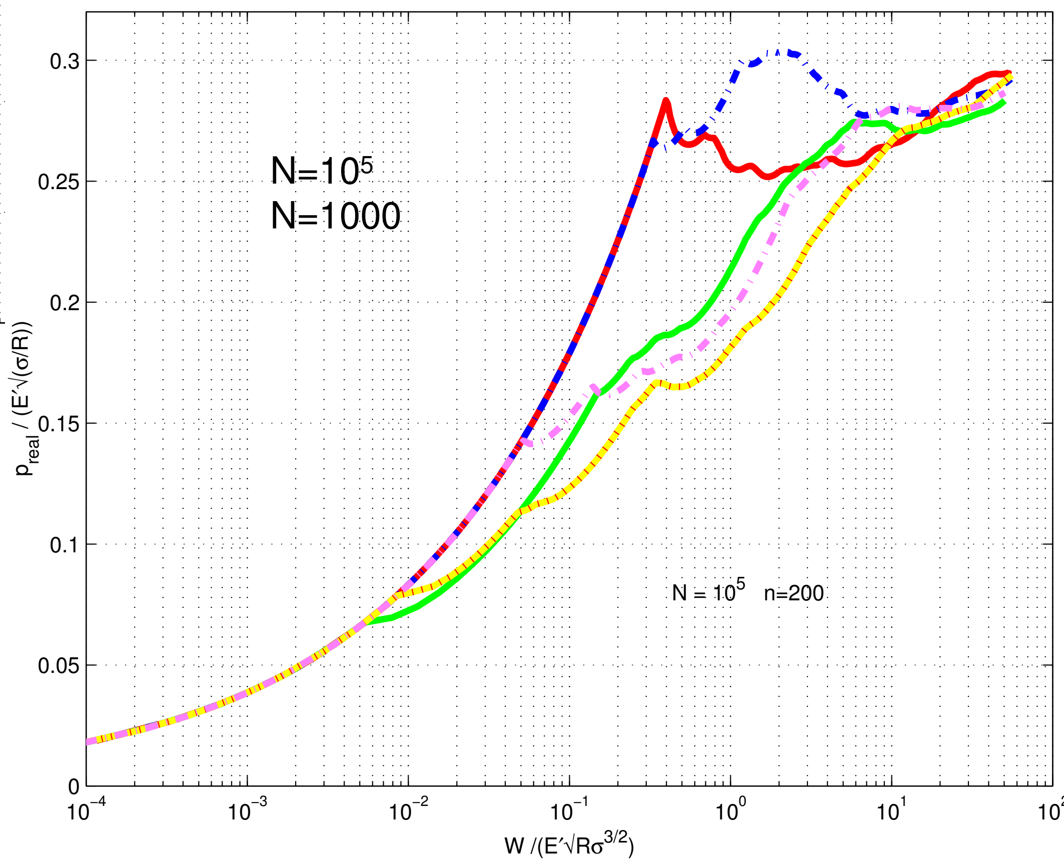
The approach for first contact varies enormously !
And with 100 contacts the scatter is still great.





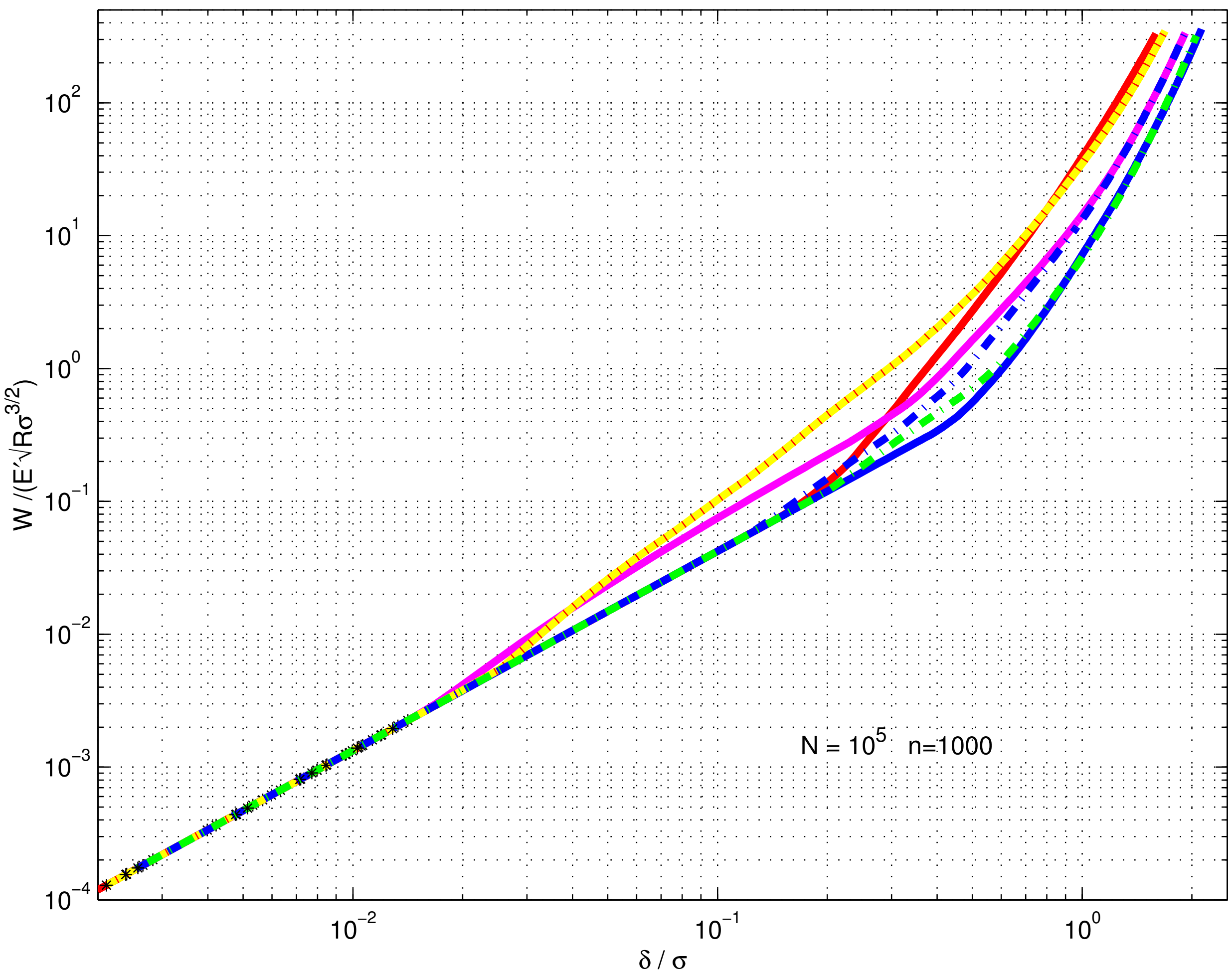
But ultimately the contact pressure begins to settle down…(to the proper value !)





Note how there may be a long delay before the second contact

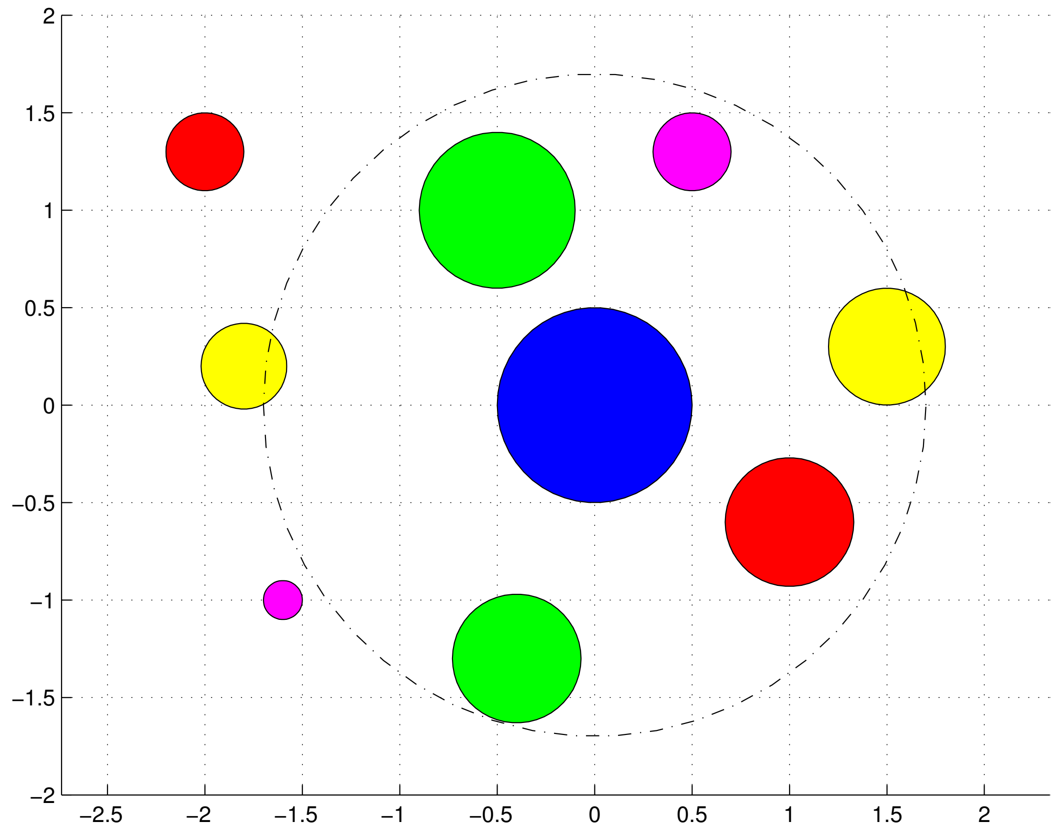
..giving just hertz behaviour $p\sim\sqrt{\delta}$



Reverting to observable quantities, the initial scatter disappears…
…but comes back later

Is the neglect of interaction by GW or BGT serious?

There is no doubt that a load on one contact spot will lower the neighbours, and may stop, or delay, them from making contact.



But is nearest neighbour interaction the real problem?

Olber's paradox: why is it dark at night?

If we live near just one of an infinite number of stars, with a density of *n* (per cubic light year?), then a spherical shell of radius R around us will contain n.(πR^2 dR) stars, each emitting light. But by the inverse square law, the illumination from a star at distance R will only be β/R^2: so the shell will contribute $n\beta/R^2.(\pi R^2 dR)$…ie $n\,\pi\beta.dR$.
So all the stars together give $n\,\pi\beta.\int dR$ ….. !

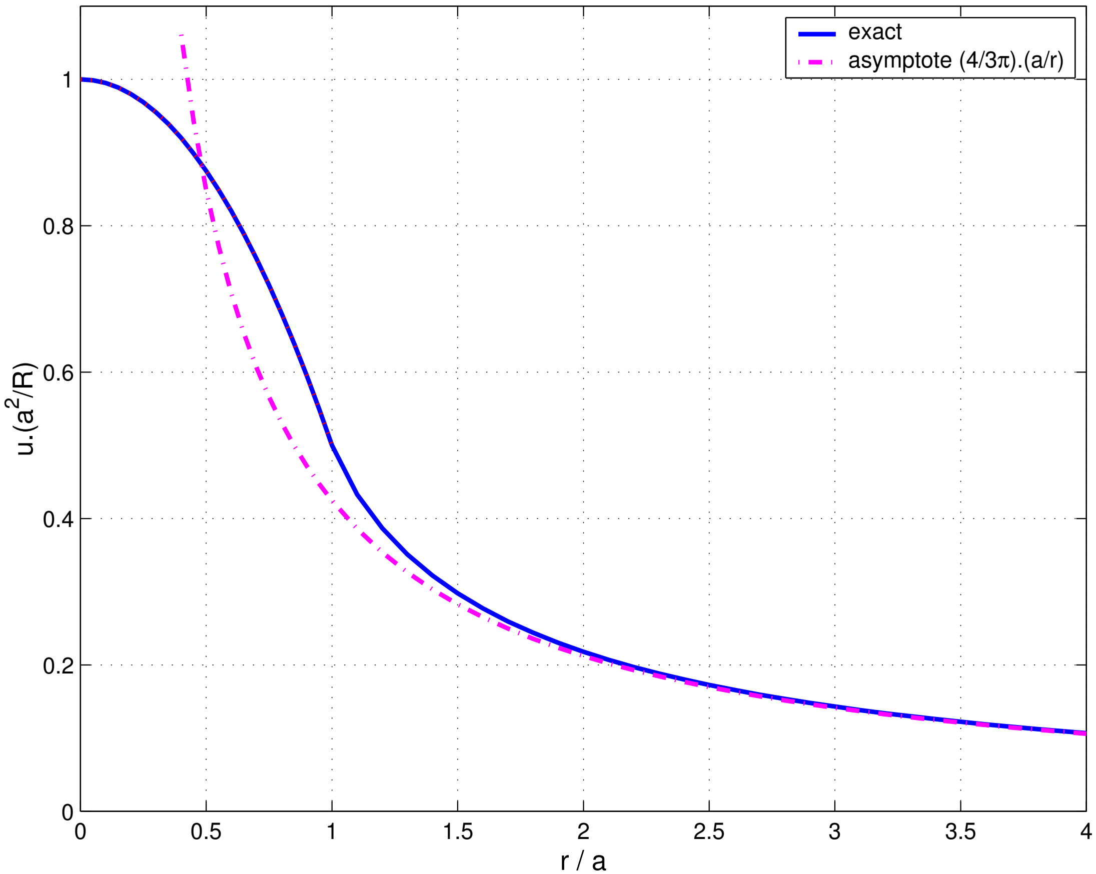
Need I go on?
A contact distant r will reduce the height by $P/\pi E^*r$. If contacts are spread over the plane with a density η (per square micron), a ring distant r will contain $\eta.(2\pi\, r\, dr)$ contacts: and lower the height by $2\eta(P/E^*)dr$.
But the load on the circle will be $\pi\, r^2 p_{nom}$: carried by $\eta\pi r^2$ contacts: so $P=p_{nom}/\eta$ and the ring of contacts will lower the height by $2\,(p_{nom}/E^*)dr$.
So the effect of the whole plane of contacts will be $2\,(p_{nom}/E^*)\int dr$….

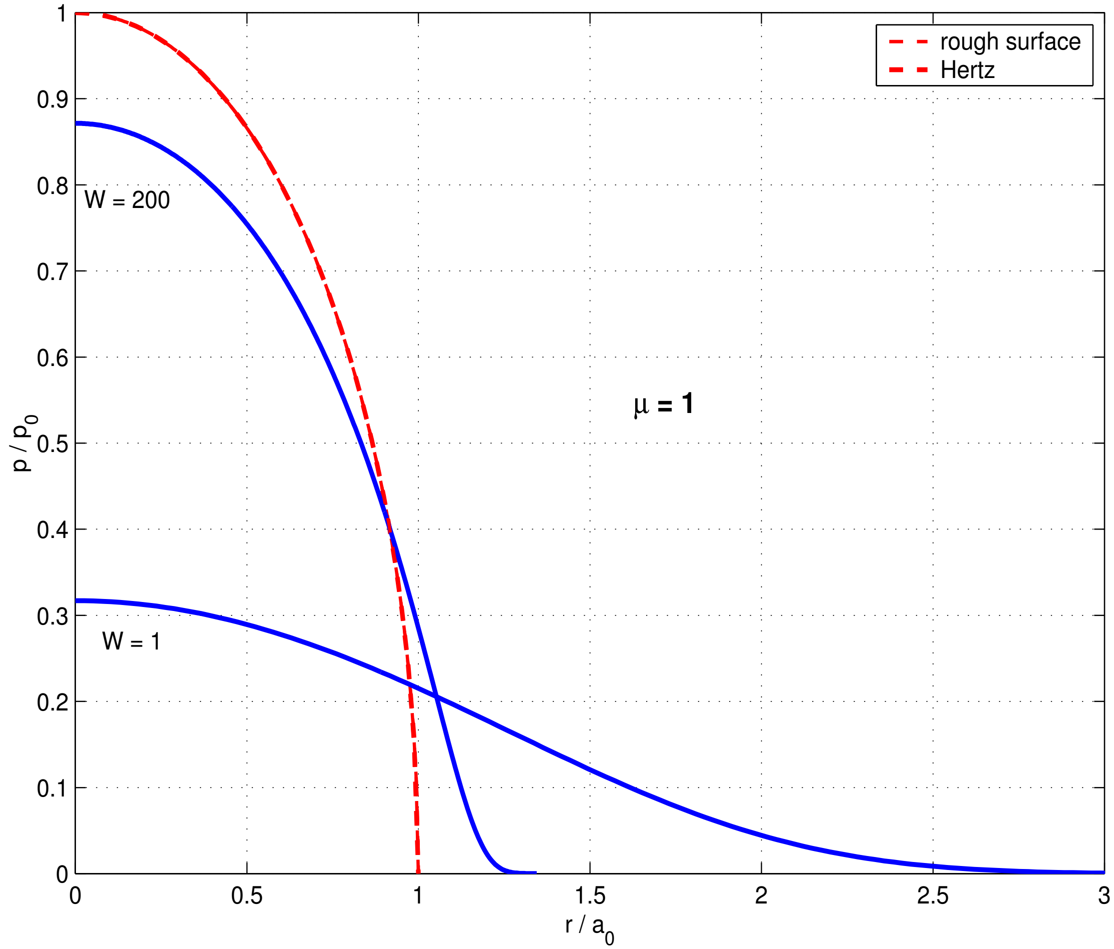
but uniform, and really just changing the datum

Its not easy to see why the effect of a ***single*** near-neighbour should matter

Hertz deformation.

The “Boussinesq” asymptote gives an excellent estimate of the interaction





$\mu=(8/3)\eta\sigma\sqrt{(2B\beta)}$

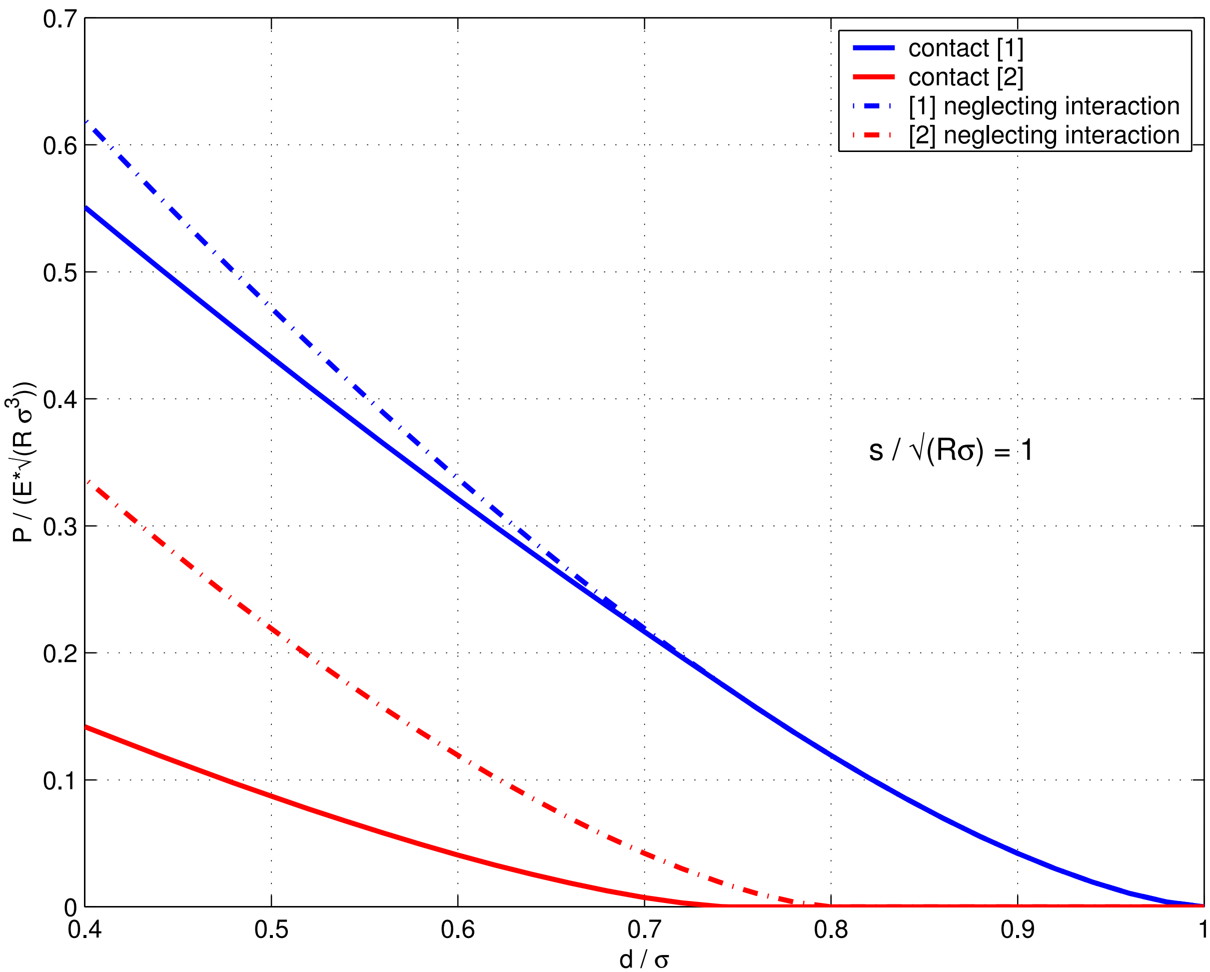
Co-operative interaction can be very real, as this study of contact between a sphere and a rough surface showed.

A Hertzian pressure distribution develops as the load increases

Greenwood & Tripp 1967

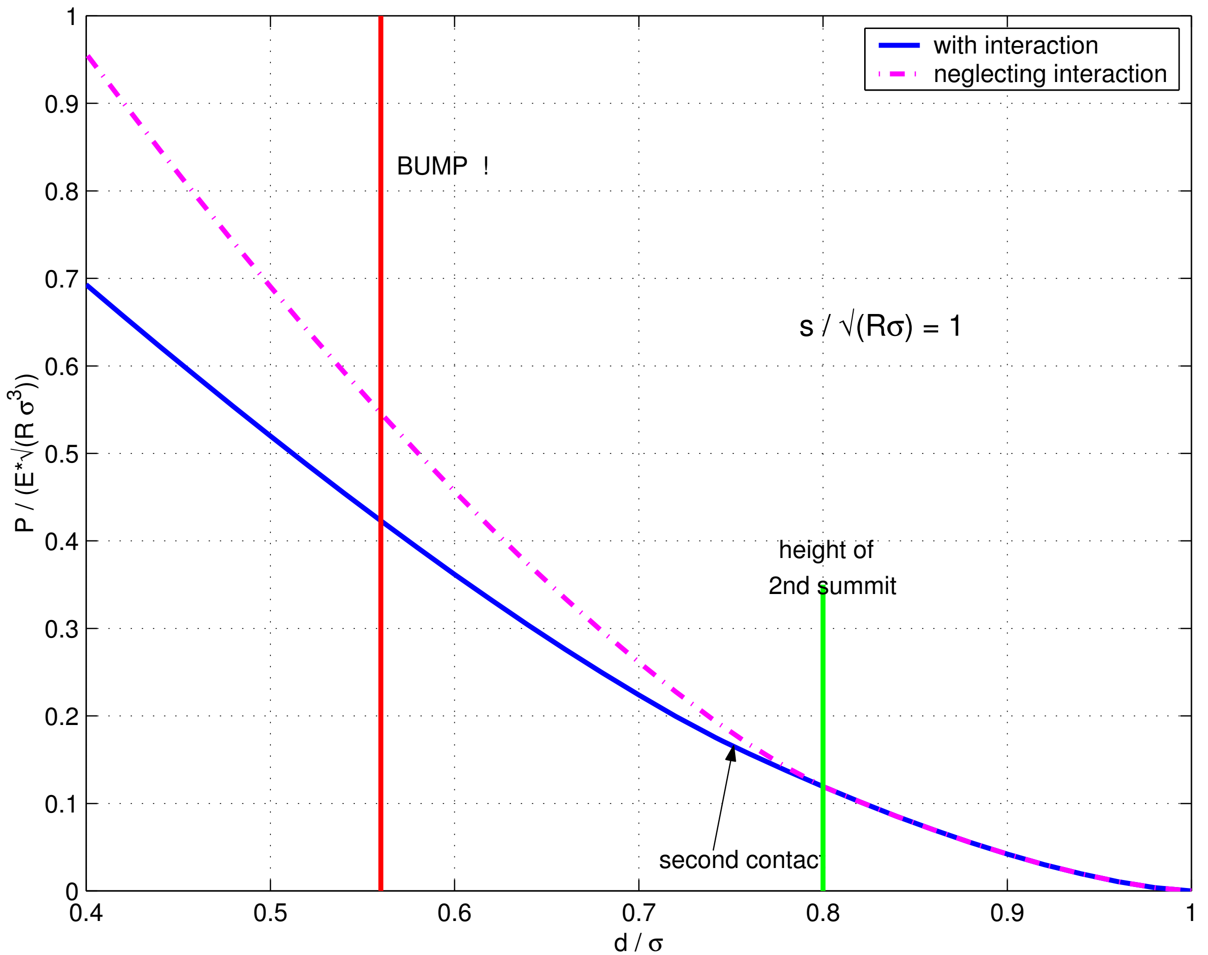
But can we really ignore local interaction ? Let's investigate :
Here is the behaviour of a pair of asperities,

height difference 0.2σ,   distance apart s = √(Rσ).



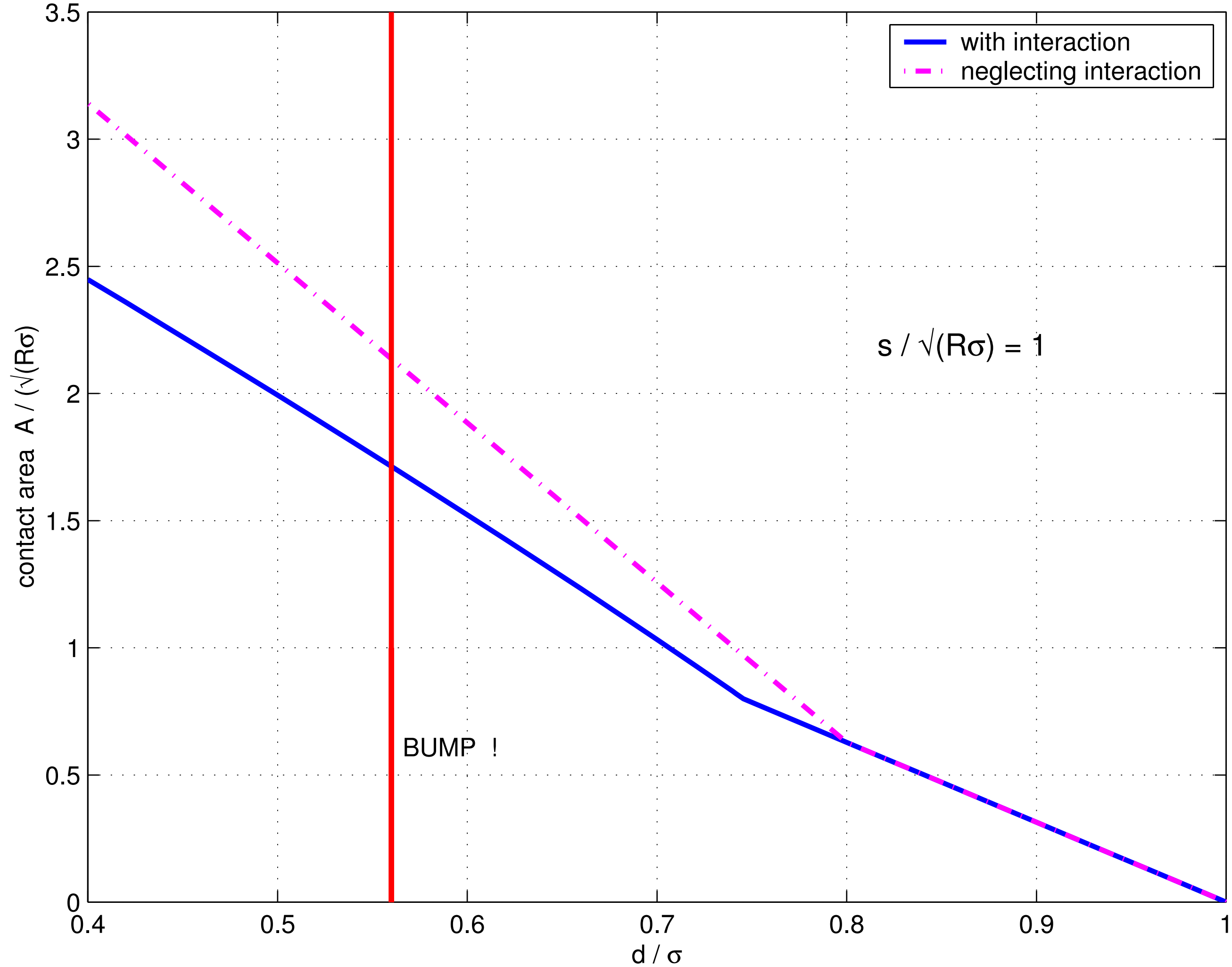
Or adding the two loads together:

Note the reduction in height of the lower one, delaying contact

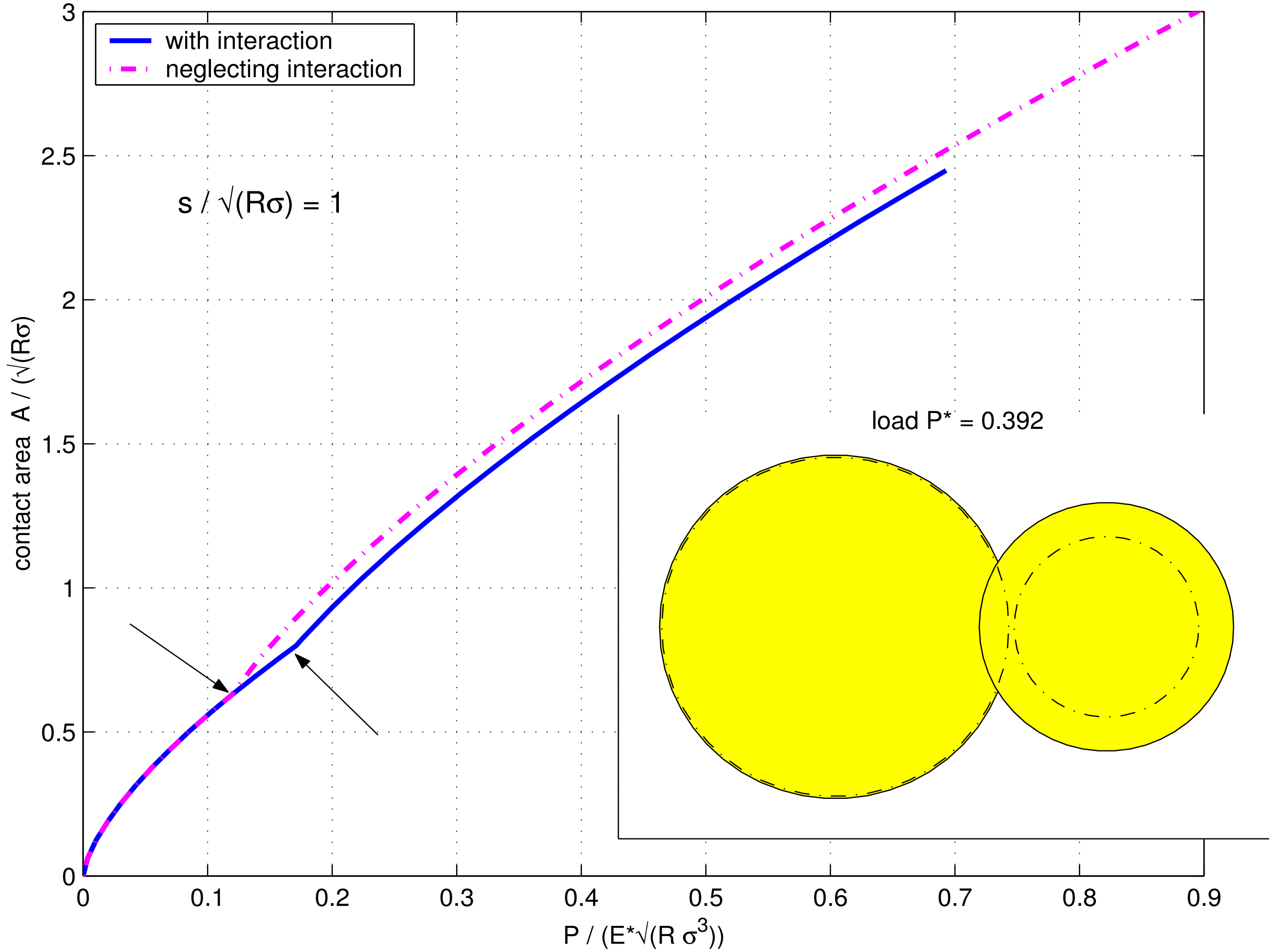


but we must stop when the two contact circles bump!

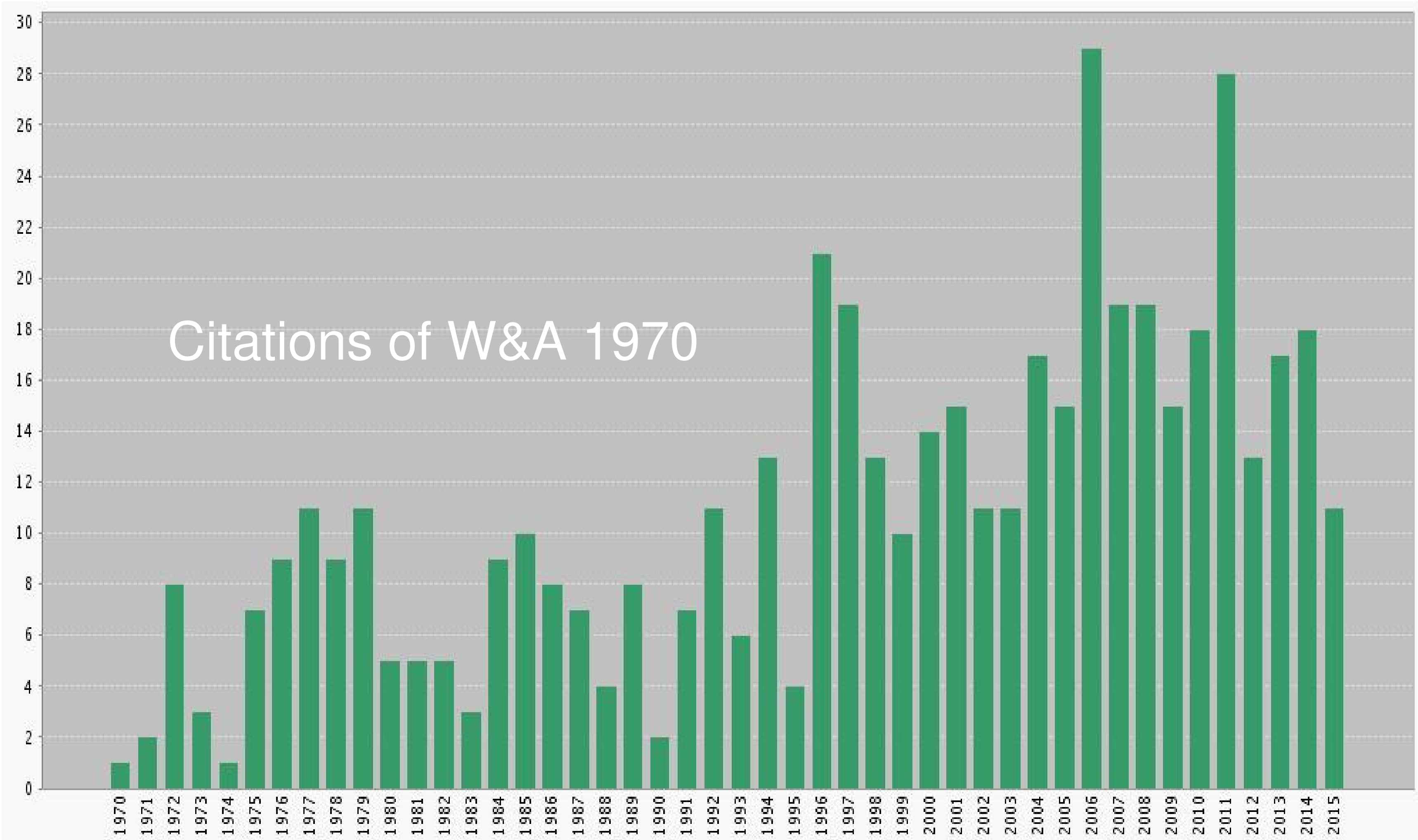
The delay before the second contact is more obvious on the area plot



But does the interaction change anything that matters?



Contact areas:
independent. **:** yellow.
with interaction **:** circles.

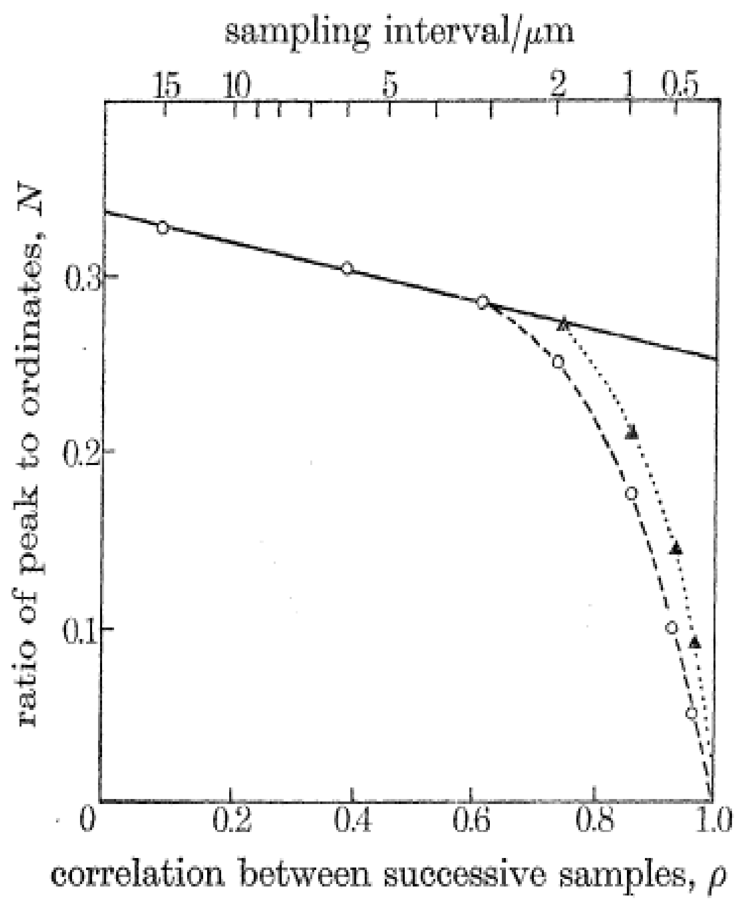


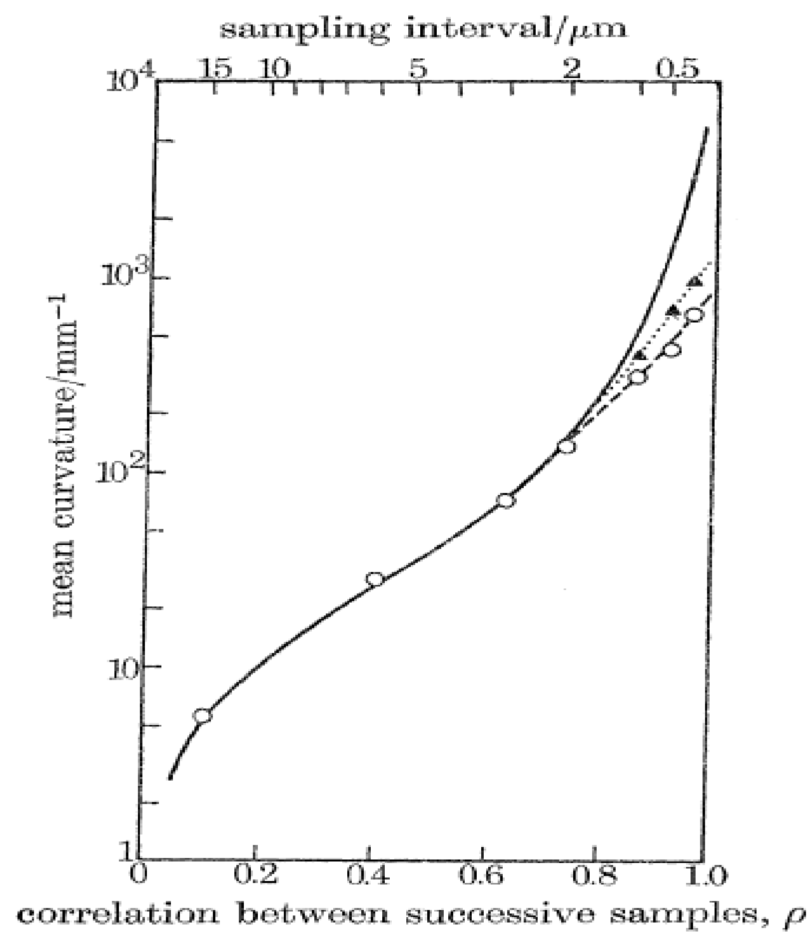
This is the paper we need to celebrate (in 5 years time). . . and it should have killed G&W dead. For W&A showed that everything depends on the sampling interval: and that all our *toys*: especially asperity density and summit curvature….and second moment $m_{2,}$ and Nayak's α, can all be anything we like: their values ***meaningless***.

So why only 500 citations ??

Whitehouse & Archard (1970) should have seen the end of GW !

For they showed it all depends on the sampling interval…with their ground surface, with an exponential autocorrelation function, between 1 in 3 and 1 in 4 of all points will be ”peaks” : and the peak curvature varies by a factor of 200





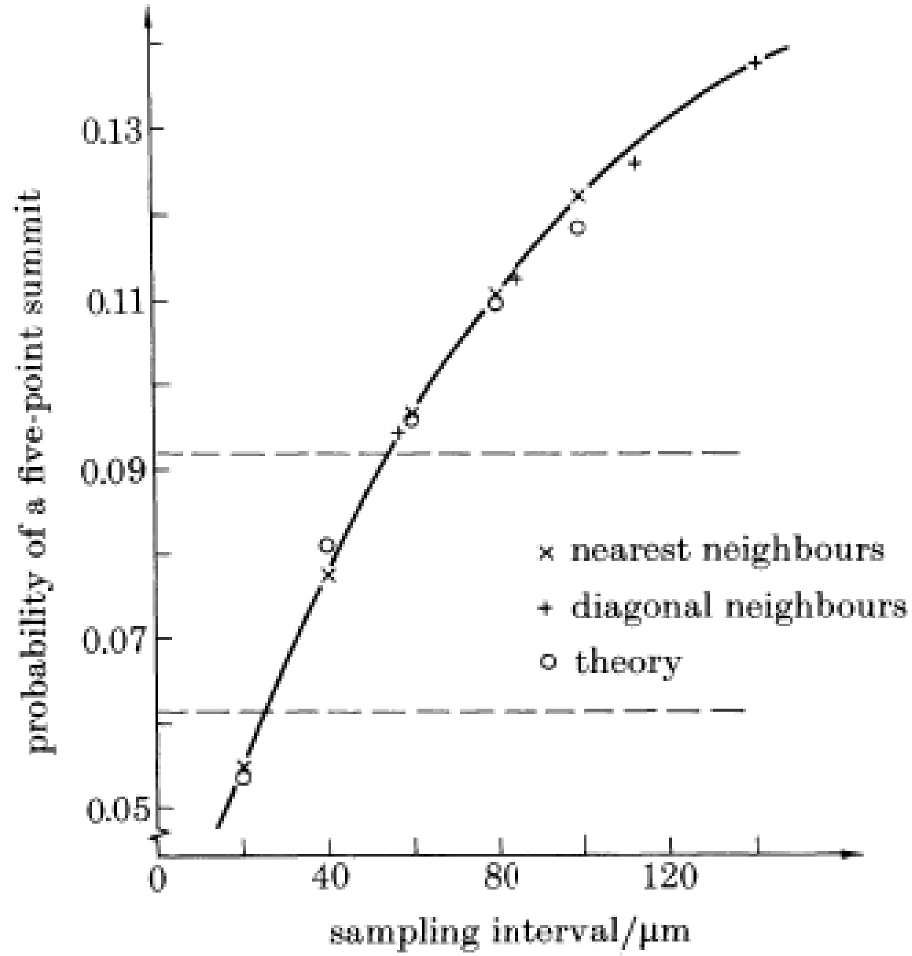


FIGURE 10. Experimental and theoretical summit densities for a grit-blasted surface (courtesy R. S. Sayles). The horizontal lines are the limits on the probability of true summits suggested by Sayles & Thomas.

Not a large variation in the summit density…until you realise this is the ***fraction*** of the sampled points which is a summit. As the sampling interval falls by a factor of 7, and the number of points / unit area increases by 50 ×**,** the summit density increased by 20 ×

So we can get nothing from measuring the surface in this way until we've chosen the relevant sampling interval.

So the spectral density approach is the better way..?

But the answers depend on m_2: the mean square profile slope. And to find that, we measure the spectral density and integrate: $m_2 = \int G(k)\, k^2\, dk$. Thus, for a power law $G(k) = A / k^p$, m_2 is infinite just as it was when we find it from the profile slopes.

We can make it finite by using a *finite sampling interval*: in the spectral density approach we do so by choosing an *arbitrary lower cut-off*.

Which do you prefer?

Ignore wavelengths shorter than $2\pi / k_1$: Sample at an interval Δ

? : ?

Only right that **Archard** should have the last word (as well as the first!).
He argued that one should worry only about the "main" structure, not about the "fine" structure.
And when you look at curve (b), would doing a G&W on those peaks (or the equivalent summits) be such a bad idea?

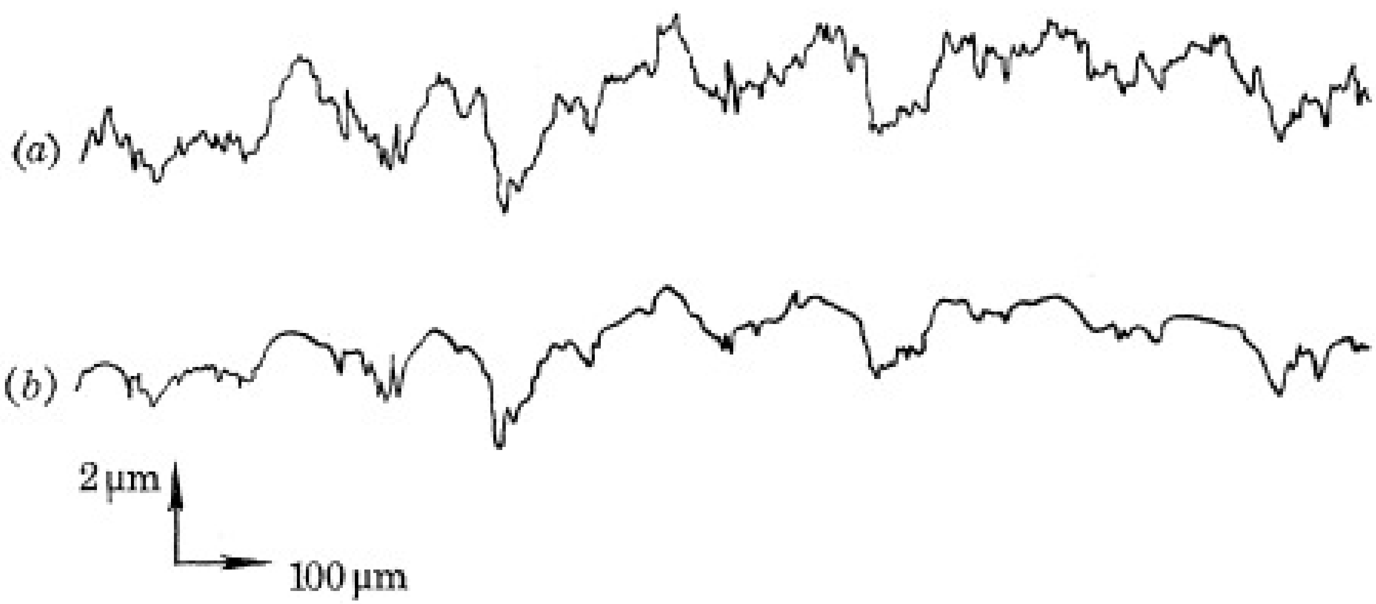


FIGURE 14. Talysurf profiles of cylindrical specimens used in lubricated friction experiments. (a) Original surface profile; (b) profile of the same line after one traversal of the load. Specimens 0.5 % C steel; 300 d.p.n.; 0.635 mm diameter; load 25 N.

Whitehouse & Archard (1970)

To sum up:

It is not clear how many engineering surfaces are Gaussian random processes: certainly not worn ones.
Then we cannot apply Nayak's results to them: so they cannot be modelled by the BGT theory.

Neither can they be modelled by McCool's ingenious adaptation of the GW theory, where the asperities are still taken to all have the same radius of curvature, and to have a Gaussian height distribution, but using Nayak's predicted values:
(summit height variance $(1 - 0.8468 / \alpha)\, m_0$; summit curvature $1.5045\sqrt{m_4}$)

And equally they cannot be modelled by Persson's theory: for while the pressure distribution for *complete* elastic contact is found using only the representation of the surface by Fourier components, his diffusion equation to find how the distribution changes as higher wave numbers are included relies on the surface being a random process.

The GW model could be used… **IF** we could make the necessary measurements of the particular surface….but how do you find enough summits to estimate the mean summit curvature? (and finding how it varies with height would be far worse)

So: pretend your surface is a Gaussian random process, ***and then select the range of wavelengths you will include?***

Measure (very tediously) the surface data, ***but choose the relevant sampling interval by some rule not yet discovered?***

ps A Carbonised Greenwood & Williamson Theory?

Suppose we first pick/guess an appropriate sampling interval: then tabulate the summit height distribution: but instead of fitting it to a Gaussian, we fit it to a *Maxwell* distribution $C\, z^2 \exp(-\tfrac{1}{2} z^2/a^2)$

(which just *happens* to be of the form of Carbone's asymptotic distribution for a Gaussian random process).

We measure the mean summit curvature κ_m: but instead of assuming all the summits have the same radius $1 / \kappa_m$, we assume the summit curvature to be proportional to its height so the radius is $R(z) = 1 / (B\, z)$

(which just *happens* to be of the form of Carbone's asymptotic distribution for a Gaussian random process).

[a will be $1.485\, \sigma_s$: and we need to take $B = 0.422\, (\kappa_m / \sigma_s)$]

Then the GW integrals become

$$P^* = (4/3)E' \int (1/\sqrt{Bz})(z-d)^{3/2}\, C\, z^2 \exp(-\tfrac{1}{2} z^2/a^2)\, dz$$

and

$$A^* = \pi \int (1/Bz)\, (z-d)\, C\, z^2 \exp(-\tfrac{1}{2} z^2/a^2)\, dz$$

and for d large we find both vary as $\exp(-\tfrac{1}{2} d^2 / a^2)$ with a constant ratio

$$p_m = 0.544 \sqrt{(\sigma_s \kappa_m)}$$

Carbone 2009

Østvik & Christensen

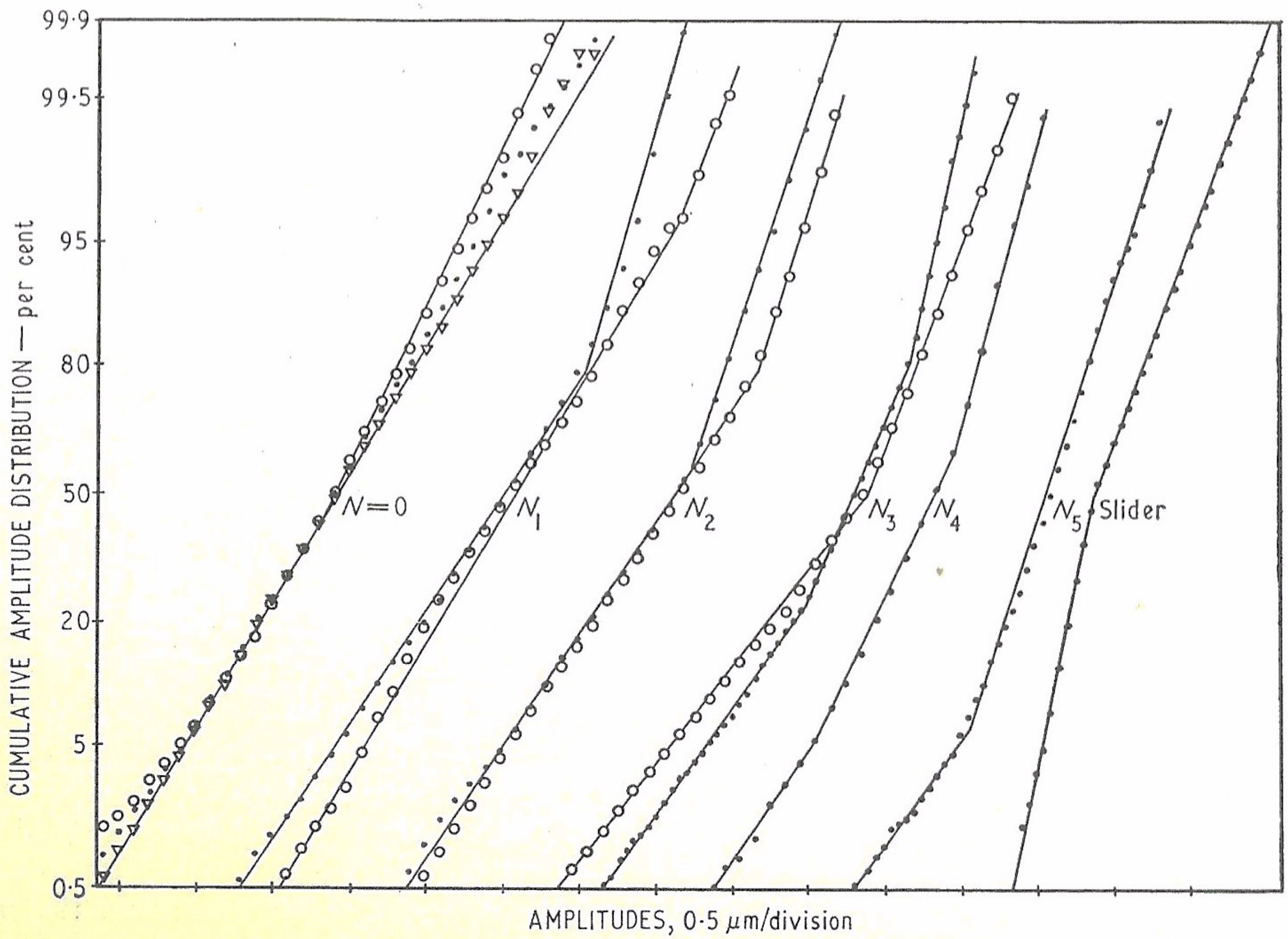
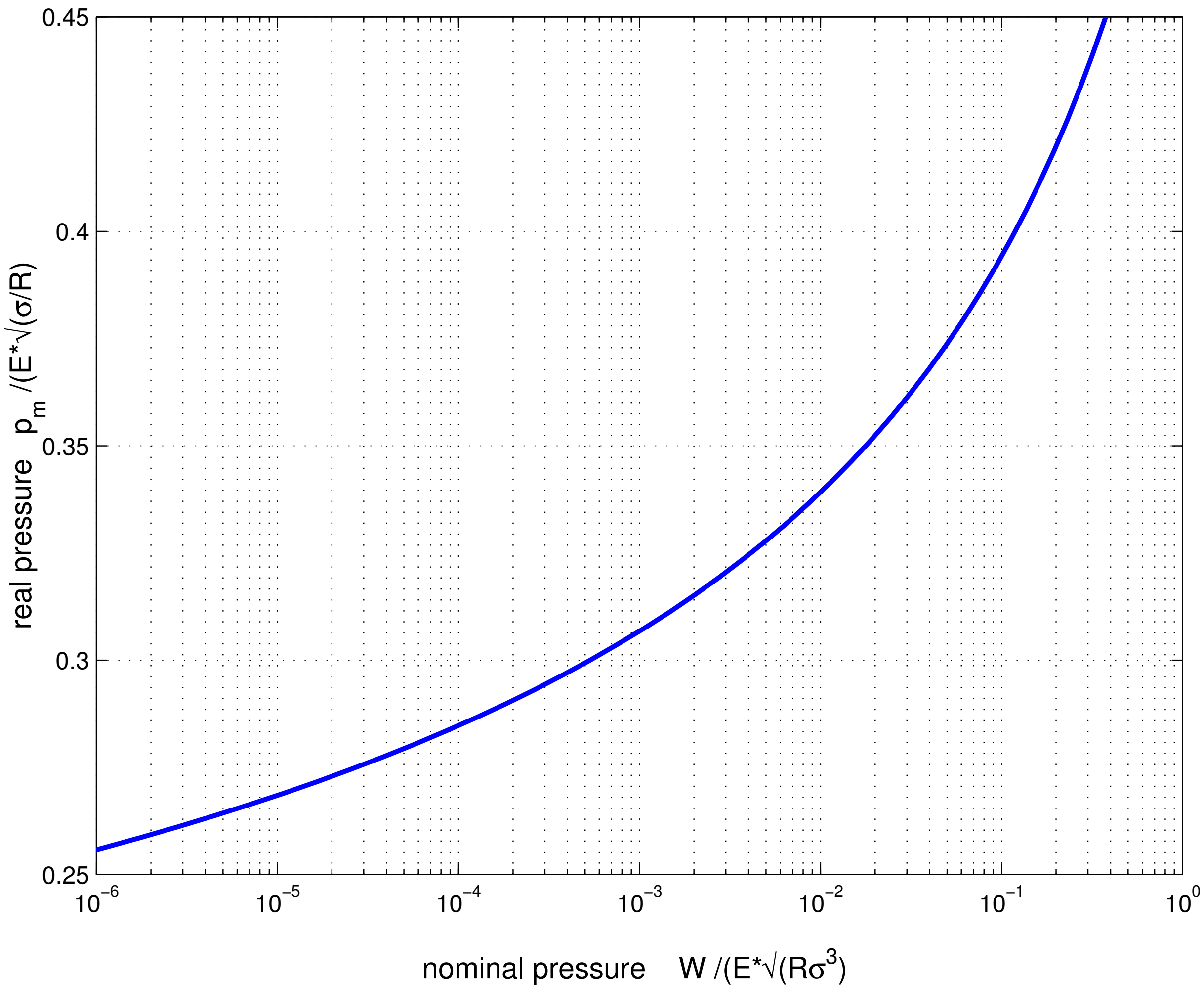
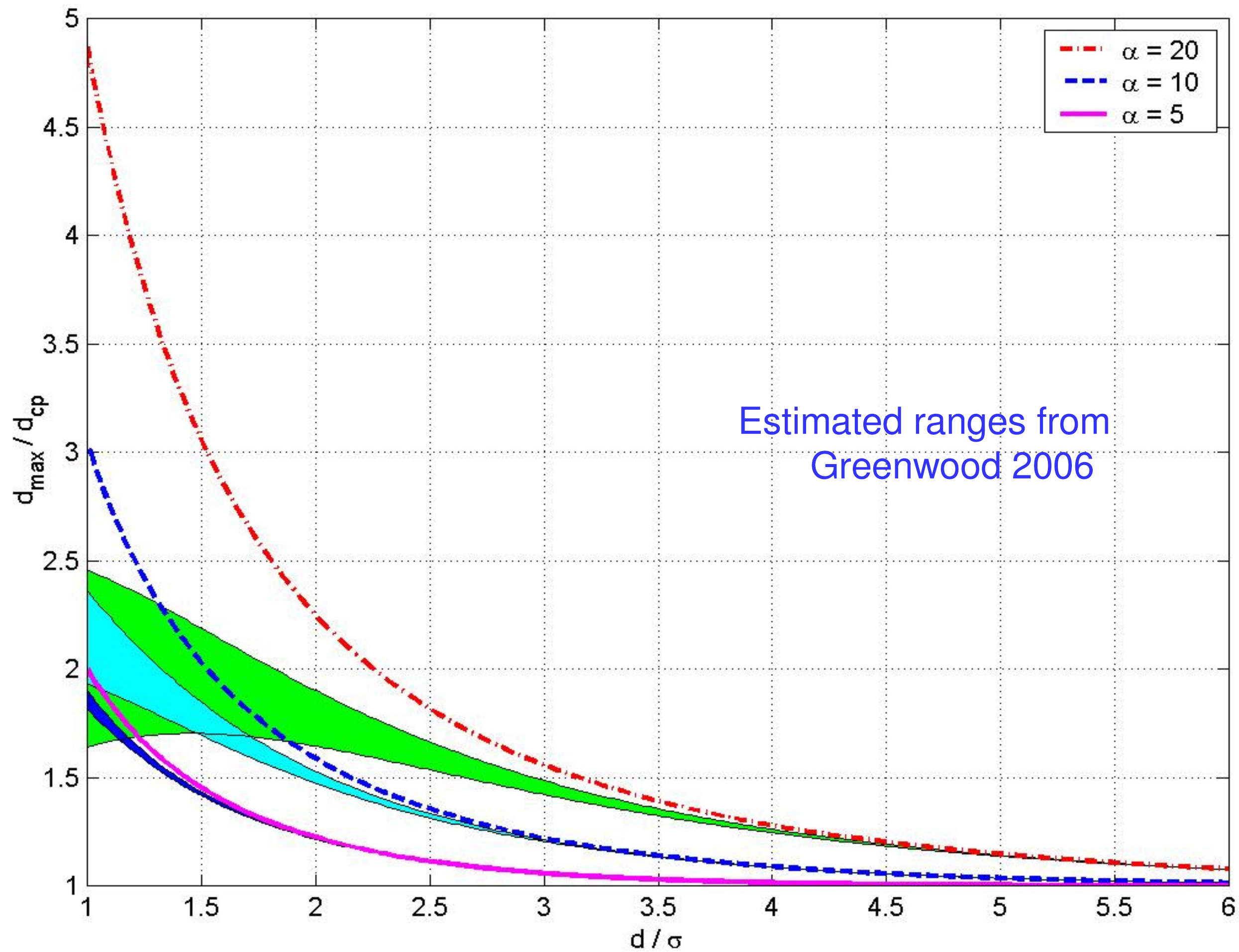


Fig. 8.3. Cumulative plot of amplitude frequency distribution in a normal distribution diagram

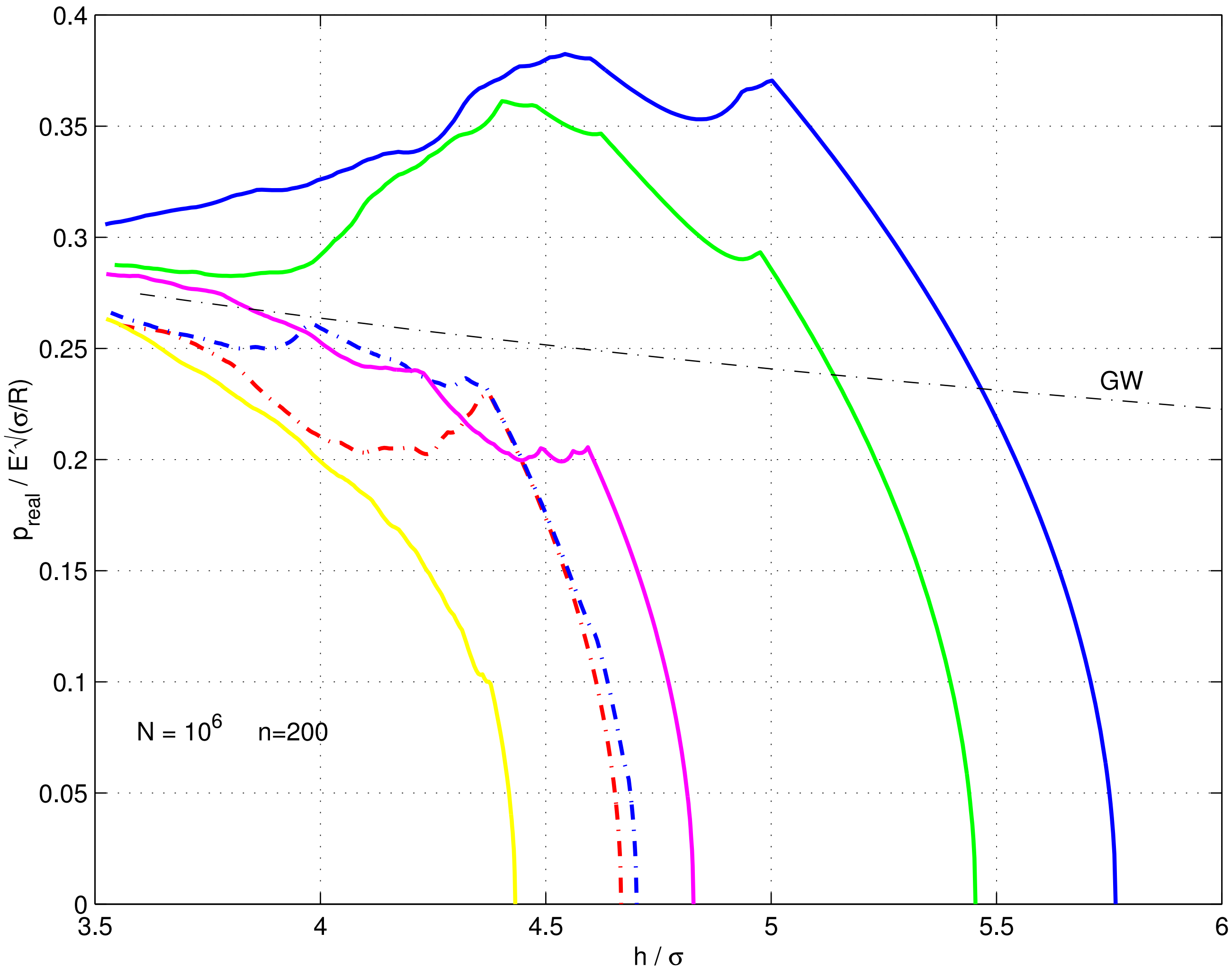


Sadly, the GW theory does not predict A ~ W….*but not far off!*

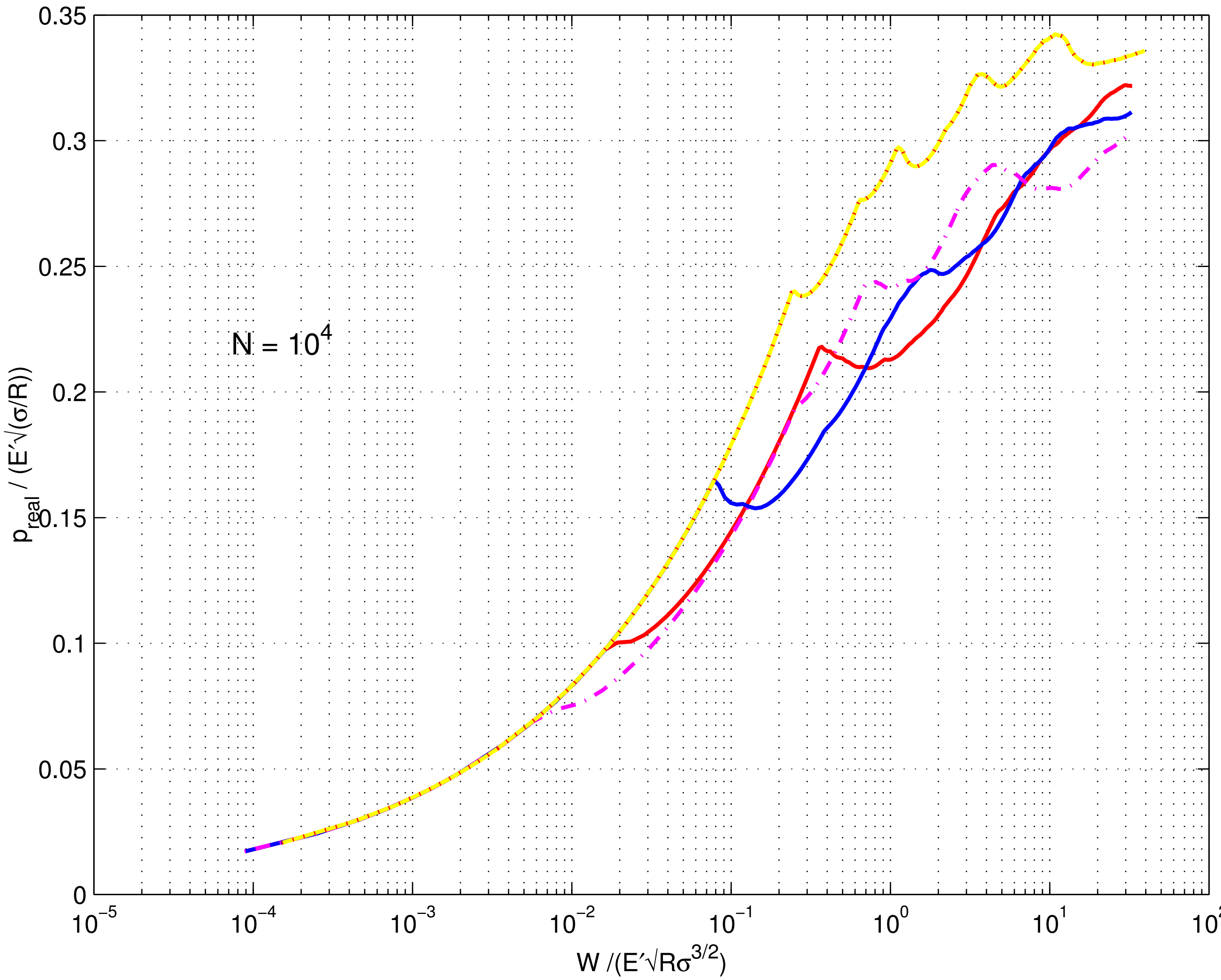
Why an upper bound? Sadly there are *clockwise* and *anticlockwise* contours: corresponding to *islands* and *lakes*: and what we know is the difference



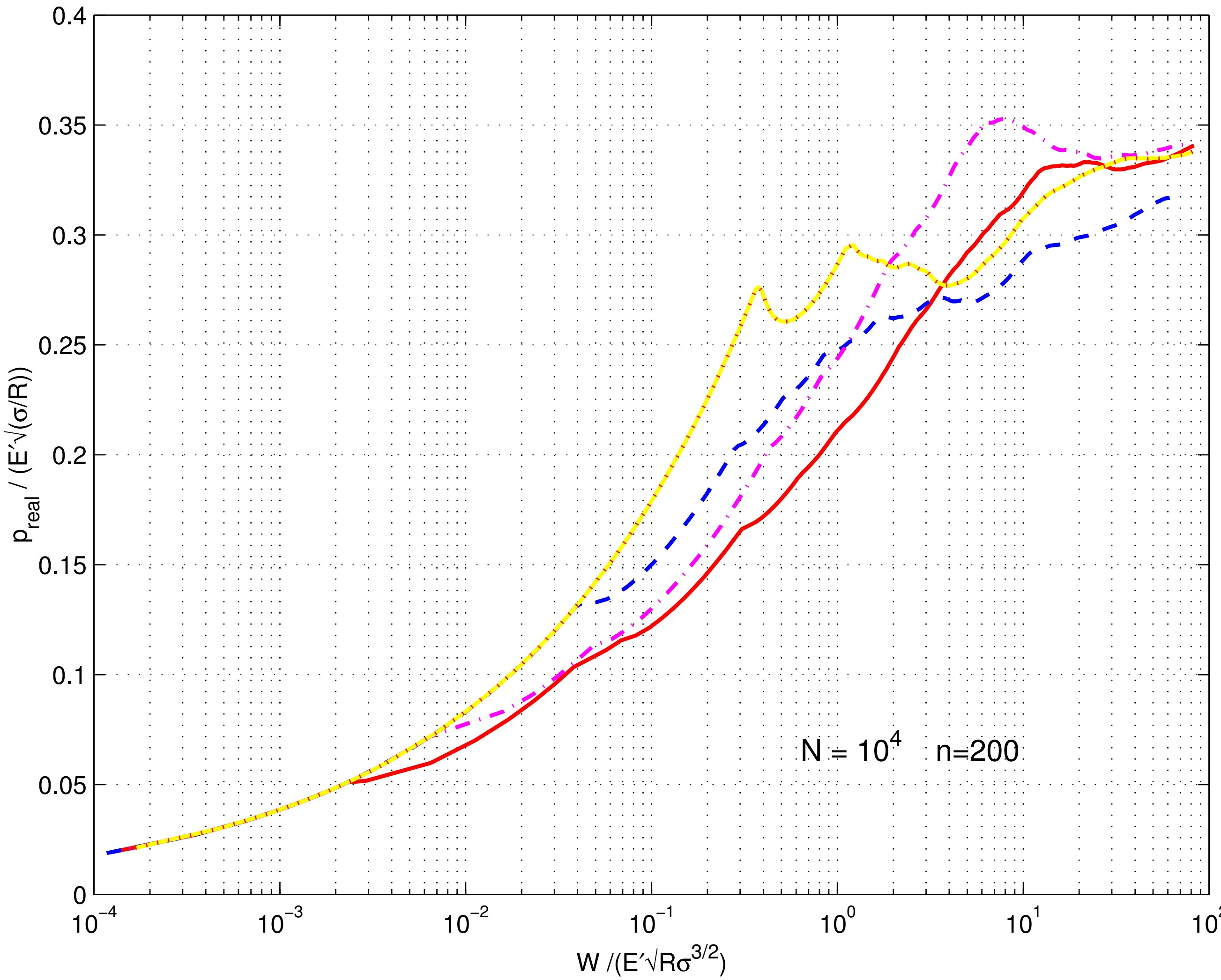
And if every contact is an isolated circular Hertzian contact, so that at each the real contact area is half the bearing area, the same is true for the totals. Which leads to Ciavarella's idea**:** forget all about asperities: just use the Abbot bearing area curve and double it



0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
3.5
4
4.5
5
5.5
6
p_{real} / E′√(σ/R)
h / σ
GW
N = 10^6 n=200



0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
N = 10^4
p_real / (E′√(σ/R))
10^−5
10^−4
10^−3
10^−2
10^−1
10^0
10^1
10^2
W /(E′√Rσ^3/2)



0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
p_{real} / (E′√(σ/R))
$N = 10^4$ n=200
10^{-4} 10^{-3} 10^{-2} 10^{-1} 10^{0} 10^{1} 10^{2}
W /(E′√Rσ$^{3/2}$)