

Guidance Notes on Lubricated Wear Testing

© 2005 George Plint

Introduction

Friction and wear are not intrinsic material properties but are properties of the system in which the materials operate.

It follows that the properties measured in an experiment using a test machine are also a system property, a property of the materials and the test machine in which the experiment is run.

The key problem is how to design experiments that allow us to predict how the materials will behave in systems other than our test machine.

Lubricated Contacts and Wear

When modelling wear and failure mechanisms involving lubricated contacts, we are most usually concerned with modelling real systems such as the contacts in mechanisms, gear-boxes, engines, bearings and machining and forming processes.

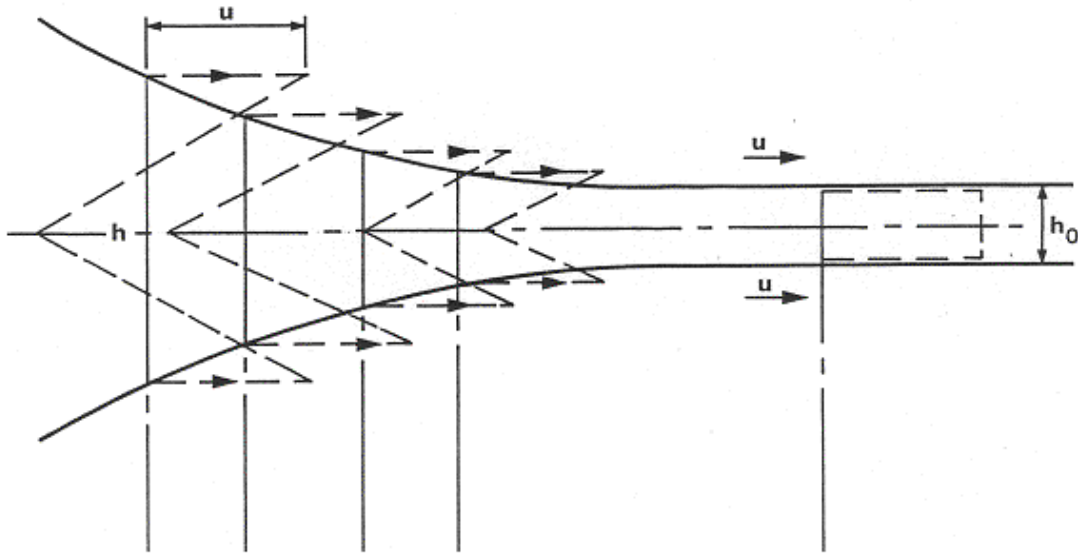
The special difficulty with lubricated tests is that in a lubricated contact we have a complex system, the performance of which depends not only on the contacting materials, the contact geometry and conditions of load and relative motion, but also the lubricant additive chemistry. So we are faced simultaneously with issues relating to engineering, materials and chemistry.

It is perhaps worth starting by reviewing those processes in which we would not expect wear to occur and then identifying examples where wear and failure does occur. We can then consider how these can best be modelled.

Providing we have a clean lubricant, free of abrasive particles or wear debris, with a contact operating under fully hydrodynamic or elastohydrodynamic conditions and providing complete separation of the material surfaces, we should not expect to generate any wear. If all types of tribological contact operated with hydrodynamic or elastohydrodynamic lubrication, then we would not have any problem with wear, we would need little in terms of lubricant additive chemistry, other than the provision of additives to provide oxidation and shear stability and we could use any moderately viscous fluid. If this were the case, we could lubricate our systems with glycerine instead of oil. This clearly is not the case.

The problem is that in order to generate hydrodynamic separation of the surfaces, any system must fulfil two requirements. Firstly, there must be a mechanism for delivering the lubricant to where it is required and in sufficient quantity to flood the inlet to the contact. Secondly there must be a sufficient entrainment velocity to carry the lubricant into the contact in order for it to generate the necessary hydrodynamic film. This latter condition, except in the case of externally pressurized bearings, by definition requires the contacting surfaces to be in motion with at least one surface moving from the direction of lubricant supply and at sufficient speed to generate a thick enough hydrodynamic film.

The following sketch illustrates what happens in an elastohydrodynamic contact between two rollers running with the same surface velocity with a fully flooded inlet.



The Entrainment (or Rolling) Velocity is defined as follows:

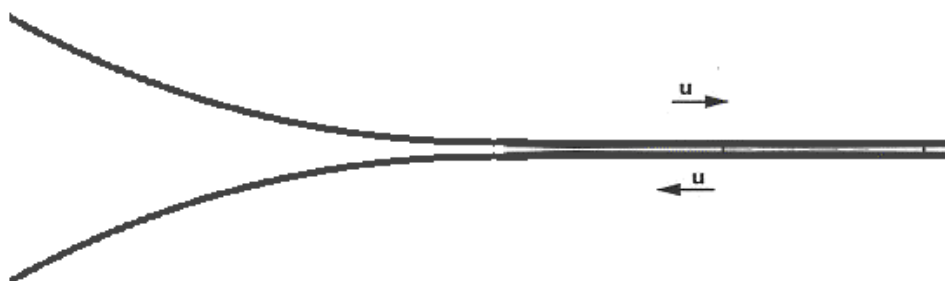
$$\text{Entrainment Velocity} = \frac{1}{2} (U_1 + U_2)$$

Where U_1 and U_2 are the velocities of the two contacting surfaces.

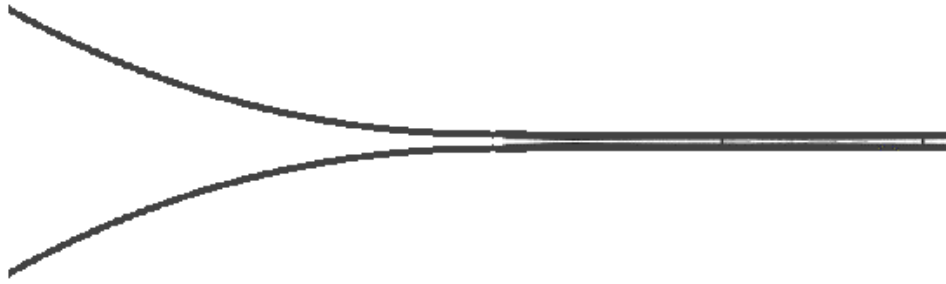
It is important to note that the system is surprisingly very weakly dependent on load, so that once a hydrodynamic film of sufficient thickness to separate the surfaces is generated, increasing the load has little impact on the level of separation.

In an ideal world, we would like to achieve this kind of lubrication regime, but with many real mechanisms and in particular mechanisms starting from rest, we are unable to do this.

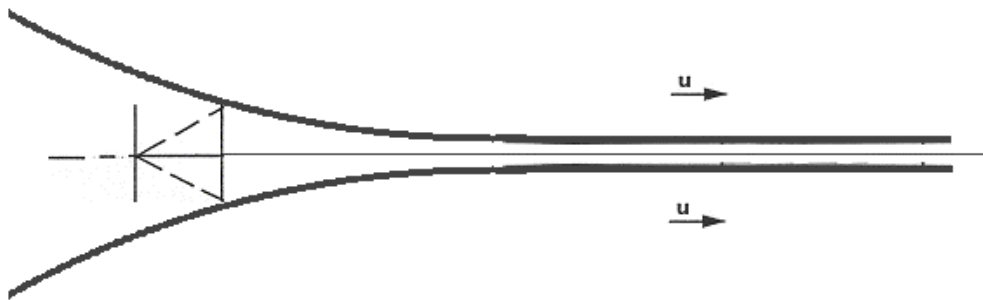
The following shows what happens if the surfaces are sliding in opposite directions. This is called the zero entrainment condition:



This is what happens when the surfaces are at rest. This is the start-up condition:



This is what happens when there is insufficient lubricant to flood the inlet. This is what is known as starved lubrication:



In the cases where we cannot prevent intimate contact between the mating surfaces we have to rely on lubricant additive chemistry to limit wear of the surfaces and to provide protection against scuffing. This is the partial welding of the contacting surfaces through the localized dissipation of frictional energy. It follows that, excluding the special case of rolling contact fatigue, which will be covered later in this presentation, all lubricated wear testing is focused on modelling contacts that are operating outside the hydrodynamic or elastohydrodynamic regimes. Under these conditions, lubricant viscosity plays little or no part and the performance of the contact depends critically on the lubricant additive chemistry.

Examples of such contacts include:

- Journal bearings at start-up or during starved lubrication because of failure of supply.

- Elastohydrodynamic bearings at start-up or during starved lubrication because of failure of supply.

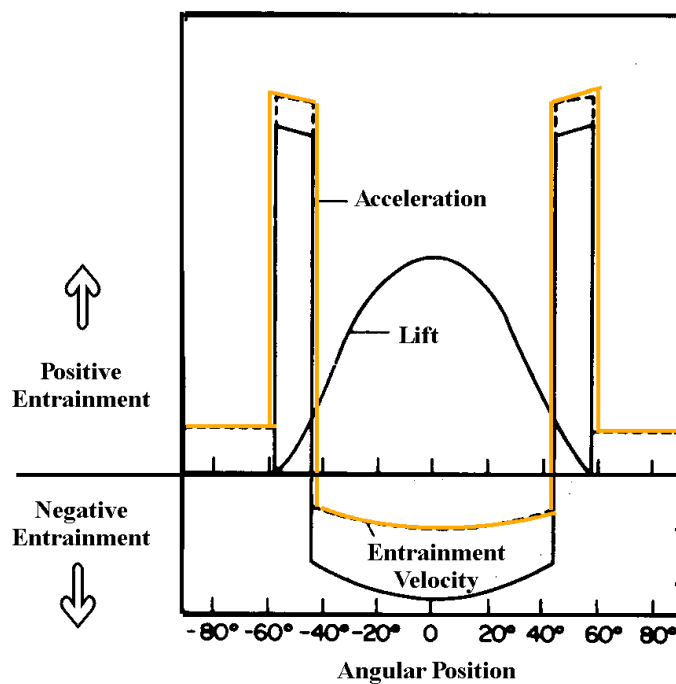
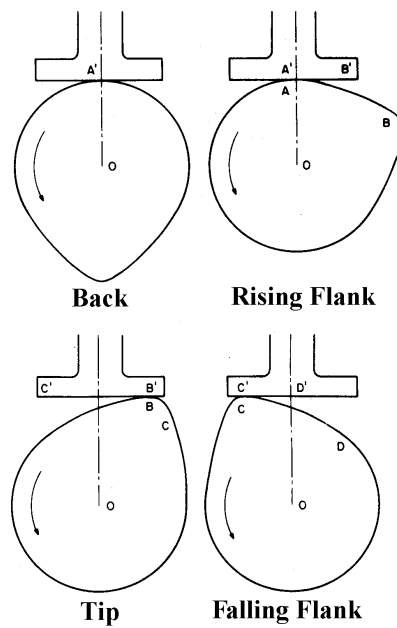
- Piston ring and liner contacts at bottom and top dead centre, where entrainment velocities fall to zero.

Piston ring and liner during running because of starved lubrication.

Gear contacts during running because of negative entrainment.

Cam and tappet or finger follower contacts during running because of negative entrainment.

As a brief example of the issue of negative entrainment consider the case of a simple cam and tappet:



Finally, you will appreciate why wear in the system is frequently associated with start-up conditions, in particular from those contacts that, when running, produce hydrodynamically separated surfaces.

Lubricant Additive Chemistry

Before considering how to model these and similar contacts in a bench test, we need to concentrate a bit on what the lubricant additives do within our non-hydrodynamically lubricated contacts.

There are two basic forms of additive protection available to surfaces that are either temporarily or permanently in intimate contact, physisorped and chemically reacted. The former, associated with the migration of polar molecules to the active metal sites on the surface, typically occurs at contact temperatures in the range of 70 to 150 C and give rise to the “hairy molecule” surface model.

The chemically reacted layers are associated with actual chemical reaction between additives and the active metal sites on the surface. These are frequently extreme pressure (or anti-scuffing) additives and are usually activated at temperature above 170 C. With conventional lubricant additive packages, these layers frequently fail at contact temperatures above 240 C.

Changes in running conditions or test parameters, for example changing load or frequency of events, will usually cause a temporary change in the consistency of these protective layers or films, indicating some kind of limited dynamic stability. The late Professor Cameron (and others) postulated that these films were partially destroyed and reformed each cycle. In other words, each specimen pass knocked the top off a number of surface asperities creating new active metal sites. The dwell time between passes allowed time for the chemistry to work, so that the film, chemical or physical, reformed before the next specimen pass. There is no way of course of telling how many asperities could be scraped off in this way, although it must be related to wear.

The rate of formation of the chemically reacted films in particular is considered to be a direct function of the contact temperature, following the Arrhenius equation, and most importantly the repetition rate. In other words, one needs temperature for activation and for controlling the rate of reaction plus a finite time for the chemistry to take effect.

The Arrhenius Equation is:

$$k = A \cdot \exp^{(-E/R \cdot T)}$$

Where:

k	=	rate coefficient
A	=	constant
E	=	activation energy
R	=	universal gas constant
T	=	temperature

Common sense and chemical intuition suggest that the higher the temperature, the faster a given chemical reaction will proceed. For tribologists, it is important simply to remember that this is an exponential relationship with reaction rates accelerating with increasing temperature.

Evidence of the process of continuous regeneration of the chemically reacted films and the time dependence issue has been reported in respect of gear contacts. At very high speeds there has been some suggestion that insufficient time elapses between the successive engagements of matching teeth for the chemistry to work, thus leaving the surfaces unprotected, giving rise to scuffing failure. However, it may be that the actual cause of this type of film failure is desorption of the reacted film caused by excessive contact temperature.

If we were able to run a test in which we ramped the contact temperature from ambient using a fully formulated mineral automotive lubricant, we would expect to observe somewhere around 70 C, evidence of formation a physisorped film and this would typically not survive at temperatures above about 150 C. There may then be a region of tribological distress, sometimes known as the Temperature Distress Gap, until chemically reacted films start to form, typically at temperatures of about 170 C. These tend to fail catastrophically at temperatures in the region of 240 C, which is a typically contact temperature at which scuffing occurs in cams and gears.

It is of course worth noting that every time you start, even with a run-in surface, there will be a degree of re-running-in. This would tie in with the observed occurrence of high wear during engine start up.

Planning an Experiment

Before attempting to design an experiment, we must ensure that we have properly analysed and understood the tribological conditions to be modelled. The contacting environment is usually defined in terms of:

- Contact pressure
- Contact speed
- Energy input
- Temperature
- Conditions of lubrication

None of these are completely straightforward to define, either for the practical contact or in the test machine model and they are not sufficient to define fully the nature of a tribological contact.

The principal contact conditions are defined by the component (shape, surface condition etc), some by the operating conditions (load, sliding speed) but others are less easy to define (lubrication conditions, contact pressure, wear mechanism) as these can vary during a machine cycle and as the components wear.

The contact pressure (loading stress in the contact) is defined by the geometry of the contact, the applied load and the material bulk properties. The geometry is defined by the shape of the two surfaces in contact.

The contact temperature results from the loading, the size, the speed of relative motion and the thermal resistance of the two halves of the contact. While measuring the bulk temperature of a material poses little problem, the measurement of the contact temperature is another matter. The frictional heating generates enormous transient temperatures in the contact (flash temperatures), which are surface effects only.

When building a test model, considerations of scale are paramount. We cannot simply define the “real life” conditions (load, speed, temperature etc) and apply them to a small test piece on a test machine.

If large amounts of energy are being dissipated in small test specimens, with supporting structures that do not allow the heat to escape, the specimens may become unrepresentatively hot and the bulk temperature may exceed that experienced in practice. This is likely to produce transitions in wear or frictional response.

Contact Temperature

The temperature reached at the surface of the contact (the flash temperature) is strongly influenced by the width of the contact and flash temperature is responsible for many wear and friction effects. So what is this flash temperature and why should it be so much higher than the bulk temperature of the bulk components?

Most analysis of contact temperature is based on the pioneering work of the late Professor Blok and is perhaps too complicated to cover in a short introduction to lubricated wear testing. It is hopefully sufficient just to cover a few basic points, the first of which is to note that the actual area of contact between two surfaces in a tribological contact is always very much smaller than the nominal contact area. This is because, even with the smoothest engineering surfaces, the load is carried by the interaction of asperities. It follows that the frictional heating within the contact during relative motion is generated at the asperities and the resulting flash temperatures can easily rise to several hundred degrees above the bulk temperature of the surrounding material, but last only as long as the individual asperities are in contact.

It follows that a very small distance from the asperities, the temperature distribution is smoothed out to give an interfacial bulk temperature from which we can derive an average contact temperature distributed over the nominal contact area. This average contact temperature will always be higher than the bulk temperature of the material as measured by some sensor embedded in the bulk material of the specimens some distance from the contacting surfaces.

Energy Dissipation

All wear processes are influenced by temperature, be they the formation of oxides on surfaces, the transformation of microstructure, the formation or break-down of lubricant additive or other tribochemical films, the melting of the surface (the PV limit of the material) or thermal stress induced failure. To be more specific, wear occurs in conjunction with the dissipation of frictional energy in the contact and this is always accompanied by a rise in temperature.

Different patterns of energy dissipation will give different wear patterns. Two more global parameters have been shown to be valuable in defining these conditions in sliding and sliding/rolling wear mechanisms.

Friction Power Intensity

The friction power intensity (Matveesky) is simply defined as the amount of energy pumped into the rubbing surfaces as they pass through the contact zone. The temperature achieved in the contact and in the bulk material is directly related to the FPI and the size and thermal characteristics of the materials and their supports.

The FPI defines only the rate of energy generation and does not take into account the timescale over which this energy can be lost to the contacting materials. This timescale clearly has implications for the amount of damage caused in the contact.

$$\text{Friction Power Intensity: } Q_F = \mu P V_s / A \quad \text{W/mm}^2$$

Where μ is friction coefficient, P is the normal load, V_s is the sliding speed and A is the apparent area of contact. Practical contacts have FPIs in the range 5,000 to 20,000 W/mm^2 .

Energy Pulse

The Energy Pulse is the product of the FPI and the contact transit time. The EP therefore takes into account the length of time during which the material is subjected to energy input during its transit of the contact zone, where t_t is the transit time in seconds.

$$\text{Energy Pulse: } E_P = \mu P V_s t_t / A \quad \text{J/mm}^2$$

The Energy Pulse is analogous to the Archard Wear Law, however, the Energy Pulse equation uses the friction force rather than the applied load. This is perhaps more logical as it takes into account the rubbing conditions (but assumes that the friction coefficient can be measured).

$$\text{Archard Wear Law: } \Delta V = k P V_s t_t / A \quad \text{mm}^3$$

Each Energy Pulse can be regarded as an incremental contribution to wear or surface damage in the contact. The sum of the Energy Pulses can be used as a measure of the total wear.

Correct analysis of the EP in the real contact and subsequent modeling in the experimental design significantly enhances the probability of achieving a satisfactory result.

Types of Tribometer

There is no shortage of choice when it comes to friction and wear test machines. Selection will depend very much on what kind of test you wish to perform. For lubricated wear tests, it is useful to divide these into three categories.

Thermally Self-Regulating Continuous Energy Pulse

These are machines in which the point of contact is stationary with respect to one of the specimens and subject to constant speed uni-directional sliding. It is important to recognize that there are few real life applications other than perhaps braking systems and un-lubricated plain journal bearings, in which this occurs. These machines are widely used because they are simple. Examples include pin on disc, block on ring, crossed cylinder and sliding 4-ball test machines. Data generated by such machines rarely correlates with field data.



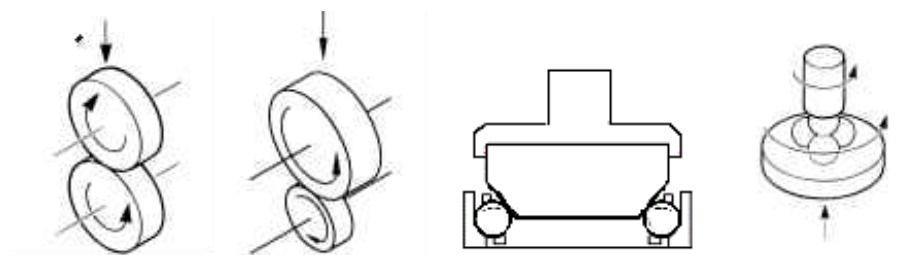
Despite their popularity, these machines are not particularly satisfactory for lubricated wear tests that are aimed at investigating lubricant additive performance, in other words, the modelling of contacts running under mixed or boundary lubrication. The entrainment conditions associated with constant speed sliding result in a requirement for heavily loaded contacts in order to overcome hydrodynamic lubrication and thus promote the kind of conditions we need to model.

These machines typically do not emulate real lubricated contacts. The Energy Pulse for the fixed point of contact specimen is continuous: it lasts for the duration of the test. This highlights one of the main limitations of these machines as models of real systems. Instead of brief rubbing episodes frequently repeated, the machines subjects one specimen to continuous rubbing and the associated temperature field dominates, frequently resulting in scuffing failure as opposed to progressive wear.

The test configuration defines the thermal conditions in the contact. The contact temperature is self-regulating and cannot be controlled as an independent variable. Uncontrolled contact temperatures results in a major problem as temperature drives the additive reaction chemistry. We have to remember that sliding speed is simply a heat generator.

Thermally Self-Regulating Cyclic Energy Pulse

These are machines where the point of contact moves with respect to both contacting surfaces and there is a close approximation to the motion in actual machine components (for example, gears, cams, joints and mechanisms). These include a number of component test machines, using idealized or standardized components such as gears, cam/follower, rolling element bearings and rollers.



These machines are essentially designed to emulate real contact conditions and typically operate under conditions broadly similar to those found in practical applications. Test piece production is no longer a simple and inexpensive process.

The most generalized version of this type of device is the two-roller machine. Here two specimen rollers are loaded together. If they are rotated at the same surface speed the motion is pure rolling and such machines are used to address particular problems in the lubrication of gears and drives in the piezo-viscous region (elastohydrodynamic). They are also used to study pitting failure (rolling contact fatigue), caused by the cyclic stressing of the surfaces. Rolling element bearing test rigs and rolling four ball machines perform a similar function.

If the rollers are rotated with an enforced surface speed difference between them, the device can be tuned to emulate the conditions found in real machine elements. We define this speed difference as the slide/roll ratio.

$$\text{Slide/Roll Ratio} = \text{Sliding Velocity} / 2 \times \text{Rolling Velocity}$$

Where:

$$\text{Sliding Velocity} = |U_1 - U_2|$$

$$\text{Rolling Velocity} = \frac{1}{2} (U_1 + U_2)$$

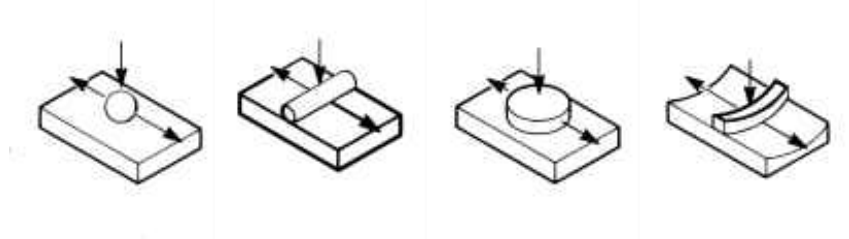
However, it is important to note that the key parameter for lubricated tests is in fact the sliding velocity. This is because the sliding velocity within the contact has a direct bearing on the contact temperature and hence what happens with the additive chemistry.

The test configuration in these machines, as with the previous category, defines the thermal conditions in the contact. The contact temperature is thus self-regulating and cannot be controlled as an independent variable.

By altering the Slide/Roll ratio and the Sliding Velocity in the contact by varying the speed of the two rollers, we can vary the wear and failure mechanism from rolling contact fatigue at pure rolling or low slide/roll ratios, to wear at moderate slide/roll ratios and low sliding speeds, through to catastrophic scuffing failure at high sliding velocities. In essence, changing the sliding speed changes the contact temperature and hence what happens to the surfaces and changing the rolling velocity changes the amount of lubricant entering the contact. By simply altering the test parameters, we can get a range of different wear and failure mechanisms and hence a range of different results from the same test configuration.

Independently Thermally Controlled Minimal Energy Pulse

These are machines in which sliding velocities are maintained at very low levels in order to minimize frictional heating and to ensure boundary lubrication at representative contact loads. These are the short stroke reciprocating rigs, in which, although sliding velocities may be low, the rate of events (specimen passes or asperity contacts) are high and by definition, cyclic. High repetition rates can be achieved without significant increases in sliding speed and corresponding loss of control of contact temperature.

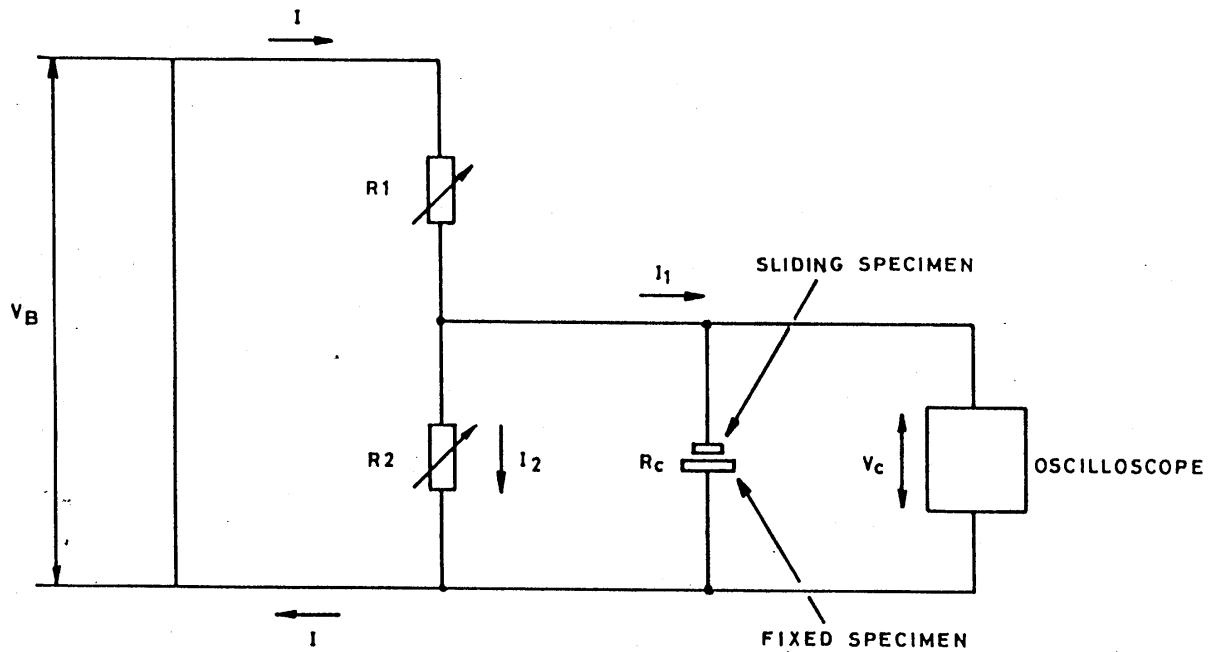


These devices (except in the case of the piston ring on liner contact near end stroke) do not model exactly real contacts, but do emulate the intimate contact conditions in a controllable and accessible way.

These rigs have an obvious similarity to the motion experienced by many practical components that have cyclic energy inputs, such as gears and cams and followers, however, by minimizing frictional heating we have the opportunity to control the contact temperature by controlling the bulk temperature of the test specimens, thus allowing contact temperature to be controlled as an independent variable. In these machines, the Energy Pulse has been replaced by controlled external heating.

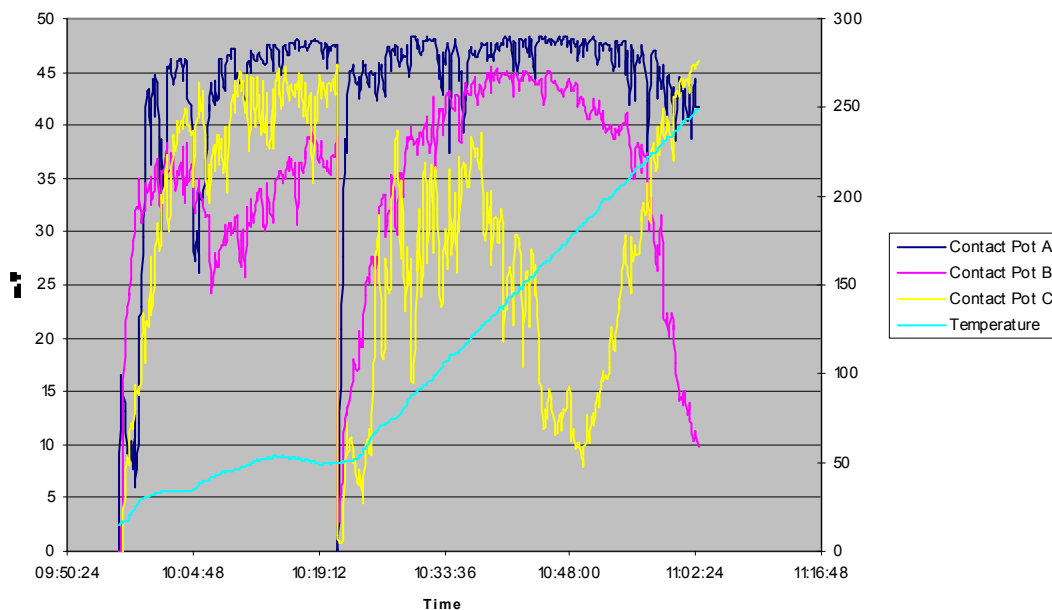
Now, as with the sliding/rolling machines, we can generate a whole range of different wear and failure mechanisms in the boundary or mixed lubrication regime, not by altering the sliding speed as with the thermally self regulating devices, but by varying the contact temperature by external heating.

A very simple test of lubricant additive chemistry is to ramp the temperature from ambient to some preset level and observe variations in friction force and the formation and failure of surface films. A common technique for doing this is electrical contact potential measurement, in which the two samples are electrically isolated and connected across one leg of a low voltage potential divider and the resulting signal monitored. When the signal is high, the surfaces are separated and when low, the surfaces are in metallic contact. Fully formed physisorbed and chemically reacted films usually give a high signal, so by monitoring the temperature at which these films form and fail we can develop an understanding of what is happening with the additive chemistry in the contact.



Lunn-Furey Contact Potential Circuit

Contact Potential - Temperature Ramp



It is worth noting that failure of the chemically reacted films usually gives rise to a scuffing failure, which is accompanied by a sharp increase in friction and surface damage. If, for example, with a given lubricant additive package and material pairs, we find that scuffing occurs at a contact temperature of, say, 240 C in our zero energy pulse bench test, we can use this knowledge to predict how the materials and lubricant will work in a real system where the contact temperature exceeds this value.

For lubricated wear tests, in which we wish to avoid catastrophic failure, we would normally choose to run a number of tests at different fixed temperatures, covering the temperatures at which different types of surface activity occur. It is worth

remembering that, depending on the lubricant additive package, different lubricants will produce different wear rates at different temperatures, so one sample might produce lower wear at low temperatures compared with another candidate, but higher wear at high temperatures. The ranking may alter, depending on the test temperature. And of course, the lubricants may have very different Temperature Distress Gaps, once again depending on the particular additive chemistry.

Conclusion

In order to perform sensible lubricated wear tests in which we aim to model the performance of materials and lubricants in real life applications, we have to choose systems that allow tests to be performed at sensible loads, while at the same time achieving mixed or boundary lubrication. Where lubricant additive chemistry is involved, we have to ensure that the resulting contact temperature, whether self generated or externally applied, is representative of the temperatures to be expected in the real system. If we wish to correlate bench test data with field data, we should aim to model the wear and failure mechanisms in actual application with our bench test model. If we cannot produce a sensible bench test, we may have to revert to full scale testing, with all that that involves!

