Wear Modeling with Sensitivity to Lubricant Chemistry

by

Benjamin C. Thomas

Bachelor of Science, Integrated Science and Technology, 2003
Master of Business Administration, 2005
James Madison University

Submitted to the Department of Mechanical Engineering
in Partial Fulfillment of the Requirements of the Degree of

Master of Science in Mechanical Engineering

at the

Massachusetts Institute of Technology

June, 2007

Signature of Author: ___

Benjamin C. Thomas
Department of Mechanical Engineering
May 18, 2007

Certified by: __

Dr. Victor W. Wong
Thesis Supervisor
Lecturer, Principal Research Scientist, Manager of Sloan Automotive Laboratory
Department of Mechanical Engineering

Accepted by: _____

Professor Lallit Anand
Chairman, Department Committee on Graduate Students
Department of Mechanical Engineering

Wear Modeling with Sensitivity to Lubricant Chemistry

by

Benjamin C. Thomas

Submitted to the Department of Mechanical Engineering on
May 18, 2007 in partial fulfillment of the requirements of the
degree of Master of Science in Mechanical Engineering

Abstract

The life of an automotive engine is often limited by the ability of its components to resist wear. Zinc dialkyl-dithiophosphate (ZDDP) is an engine oil additive that reduces wear in an engine by forming solid antiwear films at points of moving contact. The effects of this additive are fairly well understood, but there is little theory behind the kinetics of antiwear film formation and removal. This lack of dynamic modeling makes it difficult to predict the effects of wear at the design stage for an engine component or a lubricant formulation. The purpose of this research is to develop a theoretical and numerical framework for modeling the formation and evolution of ZDDP antiwear films based on the relevant chemical pathways and physical mechanisms at work.

The ability to predict the development and function of thin solid films on rough surfaces and their effect on surface wear would be a useful tool in many tribological applications, both automotive and otherwise. Completely deterministic modeling of such films may not be possible due to the complex interactions between the numerous mechanical, thermal, and chemical variables over disparate magnitudes of time and length scales. However, it is believed that useful predictions can be made by constructing a mechanistic model in which all of the most important effects are included, even if only at an approximate level.

Both the theoretical model and the numerical implementation of the concepts therein will be discussed in this work. Preliminary results from this effort are presented to illustrate feasibility and functionality on a qualitative level.

Thesis Supervisor:
Dr. Victor W. Wong
Lecturer, Principal Research Scientist, Manager of Sloan Automotive Laboratory
Department of Mechanical Engineering

Acknowledgements

This work would not have been possible without the overwhelming support that was provided from a number of sources. Because of this support, the past two years at the Massachusetts Institute of Technology have been an extremely rewarding, enjoyable, and memorable experience for me. An invaluable degree of personal and professional development has been achieved, and for this, I am eternally grateful.

First and foremost, I would like to thank my thesis advisor, Dr. Victor W. Wong, for making this research possible by organizing the research consortium in which the problem was defined, and for selecting me as a research assistant after recognizing a perfect fit between this project and my skills and interests. Additionally, Dr. Wong has offered an ideal balance between guidance and autonomy over the course of the work. He has also provided me with the direction and encouragement needed to publish this work and to present parts of the research to peers and professionals from around the world.

I would also like to extend my sincerest gratitude to the following organizations for providing financial support as well as critical feedback throughout this research:

- Caterpillar, Inc
- Chevron
- Ciba Specialty Chemicals
- Cummins, Inc
- Department of Energy
- Ford
- Komatsu
- Lutek
- Sud-Chemie
- Valvoline

The efforts of MIT and the above organizations have been combined and channeled within the Consortium to Optimize Lubricant and Diesel Engines for Robust Emission Aftertreatment Systems. This consortium was established to work toward the solving of very serious technical and environmental issues facing the industry today.

I also wish to thank a number of individuals for their collaborative input to this research effort: Nam Suh, a professor of mechanical engineering at MIT, Dr. S.M. Hsu, formerly with the National Institute of Standards and Technology and presently with George Washington University, and all of the my peers in the Sloan Automotive Laboratory.

Finally, I would like to say that I am most indebted to my caring family for providing an outpouring of love, encouragement, and patience throughout the duration of this effort. I am especially grateful to my wife, Jen, and my newborn son, Jimmy, for acting as both a solid foundation of support, and an inspirational beacon of light.

Table of Contents

List of Figures

List of Tables

1. Introduction

Zinc Dialkyldithiophosphate (ZDDP) has been used as an engine oil additive since the late 1930's [Spikes, 2004]. It was first introduced as an antioxidant, but its antiwear performance was soon recognized and attributed to the solid tribo-film (or antiwear film) formed at rubbing contacts in the presence of ZDDP. Recently, a renewed interest in ZDDP has been seen as a consequence of heightened environmental regulations relating to diesel exhaust. This is because the ash released through the combustion of engine oil with inorganic additives can clog or poison exhaust aftertreatment systems.

Many researchers and companies are now trying to better understand the processes at work so that system-wide optimization can be enhanced. The push toward low-ash oils has also created an interest in ZDDP alternatives, but it has proved difficult to find an equally effective and economical substitute. It seems that an understanding of the fundamental dynamics would be useful to improve the emission problem. If a practical model for the action of ZDDP and other antiwear additives is to be devised, then the relevant parameters must first be identified and untangled. This can be done by considering the fundamental mechanisms at work.

Over the years, ZDDP antiwear films have been studied in detail through chemical analysis and experimental data collection. However, there has been little effort in modeling the dynamics of this kind of film formation, with only a few exceptions [So and Lin, 1994, Kotvis et al, 1992]. This is largely due to the many variables at work, which span a number of scientific disciplines, but remain interdependent in the formation mechanisms. There is also a scarcity of experimental data describing the specific chemical pathways and the film formation kinetics [Spikes, 2004].

2. Background

2.1. ZDDP in Bulk

The properties of ZDDP are largely governed by the size and structure of its radical group. Aryl ZDDP is the most thermally stable, but is not generally seen in engine oil. Molecules with branched primary alkyl radicals are the next most stable, followed by those with primary, secondary, and then tertiary alkyl groups. Because the alkyl chain length affects the solubility and thermal stability of the ZDDP molecule, a distribution of high and low molecular weight alcohols is typically used to achieve the desired function over a range of operating conditions. The ZDDP in engine oil is generally composed of 85% secondary alkyl and 15% primary alkyl. ZDDP can exist as a monomer or oligomer, depending on the surrounding chemical and thermal environment. [Rudnick, 2003] The molecules may also agglomerate into micelles if the concentration and temperature are sufficient. The molecular structure of ZDDP in its most basic form is illustrated below in Figure 1.

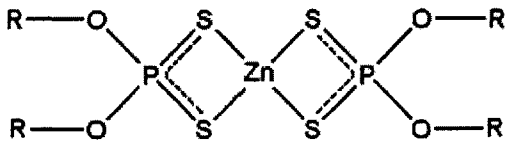


Figure 1 shows the molecular structure of ZDDP [Nicholls et al, 2005].

2.2. Migration from Bulk to Surface

Before a solid antiwear film can be created, the chemical species must first migrate to the surface. ZDDP and its decomposition products are attracted to the surface by van der Waals forces. In an engine, many chemical species compete for available space on the surface. This process can result in a layered stacking of species by polarity, with the most polar molecules adsorbed to the surface via hydrogen bonds, and the more weakly attracted molecules held at the outer surface. However, this effect is most relevant in the

hydrodynamic regime, in which there is little to no direct contact between the sliding surfaces. Friction modifiers are, by design, the dominant species on the surface in this regime. As the contact severity increases, higher temperatures cause desorption of many physisorbed species, and mechanical shearing also removes the more weakly bound molecules. In fact, a key property of friction modifiers is their ability to be easily sheared away, thereby reducing friction. [Rudnick, 2003] Under mild contact, these highly polar friction modifier molecules will continuously re-adsorb, but under more severe contact regimes, the adsorption of antiwear additives becomes preferential. Similarly, as the contact severity increases even further, the antiwear additives become less effective while the extreme pressure additives are activated. It is reasonable, then, to consider the antiwear additives alone within the window of relevant operating conditions for the initial modeling efforts.

2.3. Surface Reactions and Surface Structure

Once the antiwear additive has reached the surface, it must be assembled into an effective structural layer. In the case of ZDDP, there is an oxidation reaction on the surface, followed by a condensation polymerization process, resulting in a phosphate glass (often composed of ortho-, pyro-, poly-, and meta-phosphates). This structure is inhomogeneous in its composition and distribution, especially in the case of "tribo-films" (antiwear films formed under friction, as opposed "thermal films" formed under heat alone). Stiffer, thicker, and more elastic "pads" tend to form at the asperity peaks (where more severe contact occurs). The outer layers of these pads tend to be composed of longer-chain phosphate glass structures, whereas the bulk of the pads (and surrounding films, formed under less severe contact) are composed of shorter chain polymers. Also, in the case of an iron substrate, there is a greater concentration of iron within the antiwear film toward the substrate, and a greater concentration of zinc toward the outer surface. [Spikes, 2004] The antiwear film is generally less effective on other substrates such as aluminum and ceramic surfaces [Nicholls, 2005]. It is apparent that the chemical and mechanical nature

of the substrate plays a key role in the formation, adhesion, and composition of the antiwear film.

2.4. Wear Mechanisms

Another key element to the picture is the removal of surface material. Both the antiwear film and the metal substrate will be continuously worn away. The wear to the metal is reduced by the presence of the antiwear film, whose thickness is determined by the difference between the rate of film growth and the rate of film removal. There is a tremendous body of literature on the topic of wear, both analytical and empirical. Some methods consist of a single equation, usually in a form resembling the Archard equation. Other methods begin with a fully deterministic solution of the contact condition, such as the Elasto-Hydrodynamic Lubrication (EHL) model proposed by Zu and Hu in 2001 that can numerically analyze the state of real contacting surfaces at any fluid-film thickness. The goal in an engineering model is to use the best techniques to achieve sufficient accuracy with minimal computational effort. The antiwear film formation is dependent on asperity-level contact conditions, so it is clear that a purely macroscopic model will not suffice in a wear model that is sensitive to lubricant chemistry. Statistical contact models derived from the work of Greenwood and Williamson (1966) can provide critical inputs for approximating both the wear rates and the thermodynamic conditions.

2.5. Complexity Issues and Modeling Strategy

There has been detailed research in each of the above categories, but few attempts at modeling the overall dynamics. The ultimate goal is to achieve a useful model that accounts for all of the relevant chemical and physical drivers. A completely deterministic model of the action of ZDDP in an operating engine may never be possible. The chemistry becomes very complex when dealing with a fully formulated engine oil, combustion gasses, contaminants, reaction byproducts, and the many other uncertainties.

For example, many binary studies have been done to understand the effect of other additives on ZDDP performance. It has been shown that detergents have an antagonistic effect, while the other relationships are less clear, even when considering one pair at a time [Nicholls, 2005]. Many of the proposed pathways for the action of ZDDP in base oil alone are not entirely agreed upon. The combinatorial possibilities for chemical reactions with this many variables cannot be fully enumerated. For these reasons, only the more consequential chemical reactions can be explicitly modeled.

Another major issue is that of scale: the temporal and spatial scales are too extreme to be directly handled at all levels. For example, a molecular dynamic simulation might help to predict reaction pathways, but it cannot be carried out on the same time scale as a moving piston. Similarly, every asperity cannot be modeled over the surface area in an engine. A useful engineering model will require high efficiency in order to accommodate parametric studies. The initial modeling efforts must find the right balance between accuracy and feasibility. While the model must be built around underlying physical and chemical drivers, statistical approximations will be required for speed.

Another complication arises from the spatial inhomogeneity of antiwear films. In the literature, the antiwear film is sometimes described in terms of several functionally different layers [Williams, 2004]. The hardest layer at the metal surface helps to bear and distribute contact pressure, while the softer middle layer can reduce the shock of impacts, and outermost layer becomes an easily shearable sacrificial layer. However, despite the different functional components, it is known that the antiwear performance is more strongly related to the presence of the film than to its thickness [Fuller, 1998]. Therefore, the primary goal of this work is to determine the presence and spatial distribution of the film, while the existence of the separate functions can be taken for granted, so the explicit distinction between the different layers will not be necessary.

2.6. *Summary of Theoretical Processes*

In spite of the many variables and complications at work, a reasonably simple picture of the role of ZDDP in wear protection can be derived. There are three ways in which ZDDP can reduce wear: i) it leads to the generation of a physical barrier in the form of an antiwear film, ii) it reduces oxidation by reacting with oxygen and peroxides in solution, and iii) it "softens" otherwise abrasive particles in the engine by coating and "digesting" them [Spikes, 2004]. The antiwear film is considered to be the most significant of these functions, so its formation and removal will be the focus of this research. Figure 2 represents a conceptual framework for the major steps relating ZDDP to surface wear.

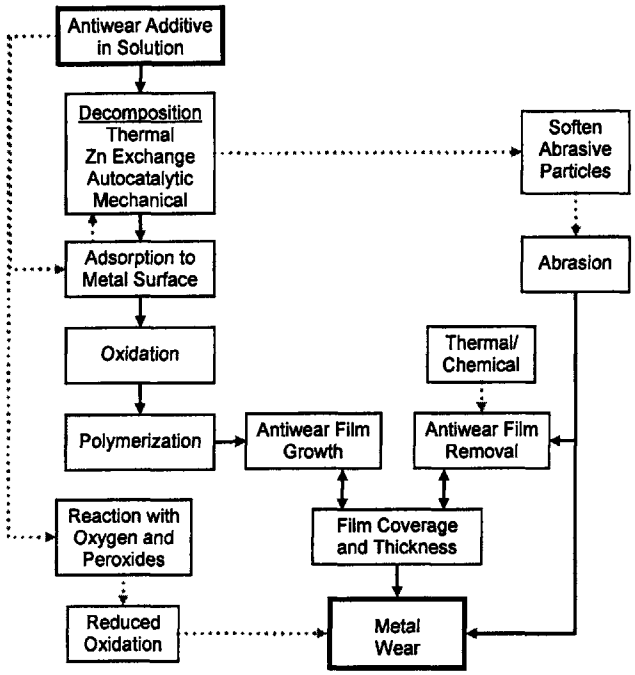


Figure 2 is a conceptual framework for the processes behind ZDDP antiwear performance.

Simply stated, the evolution of the antiwear film relies upon two competing rates: the rate of formation and the rate of removal. If the long-term rate of removal is greater, then there can be no protective film. The growth of the antiwear film can be broken down into a sequence of chemical reactions. Although the exact stoichiometry of every possible

pathway is very complex, it will still be useful to make predictions using the general form of the most significant or "rate limiting" reactions. The process begins with the decomposition of the ZDDP molecule in solution. Next, the decomposition products adsorb to the metal surface, followed by further transformation in an oxidative reaction, and finally, polymerization [Nicholls et al, 2005].

The breakdown of ZDDP into its decomposition products can happen through various pathways. [Spikes, 2004] The simplest reaction is pure thermal decomposition, though the relative importance of bulk temperature vs. flash temperature is debatable. Decomposition can also occur through a ligand exchange process in which the zinc atoms in ZDDP are replaced with iron cations, resulting in a molecule with reduced thermal stability. Also, ZDDP is known to be somewhat autocatalytic. That is, the decomposition reaction is accelerated in the presence of its own products. Mechanical stress alone may also drive decomposition, with pressure-induced molecular scission.

The subsequent adsorption process depends on the local temperature, the concentration of adsorbates, and the local surface material. It might also be possible for ZDDP to adsorb intact and then react, but any loosely bound physisorbed molecules are generally removed under friction [Rudnick, 2003]. Also, the adsorption process is not believed to be diffusion controlled (in a rate-limiting sense). It has been shown that the adsorption kinetics are the same in either a stirred or unstirred tank [Bovington and Dacre, 1984]. Of course, most antiwear applications are fairly "well-stirred" to begin with. After adsorption, there is an oxidation reaction on the surface, which depends on the local temperature and the concentration of oxidative reactants in solution. Finally, there is a condensation polymerization reaction that results in a solid phosphate glass. This last step is believed to be orders of magnitude faster than the others [Hsu et al, 1988], and so it need not be directly modeled (as it is not rate-limiting). Each of the preceding reactions can be approximated using classical chemical kinetics, with an Arrhenius-type dependence on temperature.

Given the contact severity, the wear of the film (and of the metal substrate) can be estimated. There exists a vast body of research on wear prediction, but because of the need for computational efficiency, the best approach may be a material-specific wear equation applied at the asperity level. This could be a linear estimate (with the wear volume being proportional to the local contact pressure and sliding distance [Suh, 1986]), or it may take the form of a nonlinear function that more accurately spans wear regimes (e.g. the sum of mechanical abrasive wear and thermo-chemical oxidative wear). A similar strategy that utilized a linear wear relationship in a dynamic contact model was able to successfully predict the wear of valve components [Colgan and Bell, 1989].

Within these steps, there are quite a few parameters that require values. It is believed that useful results can be generated by addressing all of the relevant mechanisms, even if some of the steps are only predicted on an approximate level. Many of the required parameters are already known, or can be calculated from fundamental principles. For example, some pre-exponential factors and activation energies can be calculated from chemical reaction theory. Other parameters can be calculated directly from experimental data, such as the order of a reaction, or wear coefficients. Some parameters might remain illusive to direct methods, and will be extracted from experimental data through parametric studies using the fully developed numerical model in combination with experiments designed to isolate certain effects. However, it should be noted that this is quite different from purely mathematical curve-fitting. The purpose of a mechanism-based model is to obtain generality, so if the relevant physics is truly captured, then all constants will be *true* constants, and can subsequently be used in predictive modeling without a need for additional experimentation.

The primary inputs to the model will include the properties of the surface, the operating conditions, and the properties of the lubricant (and additive package). From the macroscopic dynamics, the asperity-level forces can be efficiently approximated with a statistical contact model. These contact conditions will provide inputs for predicting the thermal conditions (e.g. flash temperatures). Given the chemical species present, the bulk temperatures, and the average frequency, duration, and magnitude of flash temperatures,

it will be possible to infer the chemical reaction rates on which the formation of the antiwear film relies. The growth and removal of a metal oxide layer will also be modeled using the same concepts, but different chemical reactions [Grosvenor et al, 2005]. Given the state of these films and the contact conditions, the degree of wear can be estimated.

Some of these steps occur in parallel, while other steps are sequential. If the sequential steps are orders of magnitude apart in speed, then it is only necessary to model the slowest, rate-limiting step, which will act like the bottleneck in the process. However, it is unclear at this point which step is the limiting factor. In reality, one step may be restrictive in one regime, while a different step may be rate limiting under different operating conditions. For this reason, it is necessary to account for all of the steps together. In summary, the rate of growth of the antiwear film is proportional to the final oxidation step, which is limited by the adsorption step, which depends on the decomposition step, which is comprised of several mechanisms in parallel. All the while, a wearing process works against the growth of the film. Together, these steps comprise a dynamic process even in constant operating conditions; as the reactions proceed, the chemical concentrations, film thicknesses, flash temperatures, and other boundary conditions will evolve, affecting future rates. The numerical organization of these processes is a complicated task, and will be discussed in detail in later sections of this work.

3. Theory of Antiwear Film Removal

The metal wear in an engine, as in many other lubricated sliding applications, is largely dominated by abrasion, oxidative wear, and direct asperity shearing. Third-body abrasion is caused by loose, hard particles such as dust, metal oxides, soot, and wear debris, which can all tumble, plow, and agglomerate in a sliding contact. Oxidative wear occurs when the metal surface is oxidized and scraped away. The presence of an oxide layer can also affect the mechanical and thermal contact conditions, and so it should be modeled in a similar manner as the antiwear film formation and removal, which will be discussed in the following section. The wear from direct asperity contact and shearing is most relevant in the early running-in period. Under moderate loads, the asperity contact tends to shift from plastic to elastic as the running-in is completed. After the running in, the surface roughness can be considered constant, even though wear may still be occurring. In terms of modeling, it is as if the material is flowing from the substrate to the surface and being removed. Conversely, material flows from the solution to the surface in the formation of antiwear films.

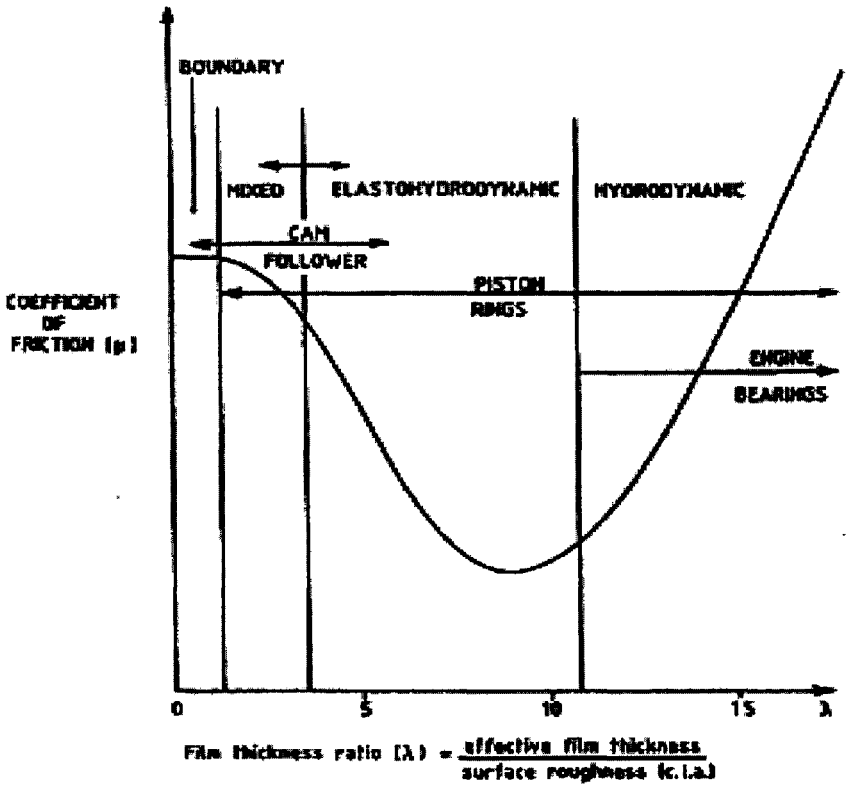
The antiwear film mechanically prevents wear from third-body abrasive particles in the same way that it reduces the severity of asperity contacts. If a general abrasive wear equation is assumed [Suh, 1986], then it can be said that the abrasive wear coefficient will depend on the pair of materials in contact. For example, the coefficient will have a higher value in the absence of the film, and a lower value in the presence of the film (where "presence" will be defined as some minimum functional thickness). The hardness will likewise be material-dependent (the antiwear film has a hardness of between 1.5 and 3.5 GPa [Spikes, 2004]). The equation below is a common approach in abrasive wear modeling

$$V_i = \kappa_{ij} \frac{LS}{H_i} \tag{1.a}$$

where V_i is the wear volume for the given material i, κ_{ij} is the wear coefficient for the material pair, L is the load, S is the sliding distance, and H_i is the material hardness. This can also be written in terms of W_i, the rate of generation of wear volume

$$W_i = \kappa_{ij} \frac{LU_s}{H_i} \tag{1.b}$$

where U_s is the sliding speed. Of the relatively few models that attempt to predict wear in automobile engines, nearly all use this equation in some form [Colgan and Bell, 1989, Priest and Taylor, 2000; Gangopadhyay, 2000; Tomanik, 2001]. A typical approach involves the adjustment of κ according to the lubrication regime from the Stribeck curve (Figure 3). The coefficient is assumed to have a constant value in boundary lubrication (where the contact load is carried entirely by the asperities) and a value of zero for full hydrodynamic lubrication (where the contact load is carried entirely by the fluid). It is further assumed to vary linearly in between, resulting in a piecewise linear representation of κ across the boundary, mixed, and hydrodynamic regimes [Colgan and Bell, 1989].



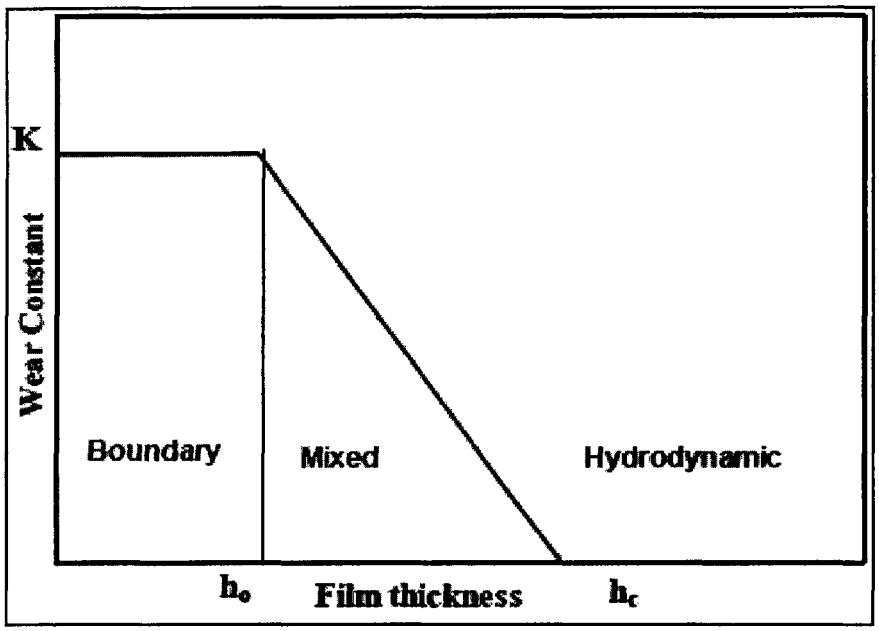


Figure 3 shows the Stribeck curve as it relates to various engine components [Priest and Taylor, 2000] and the variation of the wear constant with dimensionless oil film thickness [Gangopadhyay, 2000].

For the purposes of antiwear film modeling, it is suggested that this strategy be taken a step further, and that the constant κ might be applied directly at the asperity level using asperity contact areas and pressures. Then this transition effect will be handled automatically because L will represent the local asperity contact load rather than the nominal load (which includes the force borne by both the asperities and the lubricant). Using this strategy, κ is a constant material property of the surfaces, and L will analytically approach zero as the lubrication regime shifts into hydrodynamic (thus, an EHL or statistical contact model will be used to provide the necessary boundary condition for the wear model). Similarly, instead of nominal distance per time, U_s must represent the actual contact distance per time. The moving asperities are only in contact intermittently, so the nominal speed should be multiplied by the fraction of time that the tall asperities spend in direct contact with the opposite surface. In order to complete the wear calculation, the state of the surface must still be known, so the antiwear film coverage must be determined. When applying the abrasive wear equation to the removal of the antiwear film, it is believed that the following modification might be useful:

$$W_i = \kappa_{ij} \frac{L U_s}{H_i} \left(1 + h_{nd}{}^{w}\right) \tag{2}$$

where h_{nd} is a non-dimensional film thickness (e.g. thickness over asperity contact radius) and w is a constant. This power-law is proposed to account for the shearing effect upon thicker, patchy films, as well as any effect of the structural inhomogeneity of the film along its depth.

It should be noted that the simulation assumes that the sliding surface has already been "run in." *A priori* predictions of initial, transient wear to a surface and the changes in its roughness profile are extremely difficult and unreliable. Moreover, such predictions are outside of the scope of this research. If the wear is severe enough to drastically change the surface profile and contact condition, then it is clear that the antiwear film has failed, and further simulation is not necessary. Predictions of mild and long-term wear are more

useful at present. It is also worth noting that while abrasive wear may actually occur as the net effect of many discrete wear events, these events are collectively tracked as an average speed on a one dimensional surface partition.

Simply stated, the evolution of the antiwear film relies upon two competing rates: the rate of formation and the rate of removal. If the long-term rate of removal is greater, then there can be no protective film. This can happen, for example, under very low ZDDP concentrations, extreme temperatures, or very high contact loads. When boundary lubrication conditions are favorable for film formation, the thickness tends to increase to a steady value of around 100 nm. As stated above, the rate of removal can be approximated with a modified wear equation applied at the asperity level. The rate of formation, however, is more difficult to predict. As illustrated in Figure 2, the formation process requires ZDDP decomposition, an adsorption step, an oxidative reaction, and a polymerization reaction.

4. Theory of Antiwear Film Growth

The process of antiwear film formation begins with the decomposition of the ZDDP molecule in solution. Next, the decomposition products adsorb to the metal surface, followed by further transformation through an oxidative reaction [Nicholls, 2005]. Lastly, a rapid polymerization process takes place, resulting in a solid layer of polyphosphate glass. As indicated in Figure 2, it might also be possible for the ZDDP molecules to adsorb intact and then decompose, which should be included in a model if data indicates that this pathway is significant. The specific sequence and stoichiometry of the final step is somewhat in dispute, but it is believed to be several orders of magnitude faster than the others [Hsu, 1988], so it can be discarded from the kinetic model because it is not rate limiting. If the remaining steps can be described in a kinetic sense, then the antiwear film growth rate can be predicted in terms of the overall chemical reaction rate.

The first step in the process remains somewhat unsettled. A number of mechanisms have been proposed as drivers of ZDDP decomposition [Fujita and Spikes, 2004]. The most likely mechanisms include bulk temperature, flash temperature, autocatalysis, and ligand exchange. Some of these mechanisms will dominate the process, while others may only be relevant in certain operating regimes, if at all. These mechanisms are discussed in more detail below.

4.1. ZDDP Decomposition

4.1.1. Thermal Decomposition

The speed of a chemical reaction is generally a function of the concentration of the reactant or reactants [Missen et al, 1999]. In the case of ZDDP thermal decomposition, this might be represented by the following relationship:

$$r_T = k_T c_Z^{\alpha_1} \tag{3}$$

where r_T is the rate of generation of thermal decomposition products, k_T is the "proportionality constant" in thermal decomposition, c_z is the concentration, or molarity of reactant (ZDDP), and α_1 is a constant power that also denotes the "order" of the reaction. A typical concentration of ZDDP in engine oil is around 1% by weight, and the order of the reaction can be identified empirically (often equal to 1). Note that there are many different ways to define the units of the rate and proportionality constant [Masel, 2001]. Also, despite its name, the proportionality constant is not truly constant, but rather a function of temperature.

The decomposition of ZDDP, like most chemical reactions, is largely driven by thermal energy. It is well known that the growth rate of antiwear films is very sensitive to temperature. This is especially clear in the case of "thermal films" which are formed on a metal surface in heated oil, but in the absence of friction. The relationship is less clear in the case of "tribo-films" which can appear at room temperature when two metal surfaces are rubbed together. The Arrhenius law is used to describe the kinetics of thermally driven reactions [Missen et al, 1999]

$$k_T = A_T e^{\frac{-E_T}{RT}} \tag{4}$$

where A is known as the pre-exponential factor, E is the activation energy, R is the ideal gas constant, T is the temperature, and the subscript $_T$ indicates values specific to the thermal decomposition reaction. In cases where the behavior deviates over large ranges of temperature, a more general form might be used [Masel, 2001] to better fit the data, where an additional temperature factor is included and raised to the power n.

$$k_T = A_T T^n e^{\frac{-E_T}{RT}} \tag{5}$$

Although the Arrhenius equation is meant to represent only elementary reactions, it often applies to more complex reactions as well (which are actually the combination of multiple Arrhenius curves [Masel, 2001]). This is likely the case for the steps behind antiwear film growth, as it appears to fit the exponential pattern quite well (Figure 4). Note that this figure is presented as a qualitative representation only, as the final thickness encompasses several different steps together, which should be modeled independently.

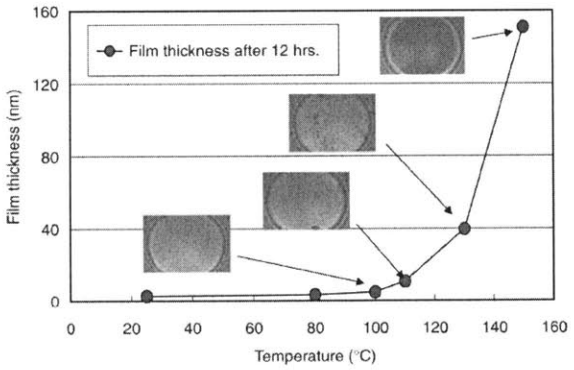


Figure 4 shows an exponential trend in the thickness of a thermal film after 12 hours as a function of temperature [Fujita and Spikes, 2004].

4.1.2. Flash Temperature

It is clear that the formation of antiwear films is largely driven by high temperatures. The decomposition step and the subsequent chemical reactions discussed below are all functions of temperature. However, it is known that tribo-films, formed between rubbing surfaces, can develop in far lower temperatures than thermal films, formed on a single surface. It may be that the same thermal mechanism is at work in either case, but on a different scale. The work done by friction between two surfaces leads to heat generation at the interface. Individual asperities can reach very high temperatures, known as *flash*

temperatures, when they collide with other asperities or are plowed along an opposing surface. Flash temperatures in rubbing conditions might have the same effect as high bulk temperatures on a single surface. Conversely, flash temperatures may be too small or short-lived to have any real effect, and tribo-films may be generated by some other mechanism. The truth is probably somewhere in between, as several driving factors could be operating in parallel. When constructing a kinetic model for antiwear film formation, it will be essential to account for all potentially significant factors, including flash temperatures. Given the frequency, intensity, and duration of such temperatures, it will be possible to quantify their effects on each chemical reaction, and ultimately show whether these predictions are consistent with experimental data. A method for estimating flash temperatures is described below.

For a circular asperity contact area of πr_a^2, and an energy source (from friction) of density q, the flash temperature, T_f, at time t can be approximated in the following form [Carslaw and Jaeger, 1959; Lim and Ashby, 1987]:

$$T_f = T_s + \frac{pqr_a}{\pi^{1/2} K_s} \tan^{-1}\left(\frac{4a_s t}{r_a^2}\right)^{1/2} \tag{6}$$

where T_s is the bulk, or "sink" temperature of the metal surface, K_s is the thermal conductivity of the material, a_s is the thermal diffusivity of the material, and p is a partition function that assigns the appropriate fraction of total heat to the asperity (with the remaining heat going to the opposing surface). The value of p, in its simplest form, depends on the thermal conductivity of the asperity surface, K_s, and that of the opposing surface, $K_{s'}$ [Suh, 1986].

$$p = \frac{K_s}{K_s + K_{s'}} \tag{7}$$

The magnitude of the heat source comes from the work done by friction, with the coefficient μ (which depends on surface materials), at a speed of U_s and pressure of P.

$$q = \frac{\mu L U_s}{\pi r_a^2} = \mu P U_s \tag{8}$$

The calculation can be made even more realistic by accounting for the insulating effect of the antiwear film and any oxide film of thicknesses h_{film} and h_{ox}, respectively. This is done by replacing the thermal conductivity of steel with an effective conductivity, K_{eff}, or the sum of the thermal resistances in series

$$\frac{l}{K_{eff}} = \frac{h_{film}}{K_{film}} + \frac{h_{ox}}{K_{ox}} + \frac{l - (h_{film} + h_{ox})}{K_s} \tag{9}$$

where K_{film} is the thermal conductivity of the antiwear film, with a value of about 1.0 W/mK [Fujita and Spikes, 2004], as opposed to K_s of about 30 for steel. Also, K_{ox} is the conductivity of the metal oxide layer, and l is a characteristic length, or an equivalent linear diffusion distance for the substrate material, which can be estimated with the following equation [Ashby et al, 1991].

$$l = \frac{r_a}{\pi^{1/2}} \tan^{-1}\left(\frac{4 a_s t}{r_a^2}\right)^{1/2} \tag{10}$$

The thermal diffusivity is defined as the thermal conductivity divided by the product of density and heat capacity [Carslaw and Jaeger, 1959], so an effective thermal diffusivity can also be calculated.

$$a_{eff} = \frac{K_{eff}}{\rho_{eff} c_{eff}} \tag{11}$$

with

$$\rho_{eff} = \frac{h_{film}}{l}\rho_{film} + \frac{h_{ox}}{l}\rho_{ox} + \left(1 - \frac{h_{film} + h_{ox}}{l}\right)\rho_s \tag{12}$$

and

$$c_{eff} = \frac{h_{film}}{l}c_{film} + \frac{h_{ox}}{l}c_{ox} + \left(1 - \frac{h_{film} + h_{ox}}{l}\right)c_s \tag{13}$$

where ρ_{eff} and c_{eff} are the effective density and heat capacity, and take the form of a weighted average (from relative thicknesses) between those of the films and the bulk substrate. Together, these relationships give the following expression for the flash temperature.

$$T_f = T_s + \frac{p\mu LU_s}{\pi^{3/2} r_a K_{eff}} \tan^{-1}\left(\frac{4a_{eff}t}{r_a^2}\right)^{1/2} \tag{14}$$

4.1.3. Decomposition by Autocatalysis

The decomposition of ZDDP is known to be autocatalytic to some degree. This means that the decomposition products act as catalysts for further decomposition, causing an increasing rate in otherwise constant conditions until reactants become scarce. It is unclear whether this phenomenon has a significant effect on the overall reaction rate, but it will be included here as a candidate. Quantitatively, the rate of additional decomposition products generated by autocatalysis, r_A, is given by

$$r_A = k_A \frac{c_Z^{\alpha_2} c_P^{\alpha_3}}{c_Z^{\alpha_2} + c_P^{\alpha_3}} \tag{15}$$

where c_P is the concentration of decomposition products. Note that the "order" of this reaction is the sum of the exponents: $\alpha_2 + \alpha_3$ [Masel, 2001]. The terms in the denominator indicate a "saturation effect" that is common to catalysis [Missen et al, 1999]. If one concentration is much greater than the other, then the rate is effectively proportional to the scarce chemical and unresponsive to changes in the abundant species. In order to keep track of the concentrations, both the rate of reaction and the rate of flow through the reactor volume must be considered. The process is an open system with bulk lubricant diluting any local changes as a function of the local residence time.

4.1.4. Decomposition by Ligand Exchange

The ZDDP molecule can react with certain species of metal, including iron cations (such as Fe^{3+}), through a ligand exchange process. The zinc atom can be replaced with iron, resulting in a molecule that has a lower thermal reactivity limit than before. Thus, the molecule will decompose at more moderate temperatures. This mechanism might therefore play a role in the ability of tribo-films to form in lower temperatures than thermal films, as the rubbing of metal surfaces can release such cations.

The ligand exchange effect will depend on the degree to which the thermal threshold is reduced based on chemical reactivity, and on the number of ZDDP molecules that have been changed (which will in turn depend on the local temperature and the amount of available iron cations with which to react). It is possible that the thermal stability is sufficiently reduced so that all ZDDP molecules that undergo the exchange will immediately decompose.

The release of various iron species through the process of rubbing is similar (if not identical) to the wear process, and can be assumed to have the same proportionality to load and speed. It will have a smaller "wear" coefficient because only a tiny fraction of

the wear volume will be in the form of reactive cations. In other words, the rate of release of iron cations, r_{Fe}, is proportional to the rate of generation of iron wear volume, W_{Fe}, or

$$r_{Fe} = \gamma W_{Fe} \tag{16}$$

where γ is a constant and W_{Fe} is given by Equation 2, but only in the case of a non-coated area of metal surface.

Observations appear to support the ligand exchange phenomenon. Antiwear films are composed of polyphosphate glass that contains mostly zinc cations, but with increasing density of iron cations toward the iron surface [Spikes, 2004]. Moreover, these iron cations are not present in thermal films, suggesting that ligand exchange is a formation driver that depends on the rubbing of the iron surfaces. Other metal cations such as copper or gold can produce the same effect, but with varying degrees of affinity [Spikes, 2004].

The ligand exchange process can be modeled as a catalytic effect. Strictly speaking, the iron cation is more of a reactant than a catalyst because the iron is not necessarily regenerated, as it might instead remain in the solid film. However, the iron is replenished through the rubbing action, and the generation of decomposition products, r_{Lg}, should have a similar form with a saturation effect as in the case of autocatalysis

$$r_{Lg} = k_{Lg} \frac{c_Z^{\alpha_4} c_{Fe}^{\alpha_5}}{c_Z^{\alpha_4} + c_{Fe}^{\alpha_5}} \tag{17}$$

where k_{Lg} is the proportionality constant for the ligand exchange (which will exhibit an Arrhenius dependence on temperature), c_Z is the concentration of ZDDP, which is raised to a power α_4, and c_{Fe} is the concentration of the replacement cation, which is raised to a power α_5. This represents a separate path for the ZDDP. After a ligand exchange occurs,

the molecule will decompose according to the same thermal decomposition equation, but with a smaller energy of activation.

The decomposition products of the ZDDP molecule from thermal, autocatalytic, and ligand exchange reactions provide the necessary ingredients for the formation of the solid antiwear film. There is an adsorption process that brings the chemicals to the surface, followed by an oxidative reaction on the surface, and then a condensation polymerization process to form the solid polyphosphate glass [Willermet et al, 1995].

4.1.5. Other Means of Decomposition

Mechanical shearing has also been suggested as a means of decomposing the ZDDP molecule [Spikes, 2004]. The theory is that pressure and friction between asperities can physically shear apart the bonds within the molecule. If this molecular scission is possible, then the number of molecules broken should be proportional to the area of contact and local shearing force, which will depend on the contact pressure and friction between asperities. This condition will also depend on the presence of any existing antiwear film, which has different properties than the bare metal surface. It is unclear whether this effect is possible, let alone significant. It has therefore been omitted from the initial modeling efforts, but will remain a candidate mechanism to be included if the data points in that direction.

Electron emission is another possible cause of ZDDP decomposition. However, this mechanism is not well understood, and it appears to be of little significance in practical applications, so it too has been omitted at present.

ZDDP is also known to decompose through a hydrolysis reaction, but an oxidative reaction usually takes place preferentially (before hydrolysis becomes significant). It is further believed that the oxidation reaction in solution does not yield products that are still viable for antiwear film formation, though it does reduce unwanted oxidative

reactions in the lubricant and on the metal surface by decomposing some reactants (e.g. oxygen and peroxides) and preventing the release of pro-oxidant ions from metals [Spikes, 2004]. It is therefore desirable to understand the rate of ZDDP loss to this antioxidant pathway for its competition with the antiwear pathway. The current modeling efforts have not yet incorporated this effect.

4.2. Adsorption Step

Adsorption occurs preferentially as the most polar molecules react with the surface first. Numerous species compete for adsorption sites in this way, and the process is driven by temperature. If we simplify the overall behavior of adsorption as a single process by considering the relevant window of boundary conditions, then its rate, r_{Ad}, can be approximated from Langmuir-Hinshelwood kinetics [Missen et al, 1999]

$$r_{Ad} = c_P \sum k_i \theta_i \qquad \sum \theta_i = 1 \tag{18}$$

where c_P is the concentration of decomposition products in solution, k_i is a proportionality constant, and θ_i represents the fraction of the local surface area that is covered by material i (i = product (P), antiwear film (*film*), oxidized surface (*ox*), or fresh substrate surface (s)). There will be a separate k for each type of surface, depending on both the temperature and material affinity. For example, at a given temperature, the proportionality constant might be at or near zero for the fraction of surface with freshly adsorbed products, low for the fraction of surface covered by the fully developed antiwear film, high for a metal surface (which may be oxidized or otherwise contaminated), and extremely high for freshly exposed, highly reactive metal surface. It has also been shown experimentally that the adsorption kinetics of ZDDP and some of its products do follow a first-order Langmuir model, and that the process is not diffusion controlled because stirring has "little or no effect on the adsorption kinetics" [Bovington

and Dacre, 1984]. At very high temperatures, or in high concentrations of certain species (including water or certain antagonistic additives), there can be a chemical removal effect similar in form to adsorption. If desorption turns out to be of significant magnitude in the operating range of interest, then it can be included with the same form as in Equation 18, but with opposite sign and a different energy of activation embedded in the Arrhenius constant, k.

4.3. Oxidation Step

Various lubricants, including esters, mineral oils, and synthetic hydrocarbons, undergo oxidation at a rate described by the Arrhenius equation [Hsu, 1994]. Again, this equation is meant to describe elementary reactions, but it happens to fit the overall reaction of lubricant oxidation over a wide temperature range. The oxidation step of antiwear film formation might also be governed by Arrhenius kinetics, though the formula is modified when chemicals in solution are reacting with a surface [Missen et al, 1999]. The rate, r_{Oxd}, is then given by

$$r_{Oxd} = k_{Oxd}\theta_P c_{Ox}^{\alpha_6} \tag{19}$$

where θ_P represents the fraction of surface area covered by adsorbed decomposition products, and k_{Oxd} is the proportionality constant for the oxidation reaction (which is governed by temperature according to the Arrhenius equation). Also, c_{Ox} represents the concentration of oxidative reactants (such as oxygen or peroxides) in solution, raised to a power according to the order of the reaction.

4.4. Polymerization Step

Once again, the subsequent polymerization step is believed to be much faster than the rate limiting steps [Hsu, 1988], so its specific chemistry and kinetics are not needed in the kinetic modeling of antiwear film formation. The reaction can essentially be taken for granted upon completion of the oxidation step. The final changes in thickness will be calculated by converting the number of moles reacted (from the above processes) using the typical molecular composition and density of an antiwear film.

5. Structure of Numerical Model

5.1. The Reactor Volume

As seen in the previous section, the growth of antiwear and other solid surface films is largely driven by chemical reactions. A "reactor volume" has been defined to systematically account for the products, reactants, and states of interest within a tribological contact. The reactor volume is defined and sub-partitioned in a manner that promotes efficient handling of the chemical and thermal variables.

The macroscopic dynamics and mechanical kinematics are considered to be known inputs. The reactor volume is defined as the volume above a macroscopically homogeneous element of surface area (Figure 5). The area can be of any shape or size provided that the macroscopic pressures, temperatures, speeds, etc are sufficiently constant at any point across the surface element for the duration of the time step being used. The degree of resolution is decided by the user, and designated via the assignment of the dimensions of the reactor volume and the length of the time step. For two surfaces in contact, the height of the reactor volume can be thought of as one half of the RMS separation between the surfaces (e.g. average thickness of the lubricant layer above each surface).

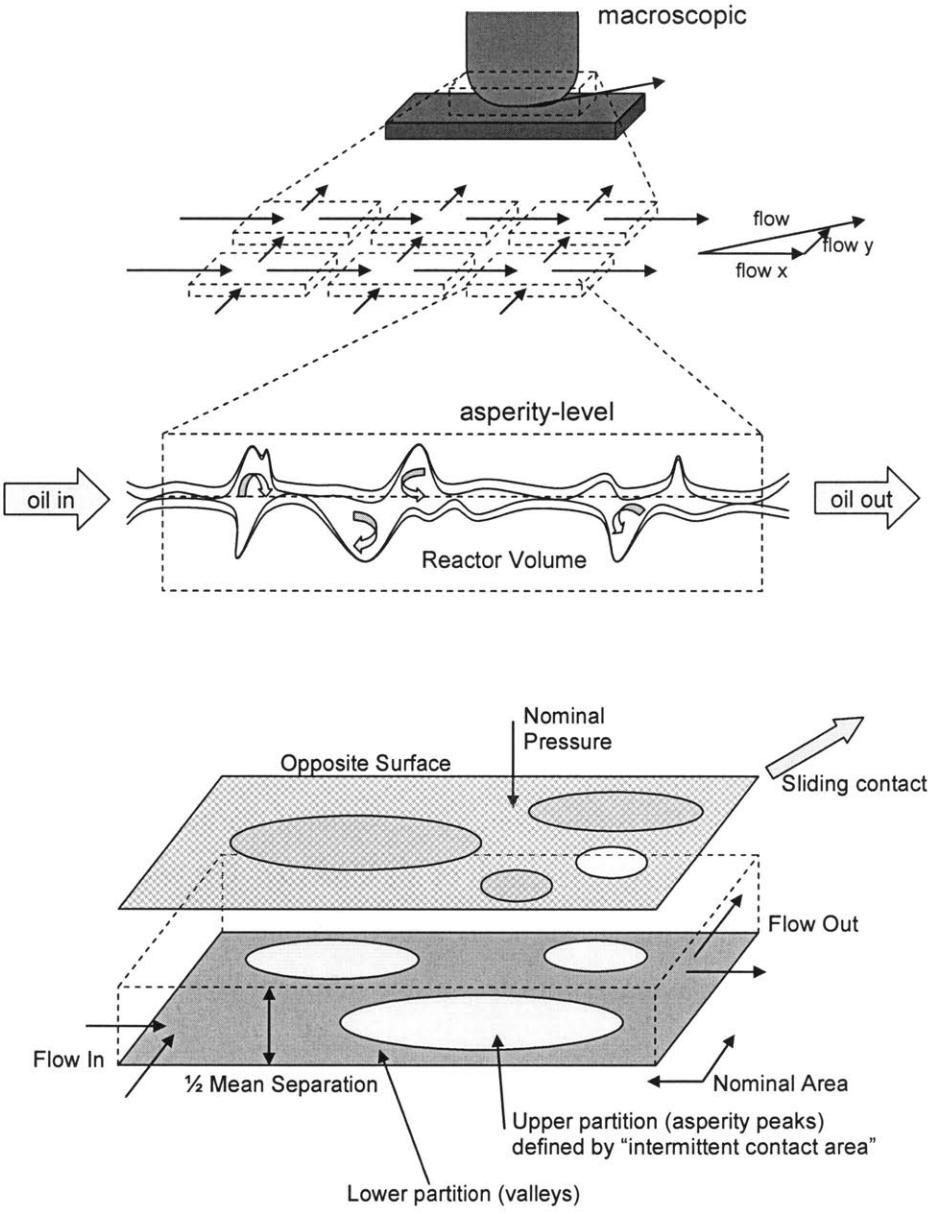


Figure 5 shows a schematic of the reactor volume as defined in this model.

Any lubricant within the reactor volume is assumed to be well mixed. The properties of the lubricant, including its chemical composition, will vary across time steps according to internal reactions, mass transport to and from the surface, and mass transport to and from the bulk lubricant outside of the reactor volume (or in neighboring reactor volumes). If necessary, a network of reactor volumes can be modeled together, with mass transport between the elements.

The length of the time step used in the model will depend on the temporal resolution required to resolve both the changes in the surface (such as the thickness of the antiwear film) and the dynamics of the moving part (such as crank-angle degrees in a reciprocating piston).

5.2. Spatial Sub-Partition of the Reactor Volume

Although the reactor volume is macroscopically homogeneous, it is understood that the microscopic, or asperity-level properties, will not be homogeneous. Each reactor volume will be sub-partitioned in both time and space in order to account for the microscopic variations. ZDDP antiwear films, for example, tend to form thicker and more elastic "pads" at asperity peaks, with a thinner layer in the valleys. This inhomogeneity will be modeled in an average sense by partitioning the surface according to its height. Two partitions can be used to describe this difference between the peaks and valleys. A greater number of partitions might be used if a greater resolution in the local contact state is desired. On each surface partition, a separate set of variables will be tracked, including contact pressure, temperature, and solid film thickness and coverage.

For the purposes of this discussion, two surface partitions will be considered. The relative size of each partition will be application-specific. The assumption of two partitions will be particularly accurate when dealing with highly plateaued surfaces, as in the case of an already run-in surface. In a sliding contact, it is well known that the real area of contact (at the asperity level) is much smaller than the nominal area of contact (at the macroscopic level) at any given instant. It is suggested that the area of the lower partition (valleys) be defined such that it is comprised of the largest area that never comes in direct contact with the opposite surface. The upper partition (peaks) is then defined as the set of points that are sufficiently tall to be candidates for contact. The area that consists of these points will be referred to as the *area of intermittent contact*. Its size will then depend on the surface profile, mechanical properties, and contact severity.

5.3. *Area of Intermittent Contact*

There is a large body of work dealing with statistical contact models for nominally flat surfaces. Since the early work in this area, it has been possible to approximate the size, number, and force of asperity contacts between two surfaces with reasonable accuracy [Greenwood and Williamson, 1966]. Such models provide a useful foundation for more complicated tribological modeling, and are used within the present modeling effort as well. However, many of these models are simplified by the assumption that two rough surfaces behave the same as a composite rough surface against a perfectly smooth surface. This may be true when estimating certain mechanics, but it does not address the needs of the present work because it implies that the same tall peaks are in continuous contact with the opposite (smooth) surface. In reality, even tall asperities are occasionally in a non-contact state as they pass over deep valleys. In other words, the real area of contact may be 5%, but it is not always the same 5% of either surface. The traditional contact models provide a snapshot of a sliding contact, but actual sliding involves a larger area of intermittent contact whose points continuously and rapidly flash in and out of direct contact with the opposite surface. When defining the upper surface partition, it is desirable to capture this area of intermittent contact, as it represents the complete set of peaks that are exposed to direct contact. In the case of antiwear films, it represents the set of thick "pads" that we wish to distinguish from the valleys.

Consider two surfaces a and b. The intermittent contact area of surface a, $A_{i,a}$, must be at least as large as the real contact area, A_r, and no larger than the nominal contact area, A_n. For a given degree of separation and roughness, all points that are tall enough to be candidates for contact are in the intermittent contact area of size A_i, by definition (the upper surface partition). All points that are never in contact are in the valley partition of size A_n-A_i (the lower surface partition). The real area of contact is given by the points at which the upper partition of one surface coincides with the upper partition of the opposite surface. Now consider an arbitrarily selected point within the nominal contact area. The

probability that the point falls within the intermittent contact area of surface a is equal to $A_{i,a}/A_n$. The probability that the point falls within the intermittent contact area of surface b is equal to $A_{i,b}/A_n$. Then the probability of contact at this point is then equal to the joint probability given by:

$$P[a \in i, b \in i] = \left(\frac{A_{i,a}}{A_n}\right)\left(\frac{A_{i,b}}{A_n}\right) \tag{20}$$

If the surfaces have identical roughness profiles, then the magnitudes of $A_{i,a}$ and $A_{i,b}$ are the same, and so the probability of an arbitrary point being in the fraction of real contact, A_r/A_n, is equal to:

$$P[a \in i, b \in i] = \frac{A_r}{A_n} = \left(\frac{A_i}{A_n}\right)^2 \tag{21.a}$$

Equivalently,

$$\frac{A_r}{A_i} = \frac{A_i}{A_n} \tag{21.b}$$

Note that as A_r approaches zero or A_n, A_i also approaches zero or A_n, respectively.

For two surfaces of identical roughness profiles, the ratio of intermittent to nominal contact areas can be approximated by the square root of the ratio of real to nominal contact areas. Therefore, if the severity of contact known in a given application, then the maximum area of real contact can be estimated using a traditional asperity contact model. The area of intermittent contact can then be estimated from the maximum real area of contact. The upper partition is then defined as the tallest x percent of the surface, where x is the ratio of intermittent contact area to nominal surface area.

If greater resolution is needed, then additional surface partitions can be implemented, and multiple areas of intermittent contact would be present within one another (to be used for varied degrees of contact severity).

5.4. Temporal Sub-Partition of the Reactor Volume

In order to model the chemical reactions and other effects, one must determine the duration of exposure to a given temperature. Each surface partition will have a unique temperature profile. Tall asperity peaks will undergo severe contact, with high flash temperatures, while the low surface valleys may never make contact at all. Flash temperatures can be very high, and are believed to play a role in the surface chemistry. Asperity contacts are very brief, but they are also very frequent.

It is necessary to estimate the overall fraction of time that a point on the upper surface partition actually spends in contact with the opposite surface. Then, when an average flash temperature is calculated for a given time step, it will only be applied to the upper surface partition, and only for the duration of actual contact. It would be computationally infeasible to handle each asperity contact event separately, so they are lumped together for each time step. Chemical reactions will be calculated twice for each partition, once for the cooler, non-contact state, and again for the warmer, contact state (for any partitions that make contact). In order for the force balance between the surfaces to remain consistent between the instantaneous and sliding cases, the fractional duration of real contact must be equal to the ratio of real to intermittent contact areas. This can be shown in terms of conservation of momentum:

$$\overbrace{\underbrace{P_{nom}A_{nom}t_{nom}}_{\text{macroscopic}} = P_{asp}A_{real}t_{nom}}^{\text{static or instantaneous}} = \overbrace{P_{asp}A_{intermittent}t_{real}}^{\text{time averaged}} \quad \text{(22)}$$

which implies that:

$$\frac{A_{real}}{A_{intermittent}} = \frac{t_{real}}{t_{nom}} \tag{23}$$

where P is the pressure (nominal or asperity-level), A is the area (nominal, real, or intermittent), and t is the time of contact (nominal or real).

Physically speaking, the intermittent area of contact is continuously changing as contact severity changes. However, its maximum value has been chosen for the purposes of surface partitioning (allowing the user to track just two points, rather than a large grid of points). This discretization is balanced by the continuously changing time of real contact, thereby enhancing the accuracy of time and temperature dependent calculations. If more than two partitions are used, then more than two states of contact should be used, where each partition has a distinct contact duration and temperature for each degree of contact severity.

5.5. Effective Material Properties

When modeling a surface with different material layers, it is desirable to capture the important mechanical, chemical, and thermal effects that each layer contributes. In the current work, there are as many as four material layers. The bottom layer is the metal substrate, such as cast iron. Next, there may be a layer of iron oxide. The next layer is the antiwear film, if present. The outermost layer is the monolayer of adsorbed

decomposition products of ZDDP. Over the course of the simulation, these layers may change in thickness or coverage area, but always appear in the same order if present.

The properties of each material are independently stored, and then a set of effective surface properties is generated based on the composition of the surface. An effective coefficient of friction is computed as the weighted average of the friction coefficients of each material present on the surface, where the weighting is scaled according to the fractional exposure of each material (material area divided by total area of partition). Chemical reactions on the surface are also scaled according to the exposure of each material. Thermal properties, such as conductivity, are calculated as a linear combination of the material properties of each layer. Specifically, a weighted average is calculated according to the relative thicknesses of each layer (as seen in the earlier discussion of flash temperatures). Mechanical properties such as hardness and wear coefficients are a function of the outermost layer, but only if it is of sufficient thickness. For each material, a minimum effective thickness must be defined. If the thickness of a layer is between zero and the minimum effective thickness, then the material properties will be linearly interpolated between those of the substrate and those of the layer. In the case of multiple layers, the effective properties are calculated from the "bottom up" by recalculating the effective properties with each additional layer.

5.6. Underdetermined Problem of Growth

Most of the theory behind the growth and removal of the different surface materials is in the form of volumetric predictions. For example, growth may be in the form of "moles of material reacting" and removal may be in the form of "volume worn away." From a given change in volume, a change in dimension must be estimated. (Note that the final growth predicted by the chemical reaction sequence is calculated in moles, and then converted to a volume using the molecular density of the antiwear film.) The surface is treated as isotropic (with average coverage values being tracked), but there still remain two spatial dimensions, area and thickness, to resolve from one parameter, volume. This is done by

defining a minimum thickness of a material. Physically, this could be thought of as the thickness of a molecular monolayer. Any material layer whose volume cannot cover the local surface area is treated as a monolayer with constant thickness. Therefore, changes in volume are handled by increasing or decreasing its coverage area. Conversely, any material layer whose volume is greater than that of a monolayer over the local surface area is treated as having a constant area, and any changes in its volume are handled by increasing or decreasing its thickness.

5.7. Changes in Concentration from Mixing

As stated before, the reactor volume, in the most general case, is an open system. In an engine, for example, the lubricating oil is steadily replaced with oil from a much larger pool. After the chemical concentrations are updated based on chemical reactions (within the solution, as well as migration to and from the surface), they must again be updated for any mixing that has occurred with outside fluids.

If we assume that the bulk volume of oil is much greater than that being analyzed, such that the degradation of bulk oil is much slower than the time scales being modeled, then the bulk oil will act as a reservoir, holding all concentrations for incoming oil constant. In order to relate this to the concentration of reactants, we must know the residence time, t_r, of the oil in the reactor volume. From the average residence time, the concentration change (caused by fluid replacement) in a newly formed species (c_p, for example) might be calculated for each numerical time step,

$$c_{P(t)} = mc_{p(t-1)} \tag{24}$$

where m is a dimensionless coefficient between zero and one that represents the degree of mixing between the reactor and the reservoir over a time Δt. If there is zero mixing, then m equals zero and the concentrations remain, but if there is a perfect exchange of fluid over one time step, then m approaches unity and all new products are washed away

(incidentally, this under-resolution may indicate that the selected time step is too large). This can be written more generally for some concentration c_t at time t in terms of the previous concentration c_{t-1} given an incoming liquid with an existing concentration, c_o:

$$c_t = (1-m)c_{t-1} + mc_o = c_{t-1} + m(c_o - c_{t-1}) \tag{25}$$

therefore, m is the percentage of fluid replaced, which is given by:

$$m = \frac{\Delta t}{t_r + \Delta t} \tag{26}$$

In this form, the depletion and replenishment of ZDDP and other species can be approximated within each reactor volume, or between neighboring reactor volumes.

6. Execution of Numerical Model

6.1. Numerical Scheme

A simple forward Euler method was used as the predictive numerical scheme in this work. Specifically, the necessary first-order rates, or time derivatives, were calculated at each time step. Changes were then computed as the product of rate and time.

This method was chosen for its speed of implementation and execution, although certain accuracy and stability concerns must be addressed separately. The overall sequence of calculations can be summarized in a set of nested loops, as shown in pseudocode in Figure 6.

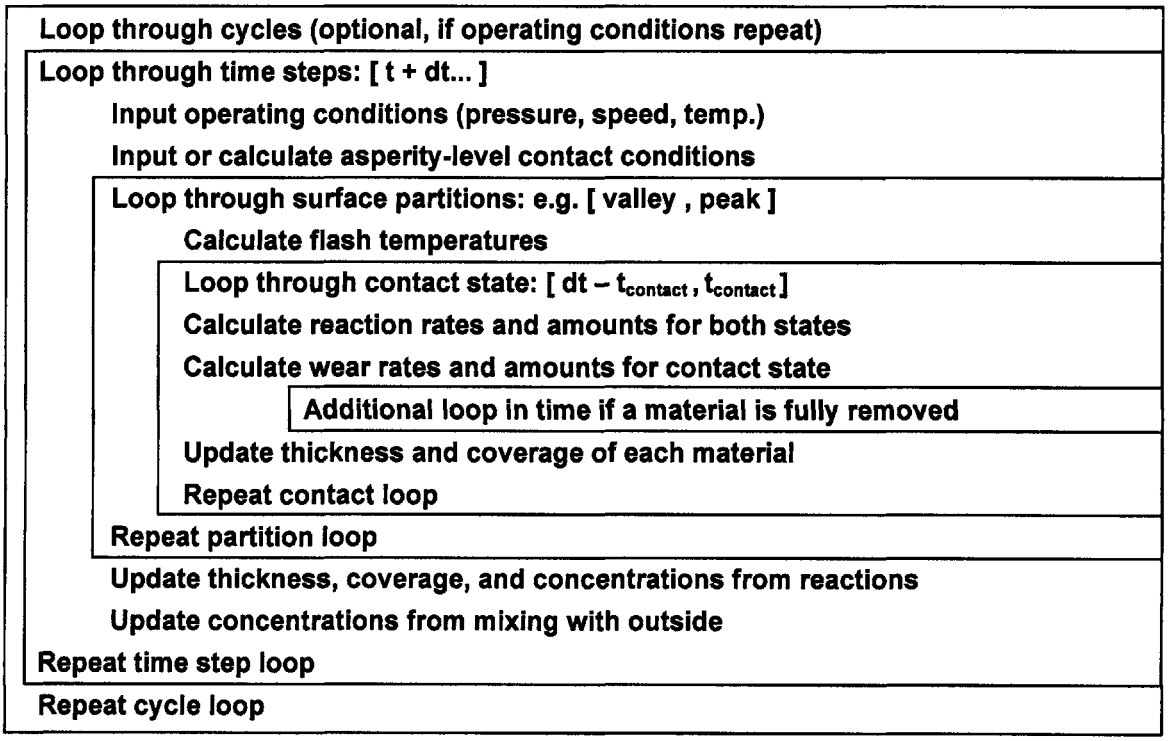


Figure 6 is a diagram of the overall loop structure of the program.

6.2. *Accuracy and Stability Considerations*

One of the main limitations of the Euler method is the potential for poor accuracy. This can generally be avoided by using a small enough time step to sufficiently resolve the nonlinearities in the evolution of the solution. In a tribo-chemical process, there exists a wide range of time scales of interest. For example, a single asperity collision might be on the order of one microsecond, while the growth of the antiwear film might be on the order of minutes or hours. The algorithm discussed above has been defined in such a way as to handle the very fine scales in an average sense so that only the slower scales need to be explicitly resolved. This allows speed and accuracy to be achieved together.

In addition to the asperity-level approximations, it should be noted that the growth of the film is defined as a set of chemical reactions that come together like an assembly line. As such, it is only necessary to resolve the slower, or potentially "rate limiting" reactions. The faster reactions do not need to be fully resolved. Instead, only their presence or absence must be determined.

Another potential limitation of the Euler method is the potential for instability. While this problem can also be mitigated using proper resolution, it will be useful to define a set of safeguards against instabilities for processes that might be under-resolved at some point during the problem execution. Again, an extremely fine resolution is not desirable because of the added computation time.

The primary source of instability in this problem, when solved using the Euler method, arises from potential violations in conservation of mass. For example, the calculated adsorption rate multiplied by the time step may predict a volume that exceeds the amount of adsorbates available. This would lead to the erroneous prediction of a negative concentration. Similarly, the product of the wear rate and time step may predict a wear volume that exceeds the volume of the surface film being worn. For these issues, a simple safeguard was implemented based on conservation of mass. In particular, when a process depends on the availability of a finite mass, two calculations are made. First, the

predicted amount is computed based on the Euler method. Second, a maximum, or limit, is computed based on the available mass. If the predicted exceeds the maximum, then the maximum is used instead. This method will have some error, as a linear approximation for depletion might be used instead of, say, a nonlinear decay toward zero (since adsorption is a function of the concentration of adsorbates). However, it is understood that the time resolution should be defined so that all processes of interest, such as rate-limiting reactions, are accurately handled.

A slight complication arises when multiple processes depend on the same mass. For example, a suspended volume of adsorbates might completely adsorb to a surface whose outermost layer is comprised of different materials. It is desirable to estimate how much of the adsorbates will migrate to each material. Since the adsorption to each material is happening in parallel with the others, a rate is computed for each based on the material and its area. If the adsorbates are depleted, then the amount that goes to each surface material will be scaled according to the normalized rates. In other words, processes occurring in parallel will be scaled according to their relative affinities.

Similarly, the wearing away of the outermost material might occur much faster than the duration of the time step. If this is the case, then it is desirable to further calculate the amount of wear imposed upon the newly exposed material below. Moreover, if the second layer is completely removed, then any wearing of the third material should be accounted for. These processes occur in series, and essentially require a smaller time step. This is handled on the fly by deducting the time step by the amount of time required to remove the first layer, and repeating the wear calculation on the redefined surface, as shown in the innermost loop of Figure 6. If a problem involves three surface layers on top of one substrate material, then there may be between one and four loops through the wear calculation in this manner. There will be some error in this method because the wear volumes are a function of the continuously shrinking material coverage of a layer that is being removed. This is why the overall time step must be chosen properly.

Once again, the above safeguards are intended to prevent instability in either parallel or series processes that are under-resolved, while maintaining sufficient accuracy in the overall computation of interest. Their activation in the program should be the exception and not the rule. If most of the processes require these modifications most of the time, then the problem as a whole is under-resolved, and a smaller time step should be used. In the context of this problem, a time step fine enough to resolve the macroscopic dynamics in a complicated contact will generally be more than adequate to handle the changing surface conditions. With proper temporal resolution, the error arising from these approximating safeguards will be negligible.

6.3. Computational Efficiency

Computational efficiency is a key priority in this work. Without an efficient approach, this type of modeling may not be feasible, much less practical. In order to conduct parametric studies of any kind, the computation speed must be high. Such studies will initially be used in combination with data to identify certain physical constants that might be difficult to isolate experimentally. This portion of the work will add to the current understanding of the theory. Once all of the relevant parameters are known with sufficient confidence, parametric studies can be used to understand the effects of independent variables, such as the lubricant formulation or component geometry. In other words, the numerical model can be used as a design tool.

Initial testing of the model as described above appears to be promising in terms of execution time. Some tests were conducted to simulate a point on the surface of a piston in a diesel engine. A time step corresponding to one crank-angle-degree was used on a desktop PC. At this resolution, a single reactor volume was modeled, and the speed was on the order of real-time. Simulations of less complicated contact dynamics can be accurately run with a much coarser resolution. For example, the simulation of a twelve hour pin-on-disk or four-ball wear experiment was executed in under five minutes.

The scaling of the computational time in this program is also favorable. For an individual reactor volume, the execution time is independent of size (and independent of the number of asperities therein). The execution time scales linearly with the number of volumes modeled. It also scales linearly with the number of time steps modeled (e.g. proportional to the overall duration modeled, inversely proportional to the length of each time step). Irregular or adaptive time stepping would also be possible, but has not yet been implemented.

7. Preliminary Results

A numerical model has been constructed based on the above principles. The model has not been entirely validated, and a number of inputs are still represented by placeholder estimates. In spite of this, the preliminary results appear to be quite promising from an initially qualitative perspective. Some examples are presented here.

The following two figures were produced by simulating the growth of thermally generated antiwear films. This type of film is produced in a hot oil bath in the absence of rubbing contact. Such a film is relatively simple to generate experimentally, and many of the variables associated with a tribological contact are removed. In other words, such a test provides a useful means of isolating some of the functionality of the model for validation purposes. Note that the film formed under higher temperatures (Figure 8) is thicker than the one formed at lower temperatures (Figure 7) after the same time duration.

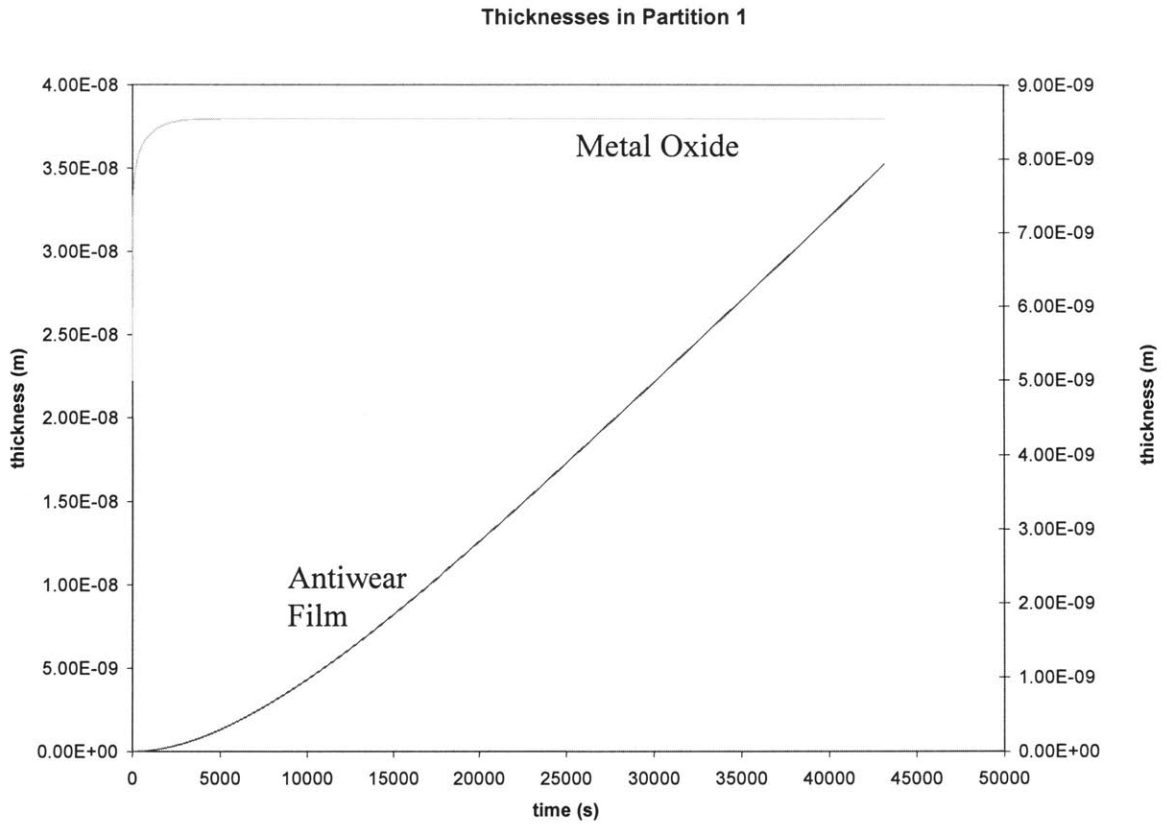


Figure 7 shows the antiwear film (left axis) and metal oxide layer (right axis) thicknesses in a thermal film simulation representing 12 hours of growth at 130°C with zero contact. A time step of 0.01 seconds was used. The initial thicknesses of the antiwear and oxide layers were 0.0 and 5.0 nm, respectively.

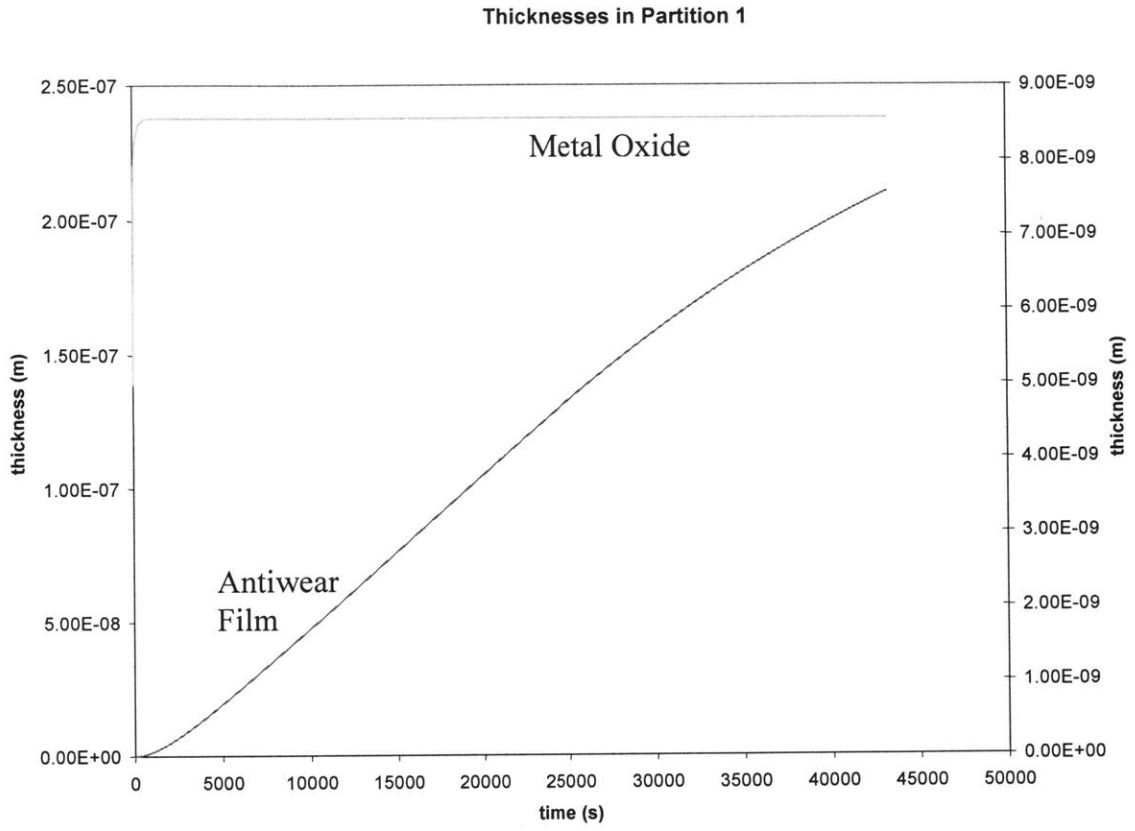


Figure 8 shows the antiwear film (left axis) and metal oxide layer (right axis) thicknesses in a thermal film simulation representing 12 hours of growth at 150°C with zero contact. A time step of 0.01 seconds was used. The initial thicknesses of the antiwear and oxide layers were 0.0 and 5.0 nm, respectively.

The next four figures were produced by simulating antiwear films grown under constant sliding contact. This is a useful test when validating both the growth and removal of the film in a tribological contact with as few complicated variables as possible. It can be compared to common bench tests, such as pin-on-disk or four-ball wear testers.

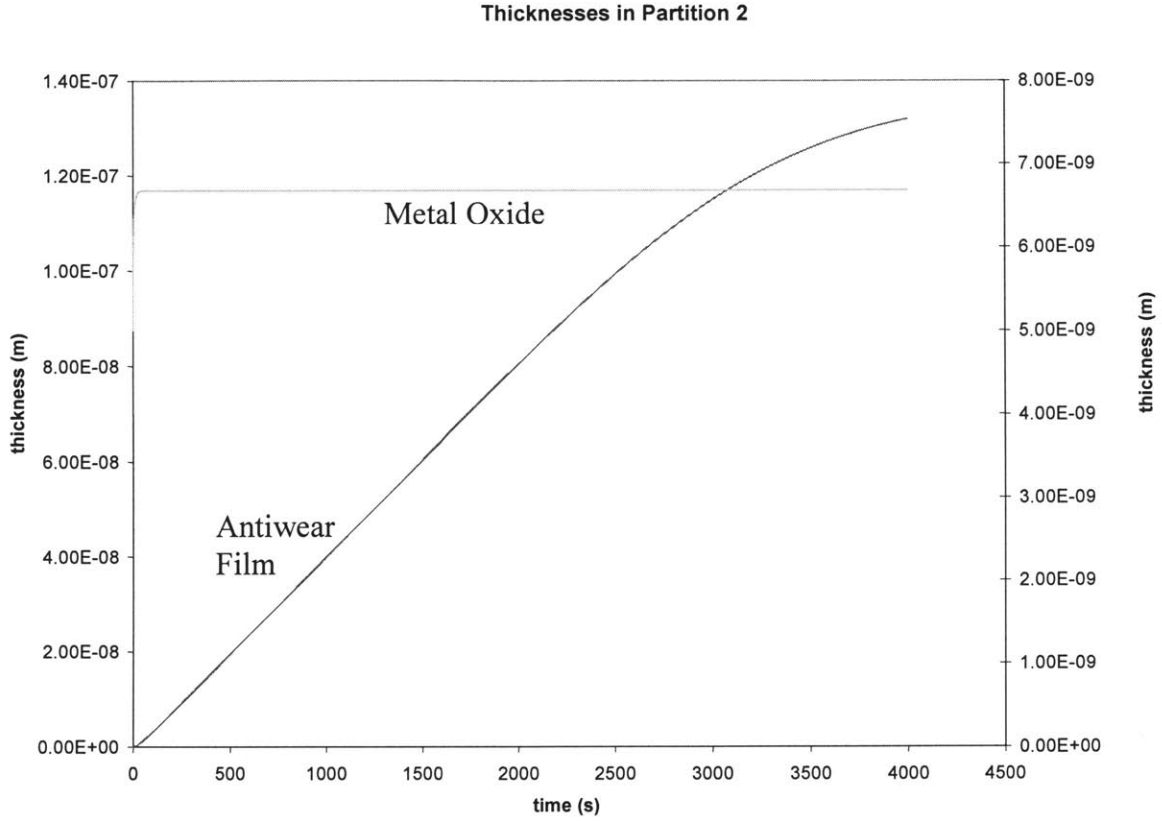


Figure 9 shows the antiwear film (left axis) and metal oxide layer (right axis) thicknesses in an antiwear film simulation representing 1 hour of growth at 80°C bulk temperature, sliding at 0.064 m/s with a mean separation of 8 nm and mean asperity contact pressure of 0.5 GPa. The surface had contacting features with a mean radius of 15 microns, corresponding to the observed size of antiwear pads. The initial thicknesses of the antiwear and oxide layers were 0.0 and 5.0 nm, respectively.

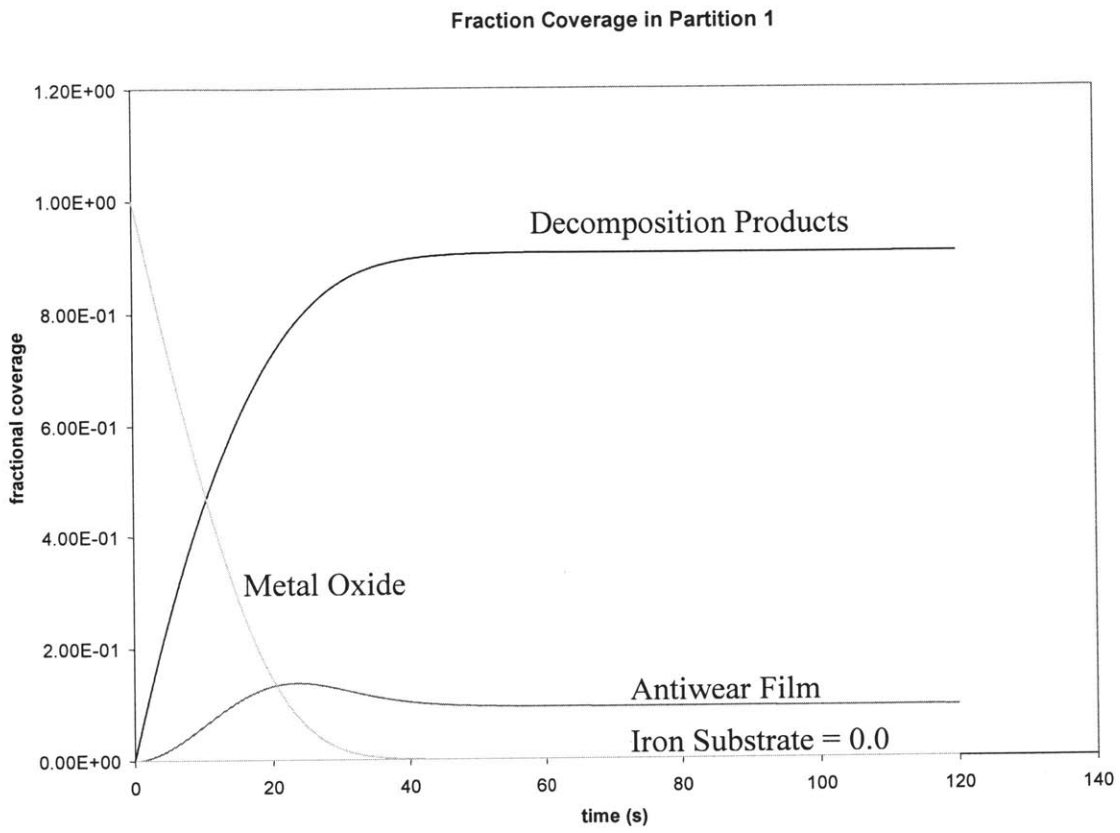


Figure 10 shows the fractional coverage area of various surface materials in an antiwear film simulation representing two minutes of growth at 80°C bulk temperature, sliding at 0.064 m/s with a mean separation of 8 nm and mean asperity contact pressure of 0.5 GPa. The surface had contacting features with a mean radius of 15 microns, corresponding to the observed size of antiwear pads. Note that the sum of fractional areas equals 1.0 at any instance in time. The initial coverage was 100% metal oxide.

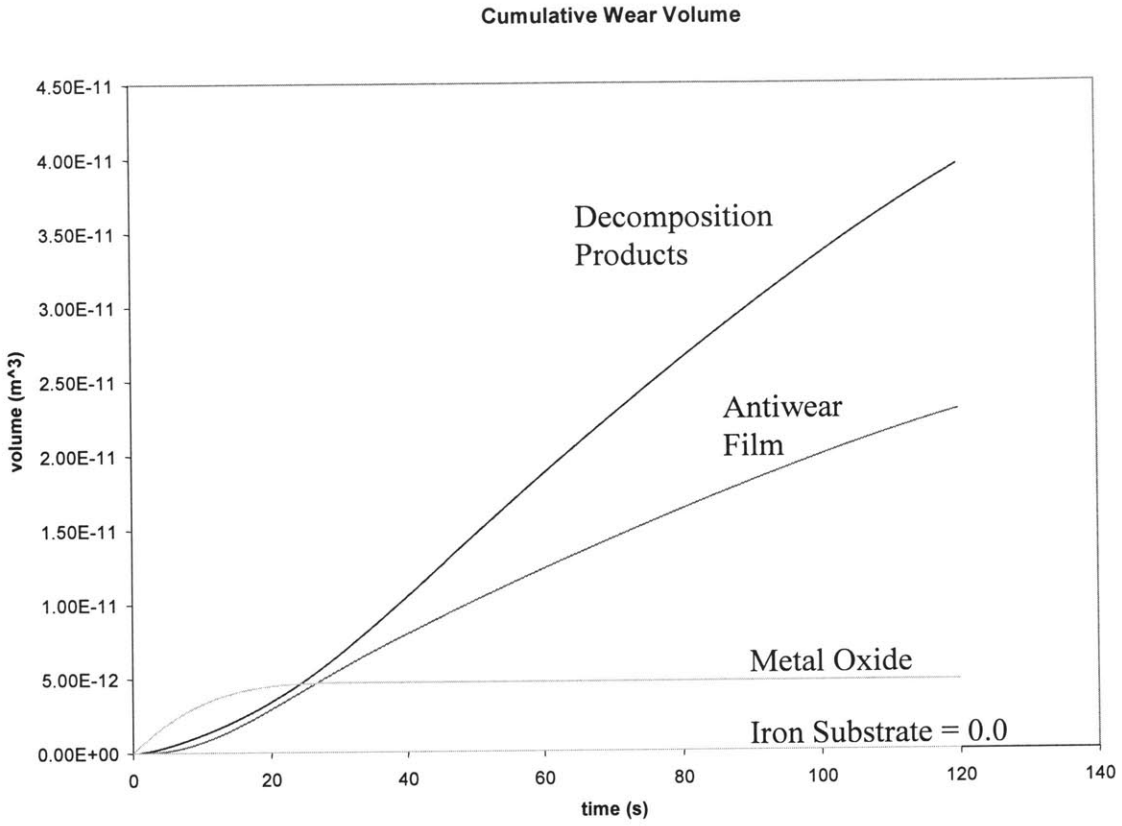


Figure 11 shows the cumulative wear volume of different surface materials in an antiwear film simulation representing two minutes of growth at 80°C bulk temperature, sliding at 0.064 m/s with a mean separation of 8 nm and mean asperity contact pressure of 0.5 GPa. The surface had contacting features with a mean radius of 15 microns, corresponding to the observed size of antiwear pads. The analysis was done for a unit-area of 1.0m^2 on an iron substrate covered with a 5 nm thick layer of metal oxide.

In Figure 12 below, the initial concentration of ZDDP is 1.5% by weight in solution in the oil. This concentration is equivalent to 37 moles per meter cubed. Because of the decomposition reaction, the concentration of ZDDP decreases as the concentration of its decomposition products increases. A steady state is reached because the system is open, with the oil steadily being replenished by oil that has the same characteristics as the initial conditions (from a reservoir). Additionally, the concentration of reactive iron cations increases early when cations are released from the surface by wear, but then decreases as they are used up in the ligand exchange reaction. They cease to be replenished when the film is developed because wear to the metal vanishes to zero.

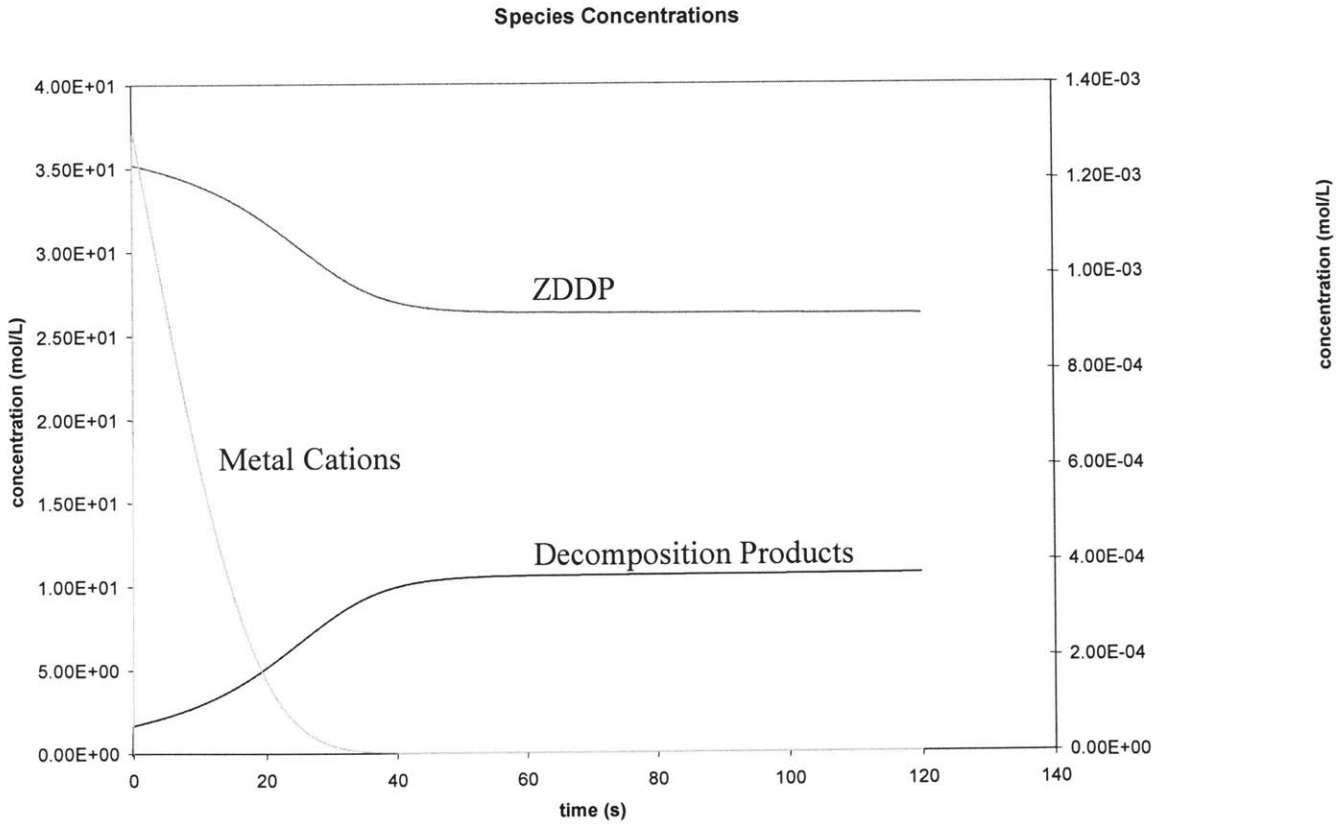


Figure 12 shows the chemical concentrations of ZDDP and its decomposition products (left axis) and of metal cations (right axis) in an antiwear film simulation representing two minutes of growth at 80°C bulk temperature, sliding at 0.064 m/s with a mean separation of 8 nm over an area of 1.0 m^2 and mean asperity contact pressure of 0.5 GPa. The surface had contacting features with a mean radius of 15 microns, corresponding to the observed size of antiwear pads.

The last two figures represent a simulation of a single reactor volume (of arbitrary area) on a piston ring sliding on a cylinder bore in a diesel engine. The complicated contact was generated using a separate program [Jocsak, 2005], and consists of a four-stroke, 720 degree cycle of pressures. This test represents the potential of the model to simulate very complex conditions, and can be validated against actual engine experiments. As illustrated in Figure 13, temperatures much higher than the bulk temperature of 100°C are predicted due to pressure and friction at the asperity level.

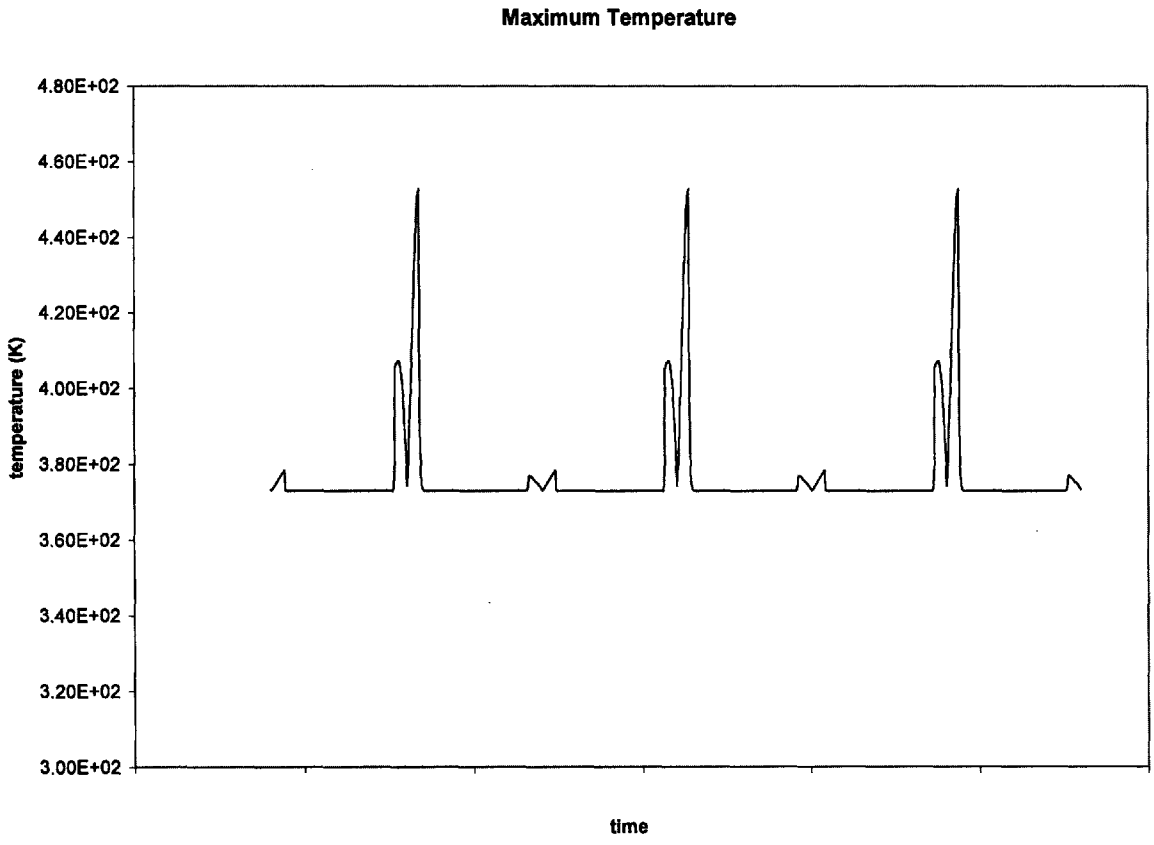


Figure 13 shows the maximum flash temperature at the surface in an antiwear film simulation representing three cycles of motion at 1500 rpm when a steady state was reached after 60 seconds of operation. The oil residence time was estimated to be 24 minutes, and a time step of one crank angle degree was used.

Under very severe contact conditions, an antiwear film may never form as the wear rate can exceed the formation rate (Figure 14). With each cycle below, the metal oxide layer starts to become covered by ZDDP decomposition products, and then the antiwear film begins to gain area, but then the high pressure event removes the outer films, exposing the oxide layer once again (as seen with each spike in the oxide coverage area). In this figure, a steady state has been reached in which a very small amount of ZDDP remains on the surface, never developing a fully protective film.

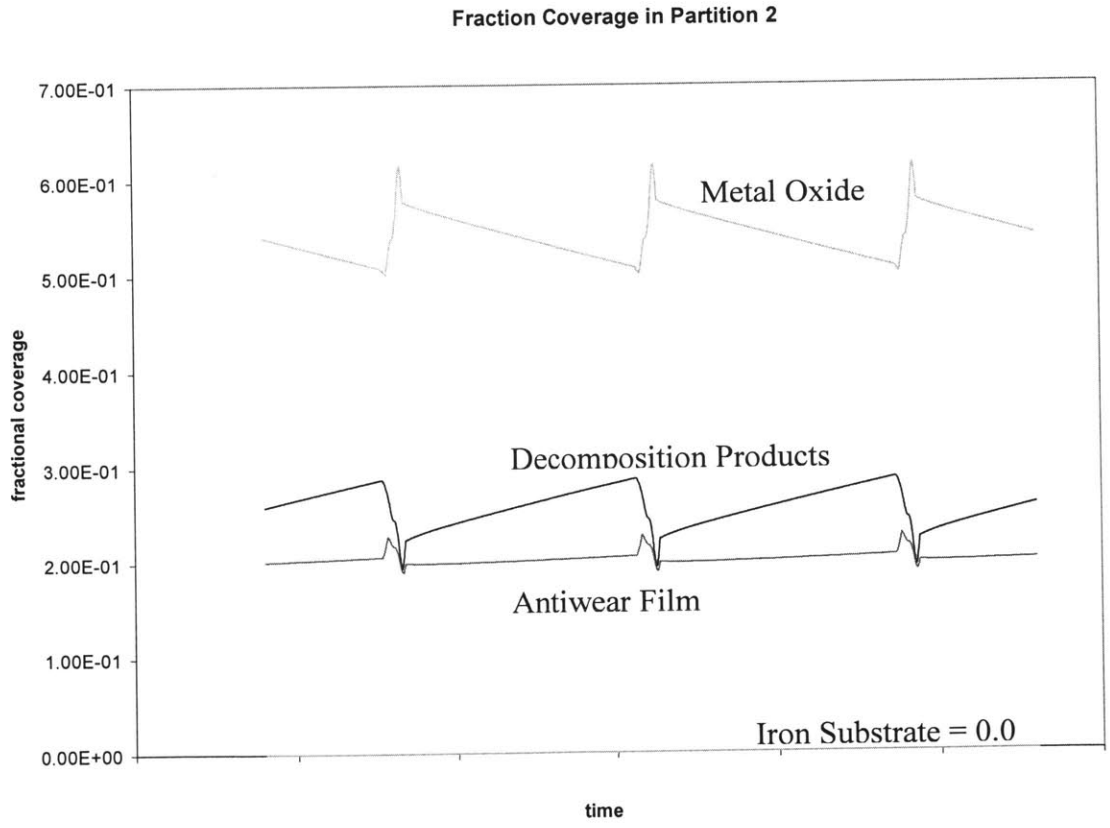


Figure 14 shows the fractional coverage area of various surface materials in an antiwear film simulation representing three cycles of motion at 1500 rpm when a steady state was reached after 60 seconds of operation. The oil residence time was estimated to be 24 minutes, and a time step of one crank angle degree was used. Note that the sum of fractional areas equals 1.0 at any instance in time.

8. Relevant Data from the Current Literature

A great deal of relevant experimental work has been published in wear literature. This work provides a useful comparison for many of the inputs and outputs of the initial modeling efforts. Some key examples are discussed below.

Figures 15 and 16, as well as Table 1 provide data regarding the growth of thermal films. Each represents a slightly different set of operating conditions including temperature, ZDDP concentration, and test duration. In Table 1, the oil bath is replaced periodically to remove the effect of ZDDP depletion. Each of these differences is useful in validating a modeling effort, but it is also difficult to directly compare work done by different researchers. In any case, the qualitative results are useful, and the experimental methods are worth considering for future work. The film thicknesses predicted by the simulation (see Figure 7 and Figure 8) fall below those predicted in the studies shown in Figure 15 and in Table 1, and above those shown in Figure 16, indicating good agreement with the available data.

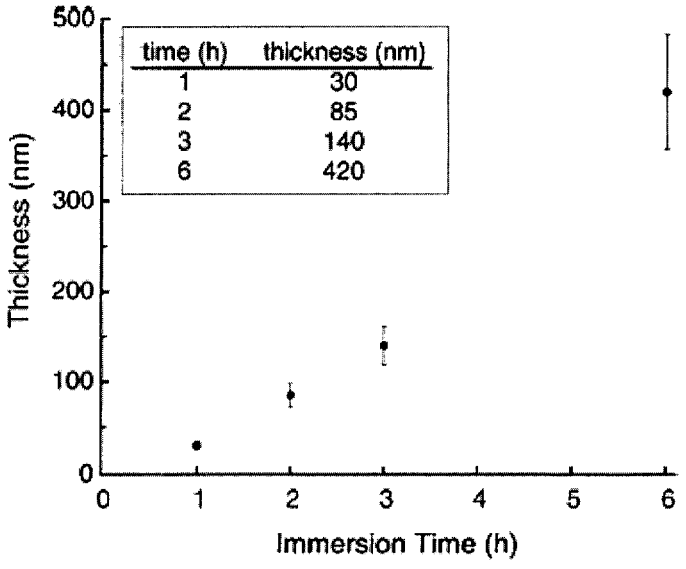


Figure 15 shows thermal film growth at 150°C with 1.48% ZDDP by weight [Aktary et al, 2001].

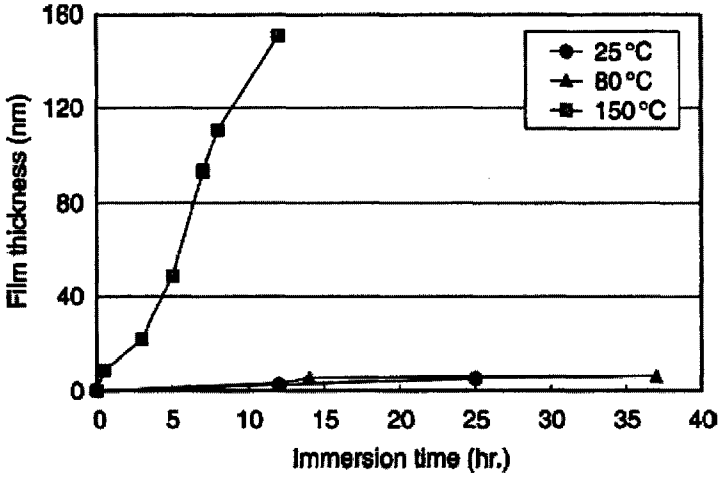


Figure 16 shows thermal films grown under various temperatures with 1.2% ZDDP by weight [Fugita and Spikes, 2004].

Film sample	P areal density[a] (atoms $\times 10^{16}/cm^2$)	Film thickness[c] (Å)	S areal density[b] (atoms $\times 10^{16}/cm^2$)	P/S
Thermal films				
8 h	1.63	110	1.09	1.5
12 h	6.62	450	1.61	4.1
24 h	12.0	820	1.80	6.7
48 h	19.2	1310	<1.0	19
72 h	31.1	2130	<1.5	21

[a] Uncertainty ±10%.
[b] Uncertainty ±30%.
[c] Assuming $Zn_2P_2O_7$ film composition.

Table 1 shows thermal film growth over long durations where the ZDDP solution is replaced every 24h [Fuller et al, 2000].

Figures 17 through 19 and Table 2 provide data regarding the growth of tribo-films that are formed under sliding contact. Again, each represents a slightly different set of operating conditions. The same independent variables present in the thermal film experiments are still present, but with the addition of contact severity, especially speed and load. Figure 17 shows the growth of tribo-films generated by two different antiwear additives. Figure 18 shows films formed (a) under different temperatures, and (b) under different degrees of contact severity (achieved by varying load, speed, and lubricant viscosity, and represented with the degree of separation given by the dimensionless variable λ). Each of these experimental films reached a steady value of around 100 nm, which is very close to the value predicted by the simulation (Figure 9). The time scale of the growth also exhibited good agreement,

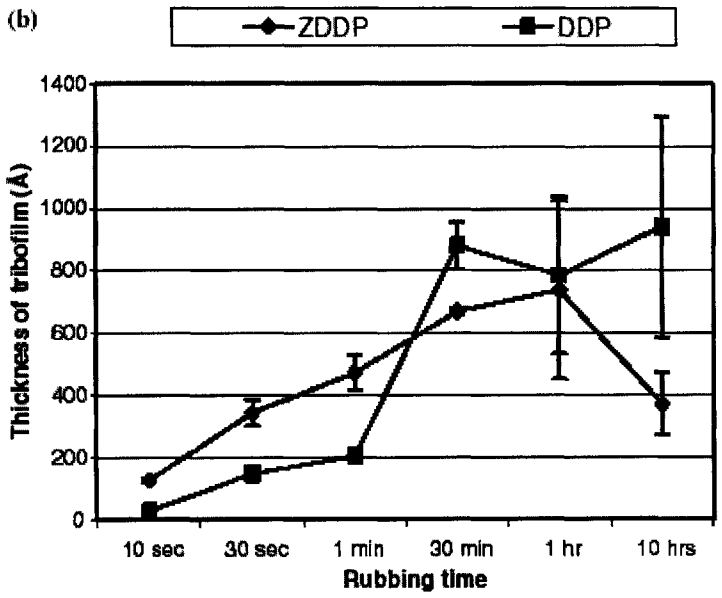
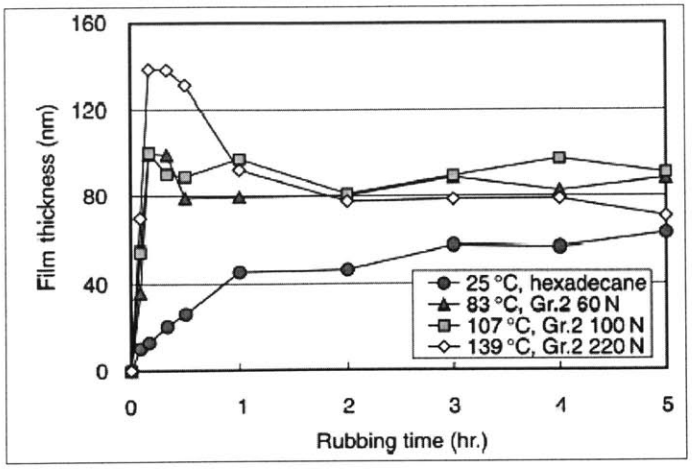
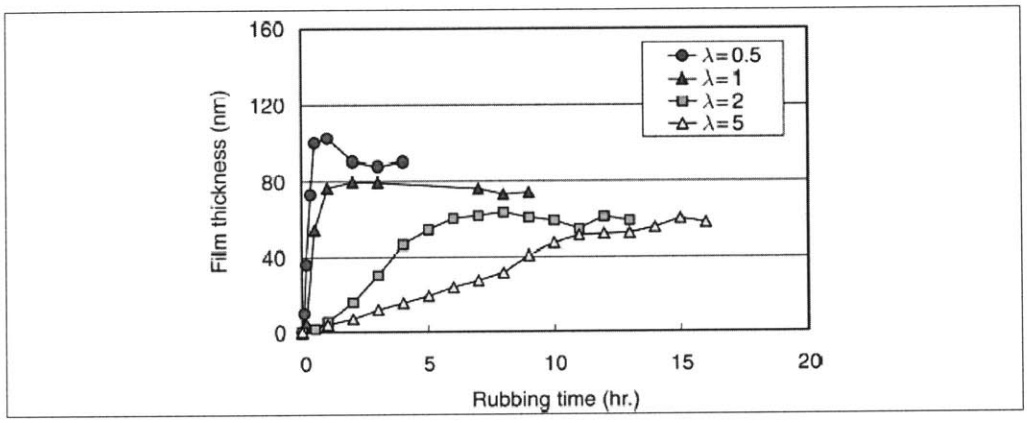


Figure 17 shows tribo-film growth from two different antiwear additives under rubbing contact [Zhang et al, 2005].



(a)



(b)

Figure 18 illustrates the temperature (a) and pressure (b) dependence of tribofilms [Fugita and Spikes, 2004].

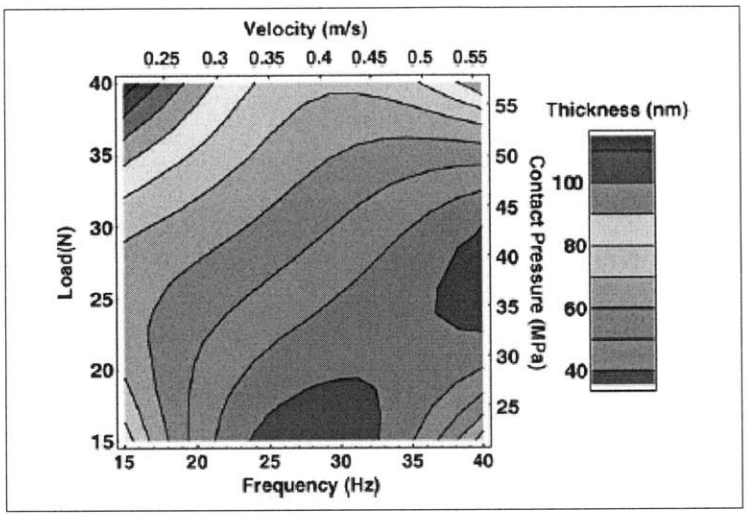


Figure 19 shows a contour plot of the antiwear film thickness after a given amount of time for various loads and speeds of sliding contact [Ji et al, 2005].

Wear test parameter set	Load (N)	Frequency (Hz)	Time (min)	ZDDP antiwear film thickness (nm)	ZDDP antiwear film average thickness (nm)	Stable coefficient of friction (μ)	Induction time[a] (s)
1	15	15	15	65.3–100.2	82.8	0.022	360
2	25	15	15	62.5–72.2	67.4	0.026	233
3	40	15	15	100.1–125.8	113.0	0.038	307
4	15	25	15	37.0–42.2	39.6	0.017	144
5	25	25	15	46.2–63.5	54.9	0.023	232
6	40	25	15	62.6–90.9	76.8	0.031	440
7	15	40	15	47.3–102.7	75.0	0.031	172
8	25	40	15	24.8–53.8	39.3	0.034	128
9	40	40	15	61.5–122.8	92.2	0.039	209
10	25	50	60	10.4[b]	10.4	0.046	107

[a] Elapsed time before contact resistance film started to build up.
[b] Based on one measurement.

Table 2 shows antiwear film thickness and friction characteristics at various loads and speeds [Ji et al, 2005].

Table 3 provides an interesting set of data related to the wear characteristics of antiwear films. Different cases are isolated, including a lack of antiwear film and a continuously replenished antiwear film (worn with ZDDP in the oil, as in most experiments), in addition to cases where a previously formed antiwear film is worn in base oil that lacks ZDDP. These cases together provide very useful information about the independent wear coefficients of the substrate and of the protective film.

Experiments	Rubbing time				
	5 min	30 min	6 h	12 h	24 h
1. coupon with no film, rubbed in base oil	283, 288	463, 582	659	–	–
2. coupon with no film, rubbed in oil + sec-ZDDP	155	158, 175	254, 344	144, 174	197
3. coupon coated with antiwear film, rubbed in base oil	–	–	165, 215	287, 320	192, 320
4. coupon coated with thermal film, rubbed in base oil	281, 313	402, 468	402	434	–
5. coupon coated with thermal film, rubbed in sec-ZDDP + oil	128, 148	149, 187	159	176	141, 157

[a] For most wear scar measurements two values are given. These refer to two independent experiments.

Table 3 provides information related to the durability of individual materials under sliding contact [Bancroft et al, 1997].

Figure 20 below provides insight into the notion of a minimum effective thickness for a thin protective film. It also provides justification for the use of a linear interpolation below such a thickness (refer to section 5.5, Effective Material Properties). The trends in the figure show that for a given load, the wear scar decreases linearly with as the thickness of the antiwear film increases, but only to a point. After a sufficient barrier has been established, wear is fairly constant for greater film thicknesses.

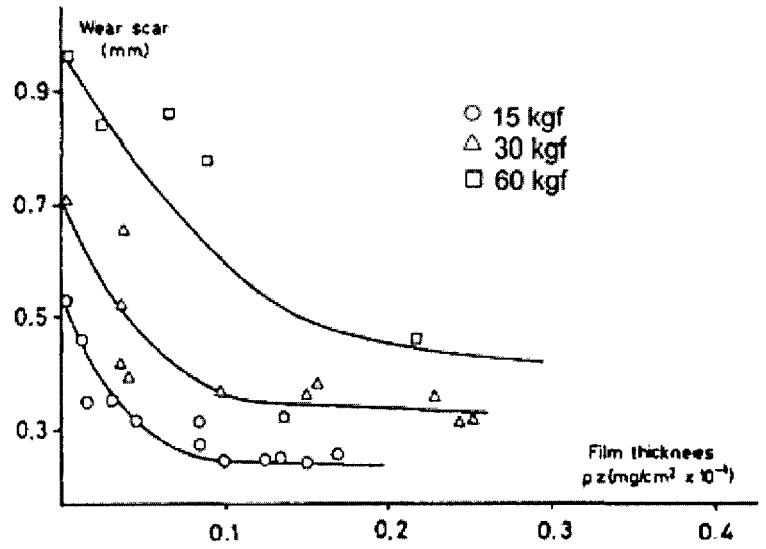


Figure 20 shows the relationship between wear and antiwear film thickness under various loads [Palacios, 1987, reprinted in Spikes, 2004].

Figures 21 and 22 show the decomposition behavior of metal dialkyl-dithiophosphates as a function of temperature. This type of data is particularly useful when defining the inputs of the model such as the Arrhenius constants.

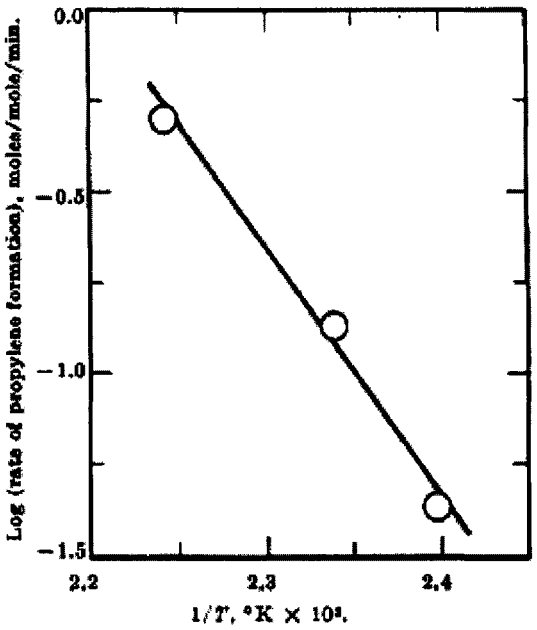


Figure 21 shows the effect of temperature on the rate of MDDP decomposition, from which the activation energy of the reaction can be inferred [Dickert and Rowe, 1966].

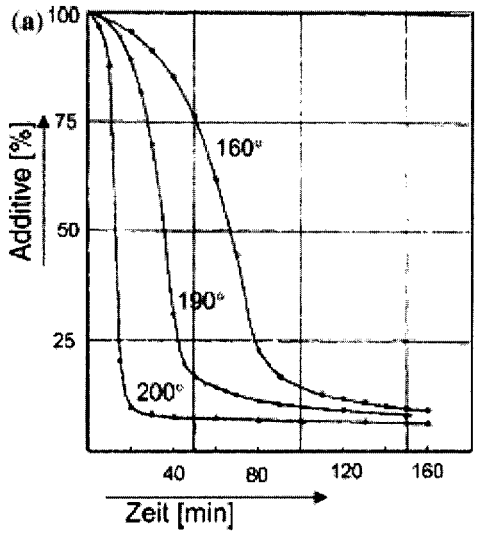


Figure 22 shows the decomposition of ZDDP as a function of time for various temperatures [von Luther and Sinha, 1964, reprinted in Spikes, 2004].

9. Future Work

The next step in the development of this model will involve a quantitative validation effort. This will begin with a comparison to published data, and continue with customized experimental work. Experimental data will be necessary to evaluate certain inputs, such as physical constants. Once these constants are known, additional data will be used to evaluate the accuracy of the model under different tribological conditions.

A formal sensitivity analysis will be conducted in order to identify the input parameters of greatest importance. For example, it is known that the growth of the film can have an exponential dependence on the value of the energy of activation terms in the chemical reactions. It is therefore important that these highly sensitive terms be known with great accuracy. Other values, such as the thermal capacitance of iron oxide, will have a smaller effect on the results of the program. Less investment is needed in finding or calculating precise values for items that have little impact. For example, it would not be worth adding an iterative solving algorithm to refine the estimate of a variable for which the final result has little sensitivity; the added refinement would be lost in the noise of the rest of the model. This would be analogous to carrying unneeded significant digits in a calculation done on paper.

After the initial validation efforts are completed, areas in need of additional improvement will be identifiable. If unacceptable error is generated as a result of poor assumptions or inaccurate approximation, then the model will be modified to more accurately represent the physical and chemical driving mechanisms (see possible improvements in Table 4). However, the preliminary results discusses in the previous section appear to indicate that the model has a potential for high accuracy under its current form.

Possible Improvement		Cost	Benefit
More detailed handling of ZDDP	Multiple species of ZDDP	low	low
	Mechanical decomposition of ZDDP	low	low
	Antioxidant behavior of ZDDP	med	med
	Consider antagonistic chemistry (surface competition)	high	med
Custom contact model	Statistical contact (number, area, pressure of contacts)	med	high
	Effect of elastic pads on pressure	med	high
	Hydrodynamic effects	high	high
More sophisticated wear prediction	Temperature dependent chemical desorption and corrosion	low	med
	Temperature dependent mechanical wear rate from softening at high temperatures	med	low
	Run-in period with surface evolution	high	med
	Discrete wear events (simulate with assumed wear geometries and probabilities)	med	med
Array of surface elements exchanging fluid		high	high
Fully generalized ability to choose number of surface area sub-partitions		high	low
Synchronization with detailed engine dynamic model (motion, pressures, temperatures, and flows)		med	high
Use model to run automated sensitivity analysis for uncertainty of inputs		med	high
Adaptive time stepping, perhaps with extrapolation from cycle loop		med	low

Table 4 lists a number of possible improvements or refinements to the existing model, as well as the perceived cost and benefit of each.

Conclusion

The mechanisms behind the formation of antiwear films have been discussed. A conceptual and quantitative framework has been laid out as the foundation for predictive modeling.

Dynamic modeling of the growth and removal of ZDDP antiwear films appears to be possible using a mechanism-based approach. Accuracy, computational efficiency, and stability can be simultaneously achieved with the proper numerical structure. The preliminary results appear promising. Upon final validation and testing of this model, it will be possible to predict antiwear film evolution and wear performance of tribological systems, such as those found in an engine. This can be a useful at the design stage of a mechanical component, the operating dynamics of a system, or the formulation of a lubricant.

The model should also shed light on the relative significance of the various theoretical mechanisms that are believed to be at work. For example, it is believed that the model will reveal that thermal energy is the primary driver of the entire formation process, and that flash temperatures and ligand exchange are the link between thermal films and tribo-films. Ligand exchange with iron cations plays a meaningful role by reducing the reactive threshold of the ZDDP molecules, while accounting for the observed compositional differences between thermal and tribo-films.

Additionally, this mechanistic framework will be useful for more general modeling purposes, including the investigation of ZDDP alternatives, or for entirely unrelated applications that involve mechanical and chemical surface interactions. The same calculations can be done with relevant pathways and reaction parameters specific to the stoichiometry of interest.

References

Aktary, M., McDermott, M.T., and Torkelson, J., 2001, "Morphological Evolution of Films formed from Thermooxidative Decomposition of ZDDP," Wear, 247, pp. 172-179.

Ashby, M.F., Abulawi, J., and Kong, H.S., 1991, "Temperature Maps for Frictional Heating in Dry Sliding," Tribol. Trans., 34, pp. 577-587.

Bancroft, G.M., et al, 1997, "Mechanisms of Tribochemical Film Formation: Stability of Tribo- and Thermally-generated ZDDP Films," Tribol. Letters, 3, pp. 47-51.

Bovington, C.H., and Dacre, B., 1984, "The Adsorption and Reaction of Decomposition Products of Zinc Di-Isopropyldiophosphate on Steel," ASLE Trans., 27, pp. 252-258.

Carslaw, H.S., and Jaeger, J.C., "Conduction of Heat in Solids," Oxford Univ. Press, London, 1959.

Colgan, T., and Bell, J.C., 1989, "A Predictive Model for Wear in Automotive Valve Train Systems," SAE Trans. J. Fuels & Lubr., S4, #892145, pp. 1221-1236.

Dickert, J.J., and Rowe, C.N., 1967, "The Thermal Decomposition of Metal O,O-Dialkylphosphorodithioates," J. Org. Chem., 32, pp. 647-653.

Fujita, H., and Spikes, H.A., 2004, "The Formation of Zinc Dithiophosphate Antiwear Films," Proc. Instn Mech. Engrs, 218, pp. 265-277.

Fuller, M. et al, 1998, "Solution Decomposition of Zinc Dialkyl Dithiophosphate and its Effect on Antiwear and Thermal Film Formation Studied by X-ray Absorption Spectroscopy," Tribol. Intl, 31, pp. 627-644.

Fuller, M.L.S., et al, 2000, "The use of X-ray Absorption Spectroscopy for Monitoring the Thickness of Antiwear Films from ZDDP," Tribol. Letters, 8, pp. 187-192.

Gangopadhyay, Arup, 2000, "Development of a Piston Ring-Cylinder Bore Wear Model," SAE, 1, 1788.

Greenwood, J.A., and Williamson, J.B.P., 1966 "Contact of Nominally Flat Surfaces," Proc. Royal Soc. London A, 295, pp. 300-319.

Grosvenor, A.P., Kobe, B.A., and McIntyre, N.S., 2005, "Activation Energies for the Oxidation of Iron by Oxygen Gas and Water Vapor," Surface Sci., 574, pp. 317-321.

Hsu, S.M., Klaus, E.E., and Cheng, H.S., 1988, "A Mechano-Chemical Descriptive Model for Wear under Mixed Lubrication Conditions," Wear, 128, pp. 307-323.

Hsu, S.M., Klaus, E.E., Cheng, H.S., and Lacey, P.I., 1994, "Mechano-Chemical Model: Reaction Temperatures in a Concentrated Contact," Wear, 175, pp. 209-218.

Ji, H., et al, 2005, "Zinc-dialkyl-dithiophosphate Antiwear Films: Dependence on Contact Pressure and Sliding Speed," Wear, 258, pp. 789-799.

Jocsak, Jeffrey, 2005, "The Effects of Surface Finish on Piston Ring-pack Performance in Advanced Reciprocating Engine Systems," S.M. Thesis, Massachusetts Institute of Technology.

Kotvis, P.V., Huezo, L.A., and Tysoe, W.T., 1993, "Surface Chemistry of Methylene Chloride on Iron: A Model for Chlorinated Hydrocarbon Lubricant Additives," Langmuir, 9, pp. 467-474.

Lim, S.C., and Ashby, M.F., 1987, "Overview No. 55 Wear-Mechanism Maps," Acta Metallurgica, 35, pp. 1-24.

Masel, R.I., "Chemical Kinetics and Catalysis," John Wiley & Sons, Inc., New York, 2001.

Missen, R.W., Mims, C.A., and Saville, B.A., "Chemical Reaction Engineering and Kinetics," John Wiley & Sons, Inc., New York, 1999.

Nicholls, Mark A., et al, 2005, "Review of the Lubrication of Metallic Surfaces by Zinc Dialkyl-Dithiophosphates," Tribol. Intl, 38, pp. 15-39.

Priest, M., and Taylor, C.M., 2000, "Automobile Engine Tribology - Approaching the Surface," Wear, 241, pp. 193-203.

Rudnick, Leslie R., "Lubricant Additives," Marcel Dekker, Inc., New York, 2003.

So, H., and Lin, Y.C., 1994, "The Theory of Antiwear for ZDDP at Elevated Temperature in Boundary Lubrication Condition," Wear, 177, pp. 105-115.

Spikes, H., 2004, "The History and Mechanisms of ZDDP," Tribol. Letters, 17, pp. 469-489.

Suh, Nam P., "Tribophysics," Prentice Hall, Inc., New Jersey, 1986.

Tomanik, Eduardo, and Nigro, Francisco E.B., 2001, "Piston Ring Pack and Cylinder Wear Modeling," SAE, 1, 572.

Willermet, P.A., et al, 1995, "Mechanism of Formation of Antiwear Films from Zinc Dialkyldithiophosphates," Tribol. Intl, 28, 177-187.

Williams, J.A., 2004, "The Behaviour of Sliding Contacts between Non-conformal Rough Surfaces Protected by 'Smart' Films," Tribol. Letters., 17, pp. 765-778.

Zhang, Z., Yamaguchi, E.S., Kasrai, M., and Bancroft, G.M., 2005, "Tribofilms Generated from ZDDP and DDP on Steel Surfaces: Part I, Growth, Wear, and Morphology," Tribol. Letters, 19, pp. 211-220.

Zhu, D., and Hu, Y.Z., 2001, "Effects of Rough Surface Topography and Orientation on the Characteristics of EHD and Mixed Lubrication in both Circular and Elliptical Contacts," Tribol. Trans., 44, pp. 391-398.

Nomenclature

Symbol	Description	Units
'	"prime" indicates property of second surface	-
a_i	thermal diffusivity (e.g. steel)	m^2/s
A	pre-exponential factor in Arrhenius equation	varies
A_i	area of intermittent contact	m^2
A_n	nominal area of contact	m^2
A_r	real area of contact	m^2
c_{Fe}	concentration of iron cations	mol/L
c_i	heat capacity of material *i*	J/kg·K
c_{Ox}	concentration of oxidative reactants	mol/L
c_P	concentration of decomposition products	mol/L
c_z	concentration of ZDDP	mol/L
dt	length of numerical time step	s
eff	subscript denoting composite or "effective" property	-
E	activation energy in Arrhenius equation	J/mol
h_i	local thickness of material *i*	m
h_{nd}	non-dimensional film thickness	-
H_i	Vickers hardness of material *i*	GPa
i, j	subscripts indicating the type of material	-
k	proportionality constant in Arrhenius equation	mol/s
K_i	thermal conductivity of material *i*	W/m·K
l	equivalent linear diffusion distance	m
L	load	N
m	mixing coefficient	-
n	power of T	-
p	heat partition function between surfaces	-
q	density of thermal energy source from friction	W/m^2
R	ideal gas constant	J/mol·K
r_a	asperity radius	m
r	rate of reaction	mol/s
S	sliding distance	m
t	time	s

t_{nom}	nominal time (e.g. full time step, dt)	s
t_r	residence time (average) of a fluid in a space	s
t_{real}	duration of real contact	s
T	temperature	K
T_f	flash temperature	K
T_s	bulk, or "sink" temperature of surface	K
U_s	sliding speed	m/s
V	wear volume	m^3
w	power term in modified wear law	-
W_i	rate of generation of wear volume of material i	m^3/s
$\alpha_1, \alpha_2 ...$	constant power or "order" of the reaction	-
Δt	time interval or time step	s
γ	coefficient relating cations to metal wear volume	mol/m^3
θ_i	fraction of the local surface covered by material i	-
κ_{ij}	abrasive wear coefficient between materials i and j	-
μ	friction coefficient	-
ρ_i	density of material i (e.g. cast iron)	kg/m^3