Corrosion, Environmental Degradation and Surface Engineering
Prof. Harish Hirani
Department of Mechanical Engineering
Indian Institute of Technology, Delhi

Lecture – 11
Role of fatigue & fracture mechanisms in wear debris formation

Hello and welcome to lecture 8A, which is an extension of lectures 7 and 8. We have covered a number of wear mechanisms, and this topic of lecture 8A is the role of fatigue and fracture mechanics in wear debris formation. So, these mechanisms that we are studying for fatigue and fracture, in my view, lead to the formation of wear debris. It is a bit different, which is why I felt that maybe we should have a complete lecture for this purpose. Now when I think about fatigue, what comes to mind is something like a reversible load, something like a reversible stress, and that is why it has been shown over here that in this case, we are showing stresses, and then here we are showing the reversible also. It is something like there is a maximum, then a minimum, then a maximum, then a minimum, and then the stress magnitude varies between the del sigma ($\Delta\sigma$).

We have the maximum value of the stress, and we have the minimum value of the stress. Now this minimum value can be 0 also; I can say σ minimum can be 0 or σ minimum can be negative of maximum also. So, there is a tensile and compressive stress that is possible, but it comes in a natural way when we think about fatigue. And then we have four known queries to explain the fatigue process.

The S-N diagram, or S-N criteria, is well known. Sometimes, instead of going for the stress and its number of cycles, we think about the strain life cycle, the strain and number of cycles, and what we are already entitled to, or as we say, the fracture. So that is a fracture mechanics point: fatigue causes some sort of brittleness in a material, and that brittleness initiates, or maybe propagates, a crack. So we should think from a fracture mechanics point of view. The last one in this case is a probabilistic approach.

What is the meaning of a probabilistic approach? There will be a number of problems in a material, and whatever the design way we are doing, there will be some sort of loading-related probabilistic approach. So inherently, in a probabilistic approach, we assume everything is dynamic; everything is a reversible kind of thing. There is a load and unload continuously happening. So that is what we understand from a probabilistic approach. We consider this to be a completely statistical approach.

Another title indicates that, in my view, I assume that this mechanism really can explain wear in a much better manner. Fatigue and fracture: these two mechanisms will be able to explain the wear phenomena in a much better manner, the reason being that these two phenomena are really trying to separate wear debris from the parent surface. So that is why these two mechanisms play an important role in my thinking. Another point that comes up when I talk about fatigue is something like this. Is it surface fatigue, or is it subsurface fatigue? In my earlier lecture, I mentioned that it may be 70 to 80 percent (70% to 80%) times the fatigue starts from a surface because of some sort of discontinuity.

However, there is also a possibility of subsurface cracks or subsurface fatigue. We will explain to some extent in this lecture as well. Another one that I mentioned is that fatigue often causes the brittle factor. It is even the ductile material because, because of the repetitive cycles, the material behaviour changes, the brittle layers increase, and it causes the fracture. So, we will deliberate to some extent on that also.

Instead of referring to the stress of the load, we sometimes describe it as sliding. It is quite possible that during this sliding, some form of contact occurs, which then causes the load. This contact allows us to represent the fatigue phenomenon in terms of reciprocal sliding, or even unidirectional sliding.

Let me give you an example using my hands. Imagine I have two hands, and I start sliding one hand over the other. What happens is that asperities on my left hand make contact with asperities on my right hand.

Contact occurs only at the asperities, where one asperity loads another or the material itself. As the sliding continues, there will be periods of loading and unloading. This means that during sliding, both loading and unloading take place. If we know the exact magnitude of the load, we can consider load-related factors. Otherwise, we can express it in terms of sliding and relative motion. As mentioned, fracture mechanics is related to crack initiation and propagation.

The question arises: what is crack initiation? We've discussed this in a previous lecture, but in my view, cracks are always present in a material. They can exist at boundaries, due to different types of crystallization, or when alloying materials, inherently creating cracks. The key question is the size of the crack—is it in nanometers or micrometers? Do we have sufficient energy, load, and velocity to impart energy to the crack, causing it to propagate or increase in size? This is our consideration. Fatigue occurs in nearly all mechanical units, leading to localized progressive degradation.

I emphasize that this degradation is localized due to asperity contacts. In this course, we focus on this localized progressive degradation, which eventually leads to mechanical unit breakdown. As mentioned, inherent cracks exist in nearly all mechanical systems, and fracture mechanics considers how these cracks propagate. This is why we connect fatigue and fracture in our lectures. Additionally, in our lecture on failure analysis, we will discuss how these factors impact the fatigue life, component life, and system life.

Now, the science of the mechanisms of wear debris creation may be understood. In my view, we can think about wear debris creation in terms of crack initiation leading to the surface and getting separated from the surface. So, in my opinion, the fracture itself is related to the wear debris. We do not want a complete failure and then say it is a fracture. But if the wear debris is getting separated from a surface, in my view, that itself is a fracture.

If the theory can effectively explain wear debris generation, it would be highly beneficial. Our goal is to understand this so that we can focus on preventing wear debris generation. We need to determine how to minimize it or achieve ultra-mild wear to improve the component's lifespan and reduce surface degradation. These are crucial considerations.

When discussing fracture mechanics, in my view, there are many factors to consider, but two very important ones include the stress intensity factor.

What is stress intensity? Can we really minimize the stress? Surface roughness itself creates a stress intensity, or maybe it is a concentration of the stress at the location. So that is why in the earlier lecture we said that please keep a smooth surface, minimize asperities, and minimize the surface roughness. And of course, when we design a surface, when we think about the coating and surface design, we will be again deliberating on this topic. Another important aspect is crack growth rates. So if the loading is done in a way where the crack is already aligned naturally, it will grow at a at a much faster pace.

So, can we really design a surface where the crack growth can be checked? These are the two important aspects from a fracture mechanics point of view. I mentioned something about a probabilistic approach. We say the probabilistic approach assesses the component or structure's dependability and safety against fatigue failure. So, we are completely going with the probabilistic approach. We feel that we do not have all the information.

So, we should go with the probabilistic approach. Now, what we consider here is that the statistical fluctuation in material property will be there. That is why we say there is a lower limit and an upper limit of material strength, and then we come up with a mean value and some sort of deviation. That is what we present in material properties that will always be important. Similarly, there will be a loading condition, a statistical distribution of the loading condition, a max value, a minimum value, and some sort of variation, a mean value, and a plus minus of variation. So that is why these are all fatigue related.

Now, even coming to the surface profile again, we have realized that the surface will not be extremely smooth. Again, there will be some sort of statistical distribution. In fact, the friction that happens between the two surfaces is statistical. There is something that happens between the two surfaces that is also statistical, and that is why many times we say that we need to repeat and repeat the number of experiments, maybe say at least 3 or 4 experiments, to come up with the right reasons, because these are the statistical properties and we should come up with a mean value and some sort of standard deviation. These are the influencing parameters for that purpose that we need to study.

Now we have covered this topic. I am just trying to repeat so that we can now connect what we are proposing in this course. So, we have already covered fatigue, and we have also covered fracture, but in two independent ways. What was in what was covered in a fatigue? We say there is some sort of the two surfaces will be coming in a contact. So, if I say this is surface 1, this is surface 2.

Now, when they come into contact and separate, the component undergoes compression and then relief. This could involve a simple asperity, a solid mass, or another shape besides spherical. When contact occurs and then relieves, the component is compressed and then relieved, which also releases any residual stress. This can lead to crack initiation or expansion. Small cracks, which are inherent in the material, exist at the nano level.

When the loading and unloading cycles are sufficient to enlarge these cracks, their size increases, and multiple cracks may merge into a larger crack. If this larger crack reaches the surface, it can result in the formation of wear debris. This process has been explained in previous lectures. Instead of generating wear debris, we apply a normal load that induces high stress at the contact point, leading to plastic deformation.

Again, this is another important aspect. Many times, we design a component and assume that only there will be elastic deformation, but that is not correct. Initially, at least there will be some plastic deformation because asperities will come into contact and the stress level will be higher than the elastic limit. There will be plastic deformation, and that is why we say there will be plastic deformation. The question now is whether plastic deformation will persist.

It is quite possible for the surface to change its configuration or deformation. Work hardening may occur, and initial high stresses might reduce to below the elastic limit after a few cycles. Initially, high asperities or surface softness may decrease with sliding, reducing asperity height. As a result, stress distribution improves, maximum stress decreases, and a more stable state is achieved.

Initially, there may be plastic deformation, which can later convert to elastic deformation. However, if plastic deformation continues without reverting to elastic deformation, the material will experience continuous deformation, leading to subsurface cracks. Any initial nanometer-level discontinuities can grow in size, leading to crack nucleation and early grain growth as cracks form.

So that already there is a discontinuity; maybe the grains are there, but because of loading, or maybe because of the plastic deformation, they are there, and that is increasing the chances of crack growth, or maybe it is really giving some sort of initiation for the crack formation. Once the crack has started, we cannot avoid it. We say the expansion of the crack will happen due to the reversal of the stresses, or maybe sliding, or maybe loading in some way. So that is possible, and finally the wear particle will come out, and this diagram was also shown to you. We found that there is some sort of sliding in this direction, and maybe I can assume there is some sort of spherical ball on the surface, and then there are some sort of initial cracks in the surface.

There is both a normal load on the surface and sliding occurring. Due to the compressive stress and the tangential load, shear stress is generated. If the coefficient of friction is negligible, the maximum shear stress will occur below the surface. However, as the coefficient of friction increases and resistance to sliding grows, the maximum shear stress shifts towards the surface. When this happens, it is quite possible for wear debris particles to be produced. This phenomenon can result in fatigue wear or pitting on gear teeth, as demonstrated by the presence of pits on the surface.

So, in this case, it is possible. Now, coming to the fracture, we again have a similar kind of thing. There is a first step that requires some sort of nucleation. Now nucleation can start from a stress concentration, some sort of discontinuity where the stresses will be high. We know that wherever there is a discontinuity, there will be a concentration of stress, which can lead to the initiation of the crack. Once the crack is initiated, propagation will happen because of the loading situation, and finally, it will lead to failure.

But we are not considering the failure of the system completely. In this course, we are assuming only the failure of the wear particle that is coming out of the surface. So we are not going to think complete failure of the system is basically a separation of a particle from a surface, and that is what we say it is connected with fatigue and fracture. So, fatigue and fracture are causing wear debris generation, that is what we are thinking.

Now the same thing will happen. We have studied some sort of fracture, the cleavage we showed in the earlier one. There is a micro void fracture also that is the dimples were shown in earlier slide also. Even the intragranular, or IG crack fracture, has shown that there are cracks, which we are able to see when the granular grains are separated, and then the crack expands across the grains. So that is what we call an intragranular crack formation. Now we also study something like radiation and heat.

What we studied is that there is a possibility that localized heating occurs when there is localized heating, and localized plastic deformation may happen. There is a possibility that some sort of atom will escape from the surface, and because of the escaping of the atoms from the surface, we study that there will be possibilities of cavities, cracking, and swelling of the surface itself. All this is going to give us a kind of initiation of the crack, or maybe an increase in the chance of the crack initiation. Of course, there is a possibility of melting because of the excess heat. When there is excess heat naturally again when the solidification happens, there will be irregularities, and that will also give us some sort of chance to form a crack, which is really harmful as such.

So, we should avoid exposure to radiation or heat, and maybe we should have a control. This diagram also shows something we call the micro-flavor crack formation due to radiation. You can see here the thin lines, and then with more exposure and significant damage, you are able to see the deep lines as well. So, this is a kind of crack that is propagated to the surface and is causing the failure, and in my view, all of these of these are leading to or related to fatigue and the fracture. Now it can be low-cycle or high-cycle fatigue.

Variations in load, microstructure, material properties, or crystallization can all contribute to crack formation, particularly when alloying materials to improve their properties. The overall aim of this course is to understand how to identify the onset of fatigue fracture starting from atomic-level discontinuities. At the atomic level, discontinuities are present and can lead to irreversible fatigue. Initially, within the elastic limit, loading and unloading are reversible. However, once a crack initiates and begins to propagate, it transitions into an irreversible state of fatigue, which needs to be closely monitored.

There is a significant knowledge gap regarding the precise impact of microstructure on fatigue fracture formation. Theories have limitations and cannot provide complete answers on their own, nor can experiments alone. Therefore, a combination of theoretical and experimental approaches is necessary to obtain accurate results. Complete theories may exist for specific components or subsystems, but they typically require extensive experimental validation. This is why both experimental studies and theoretical understanding are essential in this field.

When we consider structures from the atomic to the molecular level, we encounter significant scale differences. For example, we might discuss nanometer-level details while dealing with components measured in millimeters. The difference in scale is immense, around 10^{-9}. Creating a theory that spans from the atomic level to the micro or millimeter level is challenging. Therefore, we need to combine different theories and perform intermediate experiments to validate and connect these scales. This approach works well for specific systems but cannot be generalized.

This limitation highlights the need for additional knowledge, as discussed in the reference material provided. When examining scale differences, the magnitude is vast. At the atomic level, we're talking about 10^{-10} meters, or 0.1 nanometers. Dislocations occur at around 10^{-8} meters, and patterns of dislocations appear at approximately 10^{-7} meters. Polycrystallization can be homogeneous or heterogeneous, at the scale of 10^{-5} meters. Finally, solid materials, where we observe plastic deformation, are at the millimeter level (10^{-3} meters).

At lower levels, plastic deformation isn't visible. Even in molecular dynamic simulations, we need to start from the atomic scale (10^{-10} meters) and extend to the millimeter level. This cannot be achieved through theory alone. We need both models and experimental work.

Microstructure-sensitive computational modeling, as referenced, considers the impact of material microstructure on crack initiation and propagation. However, this model provides only partial information because the formation of fatigue fractures is influenced by the distribution of inclusions. Without accounting for inclusions and their probabilities, the model is limited to specific materials or case studies and cannot be generalized.

Then it has some sort of dislocation. Every material under different load conditions may be forming this dislocation. If you do not have knowledge about the dislocation, then there will be problems. Even the pattern of the dislocation that has been shown here can be changed, and maybe it will be different. Another thing: what is the size of the grain? It can be a smaller size, a coarse size, or a fine size, depending on the heat treatment. It depends on how we are manufacturing or processing it.

The heterogeneity among grains, such as variations in grain size or different materials, also affects the outcome. Modeling this heterogeneity is challenging. However, if we have microstructure data, such as grain boundaries or grain orientation obtained from XRD or microscopic techniques, we can incorporate this data into the model. The model needs to assume certain patterns, perform experiments, and then integrate experimental results back into the model. Models cannot be generalized and must work in tandem with experimental setups, as shown in the slide. This approach is crucial for preventing corrosion and surface degradation.

It's important to keep these considerations in mind. Detailed modeling is challenging, so we often use a simpler approach to understand material behavior under fatigue loading. We can categorize this behavior into three categories. The first category involves plastic strain accumulation, known as plastic ratcheting. This means that after a number of cycles, we observe patterns of fractures, such as surface cleavages and beach marks, caused by high stress levels beyond the plastic limit. Continuous plastic strain buildup on the surface eventually leads to a transition from ductile to brittle fracture, resulting in failure.

This is another one: reversible cyclic plasticity. Now that you can see here that the load is decreasing, we can go with this kind of pattern. It will have a lot longer life, maybe even more cycles, compared to plastic ratcheting. Well, if we further reduce the loads, then we are getting an elastic shakedown. Elastic shakedown does not mean the stress level does not come to the plastic limit at all. It comes, but what will happen in this situation? Initially, asperities are in contact; plastic deformation will occur; imperfections will change their shape and then increase the area of contact.

When the area of contact increases naturally, the maximum stress will come down. Quite possible, maximum stress comes slightly less than the yield limit. If that happens, then we say the material is mature enough to sustain the load, and it can really come up with an endurance limit. That is what happens in the fatigue limit, where we use the word endurance limit, exactly in this phase when the elastic shakedown happens, and the stage has been crossed. However, the load is, in this case, the minimum compared to reverse cyclic plasticity or plastic ratcheting.

Plastic ratcheting is at its maximum, and failure will occur. It is quite possible that it will change from ductile failure to brittle fracture in this case. So instead of going into so much detail, we can divide it into three categories. We say elastic limit and then reverse cyclic plasticity and plastic ratcheting. However, I just want to connect this with our mechanism, for which we have studied something like a bathtub curve.

You know that this is a complete bathtub curve. We know that from A to B, the wear will be maximum, or from B to A, the wear is maximum. Now why is it happening here? The A level many asperities may be the few asperities come in a contact and stress level is very high. Now, with the sliding, what happens is that asperities will change their shape. It can be, as you know, initially very sharp peaks. It will be slightly under plastic deformation; they are getting some sort of contour at the bottom. Because of the contour, the projection angle or cone angle will increase in size, which is why the wear rate is coming down.

After that, it is coming from the elastic-shakedown kind of thing. That is what I am trying to say: the BC can be a kind of elastic shakedown. It can really sustain the endurance limit. It can really save the number of cycles in this, and finally, it can come something like in a plastic ratcheting or the area of contact is much lesser. The stress has gone again very high, and the failure occurs. So that is why I am not trying to connect the fatigue and fracture with a wear phenomenon.

Sometimes, people say these phenomena are unconnected. However, there may be a connection between elastic shakedown and running-in wear behavior. Elastic shakedown leaves residual stress on the material surface, which improves wear resistance through micro-hardening. This process can result in the formation of rounded edges instead of sharp peaks, reducing wear. This micro-hardening, sometimes described as heat treatment at the local level, increases the hardness of the material.

That explains why the wear is decreasing. Regardless of the method used to reduce the wear, Now, because of the rounding of the asperities, as we understand it, there will be a sort of load redistribution. Now, because of the load redistribution, what will happen is that maximum contact pressure will come down. This explains why the distribution of contact pressure occurs more efficiently. More asperities come into contact, and then they have a round shape at the bottom, and then they are able to distribute the load in a better manner.

Instead of a single asperity, there are many asperities coming into contact. These asperities are rounded, reducing friction and significantly lowering wear. This increases the chances of elastic shakedown, meaning the material undergoes fatigue fracture rather than immediate failure. Fatigue is crucial here. The acceptable wear coefficient is generally less than 10^(-7), meaning the material can endure many cycles of contact without significant wear. This wear coefficient helps us relate to the probability of failure or wear debris generation.

By understanding these principles, we can connect fatigue and wear, which helps control fracture according to our needs. This allows us to control wear behavior and minimize stress concentration. To illustrate this, referring back to a previous lecture (Lecture 4a, Slide 3), we showed that as the cone angle (or asperity angle) increases, the wear rate decreases. Initially, a smaller asperity angle results in less contact area and higher wear. When the asperity angle increases to 100 degrees, the contact area increases, leading to stiffer asperities and some penetration, thus demonstrating the running-in behavior.

When the penetration happens, the area of contact will increase, in this case particularly because of the smooth transition. In this case, the penetration will not be characterized by peaks but rather by rounded edges. As further increases occur, the stress level will gradually decrease. This 2a level increases the maximum stress coming down. Now, the maximum normal stress is one thing, but in addition, the shear stress will also come down in magnitude. So these are favorable situations for us, and we say that in this situation, what will happen? The cracks can be controlled; whatever the initiation of cracks may not be completely controllable, propagation can be controlled because the stresses are coming down.

So it is very important to understand the connection between fatigue, fracture, and wear. In my view, they are interrelated, and to fracture, we need to think about the separation of debris particles from a surface. We do not have to think that the complete fracture or complete separation of the material through and through thickness is not possible. We need to think about the debris that is getting separated and where the debris will be getting separated. This again I am taking the slide where you see the figure from earlier slide this figure was shown.

Whenever there is loading and unloading combined with tangential load, the top layer undergoes significant deformation. The resulting high strain can initiate separation, leading to dislocation or delamination of the layer. This can be explained using fatigue phenomena. If there is a subsurface crack, the affected sublayer might be removed or delaminated. Therefore, it's crucial to minimize surface deformation to prevent this separation.

The cells are oriented vertically, while the top layers are aligned with the sliding direction. This perpendicular orientation increases dislocations, which in turn increases the chances of crack initiation and propagation. To address this, we need to focus on minimizing stress in the top layer and preventing its separation or delamination, which leads to wear debris. This topic will be continued in our next lecture. Thank you.