

TIME'S ARROW

by John Moubray

1 Introduction

Maintenance has been evolving steadily as a separate management discipline for the last sixty years or so. A remarkable feature of this evolution has been the absence of a clear understanding of or exposition of any sort of scientific basis for maintenance. As a result, it could be said that right now maintenance is literally “without foundation”. This may be one of the major reasons why so many maintenance departments still struggle to find their true place in organisations that regard maintenance as an expensive overhead that does not provide a satisfactory return on what it costs, and treat it accordingly.

In fact, such a scientific basis does exist. Not only does it exist, but it has been called “the biggest, most powerful, most general idea in all of science”. It is the Second Law of Thermodynamics.

This paper provides a brief overview of this scientific principle and shows how it clarifies many of the apparently contradictory and sometimes counter-intuitive issues now facing people who wish to formulate cost-effective maintenance strategies. The paper also shows how the Second Law demonstrates that, far from being an expensive irritation to be “designed out” wherever possible, maintenance is and will remain a vital and fundamental part of the fabric of modern industrial management for the foreseeable future.

2 The Second Law of Thermodynamics

Lambert¹ defines the Second Law of Thermodynamics as follows:

Energy spontaneously tends to flow only from being concentrated in one place to becoming diffused and spread out.

For example, when a red hot steel bar is removed from a furnace, it will cool down. This happens because the thermal energy in the bar flows out into the atmosphere. What happens to the energy level in the bar is shown in Figure 1.

The energy in all types of systems tends to dissipate in this way (unless, as we see shortly, something prevents it from doing so). For instance, when the fuel supply to a gas turbine is shut off, the rotor slows to a stop as its kinetic energy dissipates. Parachutists drift to earth much more slowly than they would without a parachute, because much of the potential energy that parachutists start with is dissipated as the parachute pushes air aside during the descent.

Most people would regard these phenomena as perfectly “natural” because they fit in with what we observe for ourselves. In our daily existence, we observe thousands of such

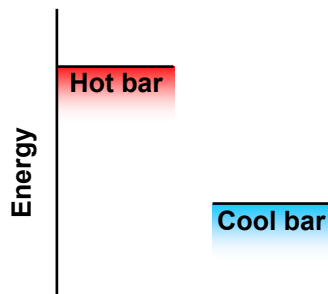


Figure 1

examples of energy being dissipated in accordance with the Second Law.

Our intuitive grasp of the “rightness” of the Second Law – based on endless amounts of personal experience – underpins our psychological sense of time. We sense time passing in the same “direction” as energy spreads out, which is why the Second Law is often referred to as “Time’s Arrow”.

(Note that we would regard it as completely “unnatural” if any of the events discussed above happened in reverse – if the iron bar spontaneously started drawing heat from the atmosphere until it became red hot, or if the turbine started spinning without fuel, or if the parachute spontaneously lifted the parachutist back to the aircraft. These events would all entail energy spontaneously becoming more concentrated rather than spreading out, and we could only imagine this happening if time ran backwards.)

3 Activation Energy

Another example of the Second Law in action occurs when anything burns (for instance, paper). Fires dissipate a great deal of energy in the form of heat and some as light. When paper burns, the cellulose in the paper reacts with oxygen to form carbon dioxide and water. The fact that this reaction produces so much energy suggests that the cellulose and the oxygen separately contain more energy than carbon dioxide and water. This is illustrated in Figure 2.

In general, chemicals tend to react if their molecules contain more energy before the reaction than the molecules formed as a result of the reaction.

If the Second Law is true, one might ask why paper doesn’t just catch fire spontaneously when it is exposed to the atmosphere. It doesn’t because the Second Law states that “energy tends to flow spontaneously ...”. The key word in this definition is the word “tends”. Energy will only succumb to the tendency to spread out if nothing stands in its way. In practice, something nearly always stands in the way, at least initially.

In the case of the paper, this “something” is the chemical bonds holding the molecules of cellulose together. Similarly, mountains don’t just collapse into a heap of sand, because chemical bonds hold the rocks together. Industrial machines don’t just fall to pieces spontaneously partly because chemical bonds hold the individual components together, and partly because fastenings – nuts and bolts, screws, welds, rivets etc – connect the components to each other.

So how does paper get to start burning? The answer of course, is by applying a naked flame. The energy in the flame is sufficient to break the bonds holding some of the hydrocarbon molecules together, in such a way that oxygen is able to

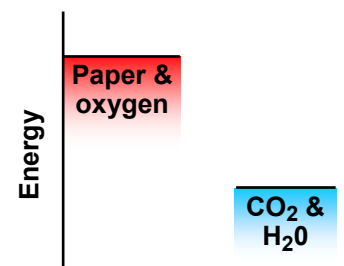


Figure 2

combine with the hydrogen and carbon atoms to form carbon dioxide and water. This reaction in turn generates more heat, – enough for the paper to continue burning on its own.

The energy needed to trigger the reaction is called *activation energy*. This is illustrated in Figure 3.

Most systems need some sort of activation energy to trigger the shift from a higher to a lower energy state. The need for this trigger protects these systems from change. (For example, if a mild steel bar is kept in a perfect vacuum and stays absolutely motionless, it will remain unchanged until the end of time. The bar will only begin to change if it is exposed to the external stresses – or activation energies – that are part of the real world.)

Another example of the Second Law in action occurs when a man breaks a wooden stick. As he applies the force needed to bend the stick, the energy level inside it builds up. In this case, energy is being transferred from the person to the stick. The stick breaks when this energy reaches a level sufficient to rupture the bonds in the stick at the breakpoint. This is the activation energy. As soon as it has broken, the energy level in the stick drops back to its previous level, – after all, two halves of a stick will make just as much of a fire as the whole stick. (The reason why almost no energy is lost in the stick itself is because very few bonds are broken at the breakpoint relative to the total number of bonds in the stick.) However, the energy level of the person will have declined by the amount needed to break the stick. This in turn means that the energy level of the whole system – person plus stick – will have declined by a similar amount, because it has dissipated in accordance with the Second Law.

So what has all this got to do with maintenance?

4 Maintenance and Failure

The maintenance function exists because things fail. In other words, if things did not fail, there would be no need for a maintenance function. So in order to establish the connection between the Second Law and maintenance, we first need to consider the relationship between the Second Law and the concept of failure.

The next section of this paper explains how the Second Law shows us that failures consist of three elements – a *failure process*, a *failure trigger* and a *failed state*. It also reviews the importance of the relationship between the initial state of a system and its failed state.

4.1 The failure process

The processes by which failures occur involve the dissipation of energy. This is illustrated by the following examples:

- *Chemical reactions:* Energy is dissipated when failures occur that entail chemical reactions, such as burning or rusting. For example, if the paper mentioned in Section 3 above was (say) a map that someone needed to read and it was reduced to ashes, it would of course become totally

illegible. This would make it a complete failure as a source of information, and we have already seen that the process by which it failed entailed moving from a higher to a lower energy state.

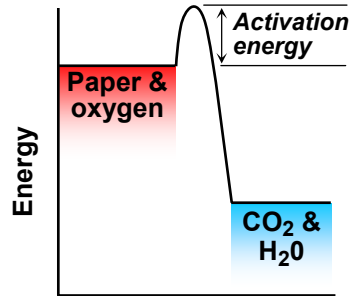


Figure 3

- *Breaking:* We have seen how energy is dissipated when things break as a result of the application of an external force, such as the act of breaking a stick. If a forklift truck smashes into a pump, the truck slows down at the moment of impact while some of its kinetic energy is used to rupture the metallic bonds that hold the pump together. During the impact, much energy is dissipated in the form of noise and heat. The failure process ends with a stationary (probably damaged) truck and a shattered pump, – a system that contains less energy than a moving truck and an intact pump.
- *Wear:* Wear entails breaking groups of atoms off solid objects. This is essentially the same process as that described in the previous paragraph, except that wear takes place a few atoms at a time. As a result, very many more bonds are broken relative to the total number of bonds in the system than is the case when something breaks at a single point. So unlike the broken stick and the bits of the shattered pump, the energy level of the wear particles and of the worn component will decline quite significantly in addition to the energy level in whatever is causing the wear.
- *Falling apart:* Energy is also dissipated when things fall apart. For instance, when flanged pipe lengths are bolted together, the act of tightening the nuts induces tension in the bolts, stretching them very slightly and clamping the flanges together. If a nut comes loose, the tension in the bolt is released (energy is dissipated) and the bolt contracts slightly. The clamping system – nut and bolt – is now failed because it no longer exerts the force that clamps the flanges together, and in failing it moves from a higher to a lower energy state.

In all the above cases, the affected systems become disorganised, and in doing so, energy is dissipated while the systems drop from a higher level of energy to a lower level of energy in accordance with the Second Law of Thermodynamics.

Note that the Second Law does not necessarily mean that in making the transition from higher to lower energy levels, systems are always broken down into smaller elements. In some cases, the operation of the Second Law entails simple systems becoming more complex, and dropping from a higher to a lower energy level in the process. For example, this occurs when the free elements of hydrogen and oxygen combine to form water, which is a more complex molecule. It could be argued that the result of this process is a more “organised” system.

However, for the purpose of this paper, let us apply the term “disorganised” to a system that is organised in some way other than it needs to be in order to function correctly (bearing in mind that this always entails the affected system moving from a higher to a lower energy state in accordance with the Second Law). This leads to two general conclusions:

- the processes by which failures occur entail the disorganisation of systems that should be organised in a way that enables them to perform satisfactorily

- the process of disorganisation entails the dissipation of energy either on the part of the system that becomes disorganised, or the system that causes it to become disorganised, or both, in accordance with the Second Law of Thermodynamics.

4.2 Failure trigger

Although the Second Law states that concentrated energy tends to spread out spontaneously, Section 3 of this paper explained (1) that more often than not, this tendency is blocked by barriers – usually bonds of some sort – that keep existing systems intact, and (2) that some kind of activation energy is needed to overcome these barriers and cause such systems to start dropping from a higher energy level to a lower energy level.

This is true of all the failure processes discussed above. The paper needed a naked flame to start burning. The pump had to be hit by a solid moving object in order to shatter. Two objects need to come into sliding contact in order to initiate wear. Some force needs to be applied to the nut in order to start loosening it (such as vibration, or alternating expansion and contraction due to periods of high and low temperature).

Let us call activation energy that initiates a failure process a “failure trigger”.

Failure triggers manifest themselves in countless other ways, such as alternating compressive and tensile stresses breaking the bonds in metallic components, causing fatigue fracture, or water freezing and thawing in the cracks in rocks, forcing the cracks to grow and the rocks – slowly but surely – to disintegrate.

Another common failure trigger is human intervention. For example, an operator might select reverse gear while a vehicle is moving forward, applying sufficient activation energy to break the metallic bonds in the gear teeth and shear them off the hub. A mechanic might apply too much torque to a nut, causing the threads to shear or the bolt to break.

Sometimes the barriers to the dissipation of energy are low enough for failure processes to take place spontaneously, without the application of any sort of activation energy. For instance, the electrical energy in stand-by batteries dissipates while they are on stand-by, albeit very slowly, until they reach a failed state. In these cases, the “failure trigger” would simply be listed as “spontaneous”.

4.3 Failed state

The above paragraphs suggest that when any system fails, what causes it to do so consists of two elements – a failure trigger and a failure process. Let us call the combination of these two phenomena – a failure trigger followed by a failure process – as a “failure mechanism”. This is illustrated in Figure 4.

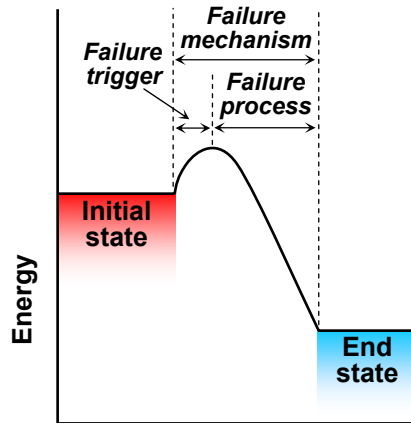


Figure 4

But what exactly is meant by the term “failed”?

Theoretically, the dissipation of energy will end when all substances in the universe have been reduced to their lowest energy state, nothing is moving relative to anything else and everything is at a uniform temperature. However, that point will only be reached a long way into the future – certainly way, way beyond the end of our working lifetimes – so it is hardly relevant to us right now.

Conversely, most systems can tolerate a small amount of disorganisation without causing any problems. An axe can be covered with a thin layer of rust and still chop wood just fine. Components such as turbine blades, piston rings, bearings, pump impellers, drill bits and crusher liners can all tolerate a small amount of wear and still perform quite satisfactorily.

So at what point does the process of disorganisation actually become relevant?

The answer to this question depends on what each system is meant to do. What any system is meant to do is determined by the people who own and/or operate it. These people will consider any such system to be failed if it gets into a state where it cannot do whatever they want it to do.

Technically, what the users of any system want it to do is defined as its “functions”. As long as a system continues to perform these functions to a standard considered acceptable by the users, the users will consider the system to be “OK”. If the performance drops below this level, they will consider it to be “failed”. So a failed state is defined as one in which the performance of a system drops below a standard that is acceptable to the users of that system, and the process of disorganisation becomes relevant when it reaches this point. This point is also known as a “functional failure”.

The failed state usually lies above the lowest (final) energy state that the system or component reaches when the failure process is complete. For example, the burning map discussed earlier will become illegible (failed from the viewpoint of anyone who wants to read it) long before the paper is reduced to ashes. This is illustrated in Figure 5.

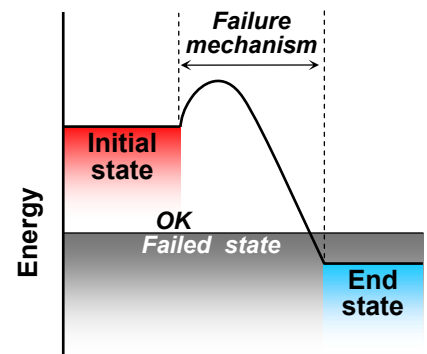


Figure 5

Since failure can only be defined in terms of the required functions of a system, a clear understanding of these functions and their associated desired standards of performance is essential before any sensible attempt can be made to analyse failures. This is illustrated by the following example.

The most obvious function of a filter element in a circulating oil system is to remove particles above a certain size from the oil. The oil pressure in this system could fail (drop below acceptable limits) because the filter has removed so many particles that it blocks up. It could be argued that the particles blocking the filter are actually in a more organised state in the filter membrane than if they are floating around freely in the oil. This might lead to the conclusion either that this failure has occurred as a result of a system becoming more organised, or even that the filter is not failed, because its function is to remove the particles. However, the filter has an

equally important second function, which is “to allow at least ... litres/minute of clean oil to pass through the filter membrane”. To do so, the membrane must have sufficiently large gaps to allow clean oil through without causing an excessive pressure drop. As these gaps get blocked by trapped particles, the differential pressure across the filter rises until the downstream pressure drops below acceptable limits. In the context of this second function, an organised system is one with large enough gaps to allow the clean oil through, and a disorganised system is one where the gaps have been blocked. (As always, this failure process involves the dissipation of energy, because the particles move from a higher energy state – moving – to a lower energy state – stationary – as they get trapped.)

4.4 Initial state

Figure 4 raises two further key points about the behaviour of systems. The first point is that for it to be possible for any system to function, the initial energy state must be above the failed state when the system is put into service. Otherwise, the system will be in a failed state right from the outset, as shown in Figure 6.

For example, consider a steel rope on a hoist intended to lift (say) 20 tons. When the rope enters service, it must be capable of lifting more than 20 tons, to allow for the deterioration that will inevitably occur when the rope enters service (mainly due to wear, a process that was discussed earlier). The attribute of the rope that enables it to be used for lifting is the chemical bonds holding the atoms of the metal together. These bonds enable the rope to transmit the tension in the rope from the hoist to the hook. If there is sufficient energy in the bonds to withstand a tension of 20 tons, the rope will perform satisfactorily. As the rope wears, its cross sectional area declines until eventually there will be too few bonds left to carry the full load and the rope will snap. However, if the cross sectional area of the rope is too small to contain enough bonds to transmit the tension when the rope enters service – in other words, if it is undersized – then it will fail immediately.

In essence, the Second Law is telling us that when any system is put into service, its initial internal energy levels must be sufficient for it to be able to fulfil the required functions. This may seem to be blindingly obvious, yet it is astonishing how often systems are encountered – or more often, sub-systems or even a single components – that are simply incapable of doing what they are supposed to do from the moment they enter service. (In the light of our earlier discussion, it could be said that such systems are not “organised” in a way that permits them to function as intended.) When this happens, something has usually gone wrong (become disorganised?) during the system design, manufacturing or installation process. Such systems break down as soon as they called upon to operate (or shortly thereafter), at which point the defect becomes a maintenance issue because maintenance people usually have to rectify it. In other words, they have to re-organise the disorganised system.

The second point concerns the size of the gap between the initial state of the system and the failed state. Clearly, a big

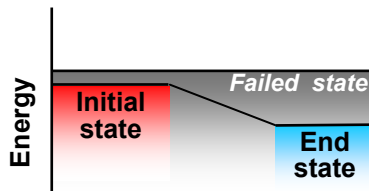


Figure 6

gap means that more energy has to be dissipated before the system descends into a failed state, so (in very general terms) it will last longer and/or fail less often than a similar system with a small gap.

For instance, in the example discussed above, a rope that is initially capable of lifting 22 tons will be able to tolerate much more energy dissipation (will last longer) than a rope made of the same material whose bonds are only capable of withstanding a tension of 21 tons to begin with.

Here the Second Law is telling us that there must be an *adequate* gap between the initial energy state of any system and whatever constitutes the failed state. This too may seem to be almost embarrassingly obvious. However, when systems fail “too soon” or “too often”, it transpires again and again that the initial gap is simply too small, so comparatively little activation energy (whether applied in a single dose or a series of tiny doses) is needed to put the system into a failed state.

5 Further Implications of the Second Law

In addition to the points discussed above, looking through the lens of the Second Law provides a number of further insights into the world of maintenance.

5.1 Age and failure

In the early days of maintenance, it was generally believed that most systems (or at least, most components) could be expected to operate reliably for a period of time, and then fail. Furthermore, it was believed that identical items performing more-or-less the same duty could be expected to fail at more or less the same age. In fact, some failures do indeed show a clear relationship between age and the likelihood of failure.

However, there is now overwhelming evidence that age-related failures are the exception rather than the rule. For example, as discussed earlier, many failures manifest themselves as soon as the affected system is put into service or very shortly thereafter because of design or manufacturing defects. Another large group of failures shows no relationship at all between how long the items concerned have been in service and the likelihood of failure (so-called “random” failures). Yet in spite of the evidence, many people inside and outside the world of maintenance still have great difficulty accepting the concept of random failure.

In fact, the concept of “activation energy” readily explains both random and age-related failures.

In the case of age-related failures, the activation energy is applied in a series of small doses. Each application lowers the internal energy of the system (what might be called its “resistance to failure”) until it reaches a failed state.

For example, when a metallic component that is susceptible to fatigue is subjected to cyclic stresses above a certain level, each stress cycle applies activation energy that weakens the bonds holding the metal together until enough of them break to cause the component as a whole to break. Identical components exposed to similar cyclic stresses (activation energies) are likely to fail after more or less the same amount of exposure – in other words, at more or less the same age.

Similar logic can be applied to failure processes like wear and corrosion, except that each small application of activation energy removes a small amount of material from the affected component until it too reaches a failed state.

On the other hand, many (most?) failures occur when a single dose of activation energy (or failure trigger) is large enough to cause the affected component to fail immediately, or very soon afterwards. The point in time at which this activation energy is applied may have nothing to do with when the affected system was put into service. So if a number of otherwise identical items is exposed to such a trigger, the likelihood that failure will occur in any one period will be same as in any other period. This gives rise to what is known as a “random” failure pattern. (Think of the forklift truck smashing the pump.)

However, some random failures do occur in situations where failure is caused by repeated small doses of activation energy. For example, properly lubricated rolling element bearings tend to fail at random, despite the fact that the failure trigger is usually cyclic stresses imposed by rollers passing over the main load-bearing section of the outer race, leading to sub-surface fatigue failure. Intuition suggests that this failure should be age-related. However, large samples of identical bearings performing more-or-less the same duty usually show little or no relationship between age and the likelihood of failure. Three of the main reasons for this are as follows:

- small defects and/or minor damage prior to or during installation lowers the initial state of one bearing relative to another, which means that a slightly damaged bearing has a smaller – sometimes a much smaller – margin for deterioration, and hence will fail much sooner than a bearing with little or no damage.
- small variations in radial load, alignment, concentricity, the presence of particles and so on greatly affect the magnitude of the activation energy that is applied in each cycle, which in turn dramatically affects the rate of deterioration of one bearing relative to another
- serial triggers which are similar in magnitude could cause one bearing to suffer much bigger changes in state than another because of minor differences in the bearing materials

Finally, situations where the initial state is below the failed state when the item is new or recently overhauled (as shown in Figure 6) give rise to the failure pattern known as infant mortality.

5.2 Terminology

At present, a major difficulty that afflicts many attempts to manage failures coherently concerns terminology. This difficulty manifests itself in two ways, as follows:

Words used to describe failures

The words we use to describe specific failures often refer to quite different aspects of failure. For example, in the context of equipment failure, the word “fatigue” calls to mind both the *failure trigger* (cyclic stress) and the *failure process* (separation of metallic bonds, leading to fracture). The same applies to words like “wear” and “corrosion”.

Other words like “break”, “shear” or “shatter” only describe the *failure process* (separation of bonds again), without giving any hint about the failure trigger. For instance, the pump casing could shatter because it was hit by the forklift truck as discussed earlier, or because the pump was massively

over-pressurised for some reason, or because a manufacturing defect in the pump casing made it incapable of containing normal pressures from the moment it entered service.

Yet another group of words tend to describe a *failure mechanism* as a whole, or even a group of failure mechanisms, without providing any information about either the failure trigger or the failure process. This group includes words like “seizes” and “fails”.

All of these terms are legitimate, so this discussion is not meant to suggest that we should stop using them. However, the Second Law provides a framework that brings much greater clarity to the meaning of the terms themselves, and also to what they mean relative to each other.

Words used to analyse failures

At this point in time, a great many terms are used in discussions about failure, such as “failure mode”, “failure cause”, “failure mechanism”, “root cause of failure”, “functional failure”, “failed state”, “potential failure”, and so on and on. Different schools of thought sometimes give different meanings to these terms, which adds to the confusion. And this is before we start talking about what could be called the by-products of failure, such as failure effects and failure consequences.

This confusion makes it very difficult for members of the physical asset management community, whether they specialise in maintenance or reliability, or both, (1) to discuss specific incidents without getting lost in thickets of confusing or conflicting verbiage and (2) to adopt universally accepted methodologies for developing failure management strategies. Perhaps the biggest single reason for this situation has been the lack of a coherent, scientific framework for considering the whole subject of failure. In fact, we have seen that such a framework does exist, in the form of the Second Law. Not only that, but looking at failures in the context of the Second Law suggests a more precise list of terms, as follows:

- *failure process*: the process by which a system makes the transition from an initial state to a failed state
- *failure trigger*: the phenomenon that initiates a specific failure process
- *failure mechanism*: a combination of a failure trigger and the resulting failure process
- *failed state*: a state in which a system is unable to fulfil a function to the satisfaction of its users
- *failure mode*: a failure mechanism or group of failure mechanisms identified at a level of detail that makes it possible to identify a suitable failure management policy
- *initial state*: the state of a system viewed from the perspective of a specific failure process, either when that system is new or immediately after it has been restored to like-new condition.

Three additional terms that apply to the analysis of failures are as follows:

- *potential failure*: a clearly identifiable phenomenon which indicates that a failure process is reaching or is about to reach a failed state
- *failure effect*: what happens when a system reaches a failed state as a result of a specific failure mode

- *failure consequence*: how and how much a specific failure mode matters.

One widely used term that does not appear in the above list is “root cause of failure”. This is so for two reasons.

Firstly, the term “root cause” implies that it is possible to “drill down” to a final and absolute level of causation when analysing failures, usually by asking “why” a number of times. In reality, finding the ultimate cause usually turns out to be impossible. What is more, it is usually unnecessary.

For instance, one might ask why the forklift truck discussed earlier hit the pump. The answer would almost certainly be because the driver drove it in that direction. Note that this action also consists of a failure trigger – turning the steering wheel – and a failure process – truck dissipating energy by moving in a direction that performs no useful work. Asking why the driver drove in the wrong direction could yield any number of answers, some relating to the state of mind of the driver, or to the configuration of the plant, or poor lighting, or whatever. Each of these answers would also involve failure triggers and failure processes. And so on and on. In fact, we are not actually drilling down, but moving sideways, from one system (smashed pump) to another (forklift truck) to another (the driver) to another (say the lighting). The point at which we stop this analytical process is not actually the root cause, but the point at which it is possible to identify a cost-effective failure management policy. In the case of the pump, this policy might simply be to move the pump to a location where forklift trucks cannot reach it, in which case, *from the viewpoint of the pump*, further analysis of the antics of forklift trucks and their drivers would be a waste of time.

Secondly, from a completely different perspective, it could be argued that all failures do indeed have a “root cause”. The discussion in Part 4 of this paper suggests that it is the Second Law of Thermodynamics.

5.3 Managing failures

Organisations acquire physical assets because they expect them to do something, – in other words, to perform a specific function or functions. We have seen that failures interfere with functions, and hence with the business processes of which the assets form part. In doing so, failures destroy value, usually by disrupting value-adding processes. Some of them breach environmental standards. A few even kill people. It also costs money to anticipate, detect, prevent or correct failures. So one way or another, failures consume time, effort and money without contributing anything.

The general uselessness of failures means that most people in industry – certainly most operations people – are hostile to them. There seems to exist a vague hope – sometimes even a belief – that if only the organisation could do something slightly different, failures would somehow go away.

In fact, failures are not going to go away, because failures are a product of the Second Law, and the Second Law is a fundamental part of the way the known universe operates. It applies to systems of every magnitude, ranging from systems of molecules, such as the sheet of paper, through industrial undertakings and geological formations to galaxies. So dealing with the results of the operation of the Second Law is a fundamental part of our existence in that same universe.

In most cases, the part of the organisation that has to deal with the impact of the Second Law on physical assets is the maintenance function. We have seen that in this context, the Second Law manifests itself as equipment failures, so the management of maintenance is all about the management of failure.

Of course, before we can set up a successful failure management program, we must determine what failures are reasonably likely to affect the physical assets in our care. As discussed in Section 4 above, an analysis of the failures that could affect any system should start with a clear definition of its functions together with the associated desired standards of performance. Defining functions clearly enables us to define how each function can fail (failed states), which then puts us in a position to identify what failures can cause each failed state. As discussed below, these “failures” should be identified in enough detail for it to be possible to identify a suitable failure management policy.

The next steps are to identify failure effects, then to assess the consequences of each failure (how failures affect safety the environment, the business process, etc.) The final step is to determine the failure management policy which deals most cost-effectively with the failure consequences.

In the minds of some people, failure management is all about failure prevention, which in turn means fixed interval overhauls or fixed interval replacements. In fact, it is now generally understood that there is far more to the management of failures than these hard-time interventions (although they are still sometimes appropriate). Other options include:

- *predictive or condition-based maintenance*, which entails checking for potential failures. Where appropriate, predictive techniques include the application of the human senses, the use of specialised condition monitoring equipment, product quality monitoring and the direct monitoring of equipment performance
- *failure-finding*, which entails checking whether hidden functions (protective devices which can fail in such a way that no-one knows they have failed) are still working
- *change the physical configuration of a system to reduce the probability of or to eliminate a specific failure process or failure trigger* (substitute stainless steel for mild steel to eliminate rust, cut a radius in a corner to reduce the likelihood of fatigue, relocate the pump to a place where the forklift truck can't hit it, etc)
- *change the behaviour of the people* who interact with the system by training them and/or by strengthening procedures (ensure that maintainers use torque wrenches where appropriate, train forklift drivers to drive more carefully, etc)
- *change the design of a physical system or the way in which it is operated* in order to reduce or eliminate the consequences of the failure.

When establishing failure management policies, it is not necessary to identify every failure process and every failure trigger that might cause every system to get into a failed state. To try to do so would not only be ruinously expensive but it also goes way past the point at which the law of diminishing returns begins to apply. It is not even necessary to identify

every failure mechanism. The key to cost-effective analysis is to identify all the phenomena which could put the system into a failed state, at a level of detail which makes it possible to identify suitable failure management policies. (“Phenomena” identified in this way were defined earlier as “failure modes”.)

Sometimes this level will be individual failure mechanisms or even failure triggers. At other times it will be groups of different failure mechanisms that could all contribute to the same failed state.

6 Conclusion

This paper has sought to make three key points:

- the Second Law of Thermodynamics provides much greater clarity than hitherto about the concept of “failure”. Specifically, it shows that any failure is not a single incident, but is actually a surprisingly complex system that embodies two steps which add up to a third (a *failure trigger* and a *failure process*, which together amount to a *failure mechanism*) and three states (*initial*, *end* and *failed*)
- the Second Law offers a foundation that could be used to consolidate the wide range of overlapping and at times conflicting failure analysis techniques currently in use around the world (RCM, RCFA, HAZOP, FMEA, FMECA, RBI and so on) into much fewer, more coherent, and universally understood processes. Perhaps even just one process.
- the Second Law provides a solid, scientific rationale for the existence of the maintenance function. The extent to which this Law governs the behaviour of the known universe in general and of physical assets in particular, means that maintenance is and will remain as much a part of the fabric of organisations that use physical assets as the assets themselves.

¹ “The Second Law of Thermodynamics”, F L Lambert, <http://www.seconclaw.com>, March 2003.

(The author would like to acknowledge the extent to which Professor Lambert’s paper influenced his thinking on this subject, especially parts 2 and 3 of this paper. Professor Lambert’s paper is strongly commended to anyone who wishes to start finding out more about the Second Law.)