MECHANICAL FAULT DIAGNOSIS PART 1

MENG 472
CHAPTER 1
MACHINERY MAINTENANCE: AN OVERVIEW

1.1 INTRODUCTION

A machine that has been designed and manufactured to perform a certain function, is expected to do so when installed in a plant for its designed life span. However, for reasons beyond one’s control, such a machine may fail to do so for several reasons. Some of the reasons could be a faulty design of the machine, inferior material and workmanship, incorrect installation and wrong operational procedure, among many others. However, in a plant where the output could be finished goods, the failure of the installed machine would lead to a loss of sales and loss in the earnings made by the plant. So, if care is not taken of the machine to avoid failures, a plant owner could incur serious financial loss and lead to bankruptcy.

Maintenance and repair of machinery in different industries process plant was defined simply as “defending machinery equipment against deterioration.” The machine thus has to be maintained to avoid such failures. All over the world, plant operators adopt three different types of maintenance techniques for machines, known as the reactive maintenance, preventive maintenance, and predictive maintenance. The benefits of planned maintenance are as follows:

- Eliminate unnecessary maintenance
- Reduce rework costs
- Reduce lost production caused by failures
- Reduce repair parts inventory
- Increase process efficiency
- Improve product quality
- Extend the operating life of plant systems
- Increase production capacity
- Reduce overall maintenance costs
- Increase overall profit

1.2 MAINTENANCE STRATEGIC AND TECHNIQUES

Four strategies within the failure fighting role of maintenance were defined:

1) **Reactive or corrective Maintenance**
2) **Preventive Maintenance (PM)**
3) **Predictive Maintenance (PdM)**
4) **Proactive-Centred Maintenance (PCM)**

### 1.2.1 CORRECTIVE OR REACTIVE MAINTENANCE

Reactive maintenance is a fire fighting approach where maintenance is performed after a failure of the asset. In this strategy, machines are operated in a **run-to-failure** (RTF) mode and the maintenance is carried out only after the functional failure of the equipment. Usually, corrective maintenance is an unscheduled maintenance action, basically composed of unpredictable maintenance needs that cannot be preplanned or programmed on the basis of occurrence at a particular time. The action requires urgent attention that must be added, integrated with, or substituted for previously scheduled work items. This incorporates compliance with “prompt action” field changes, rectification of deficiencies found during equipment/item operation, and performance of repair actions due to incidents or accidents. A substantial part of overall maintenance effort is devoted to corrective maintenance, and over the years many individuals have contributed to the area of corrective maintenance.

This type may be classified into five major categories as shown in Fig. (1.1). These are: fail-repair, salvage, rebuild, overhaul, and servicing. These categories are described below.

![Fig.1: Types of corrective maintenance.](image)

**Fail-repair:** The failed item is restored to its operational state.

**Salvage:** This element of corrective maintenance is concerned with disposal of no repairable material and use of salvaged material from no repairable equipment/item in the repair, overhaul, or rebuild programs.

**Rebuild:** This is concerned with restoring an item to a standard as close as possible to original state in performance, life expectancy, and appearance. This is achieved through complete disassembly, examination of all components, repair and replacement of worn/unserviceable
parts as per original specifications and manufacturing tolerances, and reassembly and testing to original production guidelines.

**Overhaul:** Restoring an item to its total serviceable state as per maintenance serviceability standards, using the “inspect and repair only as appropriate” approach.

**Servicing:** Servicing may be needed because of the corrective maintenance action, for example, engine repair can lead to crankcase refill, welding on, etc. Another example could be that the replacement of an air bottle may require system recharging.

So in a plant, machines that are very critical and expensive obviously cannot be left to fail by performing reactive maintenance. Usually, the less expensive and noncritical machines can be good candidates for reactive maintenance. For example, in a steel plant one obviously cannot afford to have the blast furnace under a reactive maintenance program, but perhaps a water cooler in the workers’ cafeteria may be a candidate for reactive maintenance. Following are the attributes of a reactive maintenance program:

- High expenses involved
- High spare parts inventory cost
- High overtime labor costs
- High machine downtime
- Low production availability

### 1.2.2 PREVENTIVE MAINTENANCE (PM):

Preventive maintenance strategy is a template based strategy where periodic maintenance is carried out as per OEM recommendation/ organizational recommendations. The schedule of the intervals is based on average statistical/anticipated lifetime to avoid failure. This includes inspection, service / replacement of the defective parts of the machine. PM may be described as the care and servicing by individuals involved with maintenance to keep equipment/facilities in satisfactory operational state by providing for systematic inspection, detection, and correction of incipient failures either prior to their occurrence or prior to their development into major failure. Some of the main objectives of PM are to: enhance capital equipment productive life, reduce critical equipment breakdowns, allow better planning and scheduling of needed maintenance work, minimize production losses due to equipment failures, and promote health and safety of maintenance personnel. There are seven elements of PM as shown in Fig. (1.2).
**Fig.1: 2 Elements of preventive maintenance.**

**Inspection:** Periodically inspecting materials/items to determine their serviceability by comparing their physical, electrical, mechanical, etc., characteristics (as applicable) to expected standards.

**Servicing:** Cleaning, lubricating, charging, preservation, etc., of items/materials periodically to prevent the occurrence of incipient failures.

**Calibration:** Periodically determining the value of characteristics of an item by comparison to a standard; it consists of the comparison of two instruments, one of which is certified standard with known accuracy, to detect and adjust any discrepancy in the accuracy of the material/parameter being compared to the established standard value.

**Testing:** Periodically testing or checking out to determine serviceability and detect electrical/mechanical-related degradation.

**Alignment:** Making changes to an item’s specified variable elements for the purpose of achieving optimum performance.

**Adjustment:** Periodically adjusting specified variable elements of material for the purpose of achieving the optimum system performance.

**Installation:** Periodic replacement of limited-life items or the items experiencing time cycle or wear degradation, to maintain the specified system tolerance.

**1.2.2.1 DESIGNING A PREVENTIVE MAINTENANCE PROGRAM**
Valid failure data provide the intelligence for an effective preventive maintenance program. After all, the objective is to prevent those failures from recurring. A failure reporting system should identify the problem, cause, and corrective action for every call. An action group, prophetically called the Failure Review and Corrective Actions Task Force (FRACAS), can be very effective for involving responsible organizations in both detailed identification of problems and causes and assignment of both short- and long-term corrective action. The following are typical factory and field problems and codes that shorten the computer data entry to four or fewer characters:

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOOP</td>
<td>Not Operable</td>
<td>OTHR</td>
<td>Other</td>
</tr>
<tr>
<td>BELR</td>
<td>Below rate</td>
<td>PM</td>
<td>Preventive task</td>
</tr>
<tr>
<td>INTR</td>
<td>Intermittent</td>
<td>QUAL</td>
<td>Quality</td>
</tr>
<tr>
<td>LEAK</td>
<td>Leak</td>
<td>SAFE</td>
<td>Safety</td>
</tr>
<tr>
<td>MOD</td>
<td>Modification</td>
<td>WEAT</td>
<td>Weather</td>
</tr>
<tr>
<td>NOIS</td>
<td>Noise</td>
<td>NPF</td>
<td>No problem found</td>
</tr>
</tbody>
</table>

**The typical action codes are as follows:**

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/A</td>
<td>Adjust/align</td>
</tr>
<tr>
<td>CAL</td>
<td>Calibrate</td>
</tr>
<tr>
<td>CONS</td>
<td>Consumables</td>
</tr>
<tr>
<td>DIAG</td>
<td>Diagnose</td>
</tr>
<tr>
<td>REMV</td>
<td>Remove</td>
</tr>
<tr>
<td>R/R</td>
<td>Remove and replace</td>
</tr>
<tr>
<td>R/RE</td>
<td>Remove and reinstall</td>
</tr>
<tr>
<td>INST</td>
<td>Install</td>
</tr>
<tr>
<td>INSP</td>
<td>Inspect</td>
</tr>
<tr>
<td>REF</td>
<td>Refurbish</td>
</tr>
<tr>
<td>REB</td>
<td>Rebuild</td>
</tr>
<tr>
<td>LUBE</td>
<td>Lubricate</td>
</tr>
<tr>
<td>MOD</td>
<td>Modify</td>
</tr>
<tr>
<td>PM</td>
<td>Preventive task</td>
</tr>
<tr>
<td>RPR</td>
<td>Repair</td>
</tr>
<tr>
<td>TRN</td>
<td>Train</td>
</tr>
<tr>
<td>NC</td>
<td>Not complete</td>
</tr>
<tr>
<td>NK</td>
<td>Not known</td>
</tr>
</tbody>
</table>

**Preventive maintenance design based on:**

- Equipment Data
- Equipment Types
- Failure Types

---

**The necessary items for establishing an effective preventive maintenance program**
are:
1. Every equipment uniquely identified by prominent ID number or serial number and product type
2. Accurate equipment history records
3. Failure information by problem/cause/action
4. Experience data from similar equipment
5. Manufacturer's interval and procedure recommendations
6. Service manuals
7. Consumables and replaceable parts
8. Skilled personnel
9. Proper test instruments and tools
10. Clear instructions with a checklist to be signed off
11. User cooperation
12. Management support.

**Designing a Preventive Maintenance Program**

<table>
<thead>
<tr>
<th>Preventive maintenance parameters</th>
<th>Yes</th>
<th>No</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Standardization</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>▪ Is equipment already in use that provides the desired function?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>▪ Is this the same as existing equipment?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>▪ Are there problems with existing equipment?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>▪ Can we maintain this equipment with existing personnel?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>▪ Are maintenance requirements compatible with our current procedures’?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Reliability and Maintainability</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>▪ Can vendor prove the equipment will operate at least to our specifications?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>▪ Warranty of all parts and labor for 90+ days?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>▪ Is design fault-tolerant?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>▪ Are tests go/no go’?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Service Parts</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>▪ Is recommended replacement list provided?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>▪ Is the dollar total of spares less than 10%, of equipment cost?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>▪ Do we already have usable parts?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Can parts be purchased from other vendors’?

Are any especially high quality or expensive parts required’?

**Training**
- Is special technician training required’?
- Will manufacturer provide training?
  - At no additional cost for first year’?
  - At our location as required’.

**Documentation**
- All technical manuals provided’?
  - Installation
  - Operation
  - Corrective and preventive maintenance
  - Parts

**Special Tools and Test Equipment**
- Do we already have all required tools and test equipment’?
- Can at least 95% of all faults be detected by use of proposed equipment’?
- Are calibration procedures minimum and clear’?

**Safety**
- Are all UL/SCA, OSHA, EPA and other applicable requirements met’?
- Are any special precautions required’?
- Can one person do all maintenance’?

### 1.2.3 PREDICTIVE MAINTENANCE (PDM):

Like preventive maintenance, predictive maintenance has many definitions. To some workers, predictive maintenance is monitoring the vibration of rotating machinery in an attempt to detect incipient problems and to prevent catastrophic failure. To others, it is monitoring the infrared image of electrical switchgear, motors, and other electrical equipment to detect developing problems. The common premise of predictive maintenance is that regular monitoring of the actual mechanical condition, operating efficiency, and other indicators of the operating condition of machine-trains and process systems will provide the data required to ensure the maximum interval between repairs and minimize the number and cost of unscheduled outages created by machine-train failures.

*Prepared by: Dr. Essam Bahgat*
Predictive maintenance is much more, however. It is the means of improving productivity, product quality, and overall effectiveness of manufacturing and production plants. Predictive maintenance is not vibration monitoring or thermal imaging or lubricating oil analysis or any of the other nondestructive testing techniques that are being marketed as predictive maintenance tools. Predictive maintenance is a philosophy or attitude that, simply stated, uses the actual operating condition of plant equipment and systems to optimize total plant operation.

A comprehensive predictive maintenance management program uses the most cost effective tools (e.g., vibration monitoring, thermography, tribology) to obtain the actual operating condition of critical plant systems and based on this actual data schedules all maintenance activities on an as-needed basis. Including predictive maintenance in a comprehensive maintenance management program optimizes the availability of process machinery and greatly reduces the cost of maintenance. It also improves the product quality, productivity, and profitability of manufacturing and production plants.

**Condition Monitoring Methods**

Condition monitoring is based on being able to monitor the current condition and predict the future condition of machines while in operation. Thus it means that information must be obtained externally about internal effects while the machines are in operation.

Predictive Maintenance (PdM) monitors the performance and condition of equipment or systems to detect/trend degradation. Techniques include:

1. Vibration monitoring
2. Thermographic inspection
3. Oil analysis
4. Visual inspection
5. Shock pulse
6. Ultrasonic leak detectors
7. Electrical insulation A. Megger tests B. Surge comparison
8. Performance testing
9. Wear and dimensional measurements
10. Signature analysis, time and frequency domain
11. Nondestructive testing A. Ultrasonic B. Borescope inspections C. Eddy current

<table>
<thead>
<tr>
<th>Monitoring Techniques</th>
<th>Use</th>
<th>Problem Detection</th>
</tr>
</thead>
</table>

**Prepared by: Dr. Essam Bahgat**

essbahgat@gmail.com
<table>
<thead>
<tr>
<th>Ch 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vibration</strong></td>
</tr>
<tr>
<td><strong>Shock Pulse</strong></td>
</tr>
<tr>
<td><strong>Fluid Analysis</strong></td>
</tr>
<tr>
<td><strong>Infrared Thermography</strong></td>
</tr>
<tr>
<td><strong>Performance trending</strong></td>
</tr>
<tr>
<td><strong>Electrical insulation tests</strong>, e.g., megger tests, polarization index, surge comparison testing, rotor impedance testing, DC high potential testing</td>
</tr>
<tr>
<td><strong>Ultrasonic leak detectors</strong></td>
</tr>
<tr>
<td><strong>Fault gas analysis and insulating liquid analysis</strong></td>
</tr>
<tr>
<td><strong>Protection relay testing and time travel analysis</strong></td>
</tr>
<tr>
<td><strong>Stereoscopic photography, hull potential measurements, diving inspections.</strong></td>
</tr>
</tbody>
</table>

*Prepared by: Dr. Essam Bahgat*

essbahgat@gmail.com
Material (non-destructive) testing, e.g., ultrasonic, eddy current, borescopic inspections

Hull structure, shipboard machinery and associated piping systems and mechanical components

Corrosion, erosion, fatigue cracking, delamination, wall thickness reduction

Signature analysis, time domain and frequency domain

Rectifiers, power supplies, inverter, AC and DC regulators, generators

Degraded solid state circuits and other electrical components

Wear and dimensional measurements

Sliding, rotating and reciprocating elements

Excessive wear and proximity to minimum acceptable dimensions which affect performance

The advantages of the predictive maintenance technique over the other two are many. For instance, it is economic in the long run, it provides a scope for fault prognosis, the maintenance schedule can be controlled according to the availability of resources, the spare parts inventory can be reduced, and the faults in a machine can be minimized. In the long run, this type of maintenance leads to high production rates and increased profitability. Thus the benefits of predictive maintenance can be summarized as follows:

- Lower maintenance costs
- Fewer machine failures
- Less repair downtime
- Reduced small parts inventory
- Longer machine life
- Increased production
- Improved operator safety

One of the major disadvantages of predictive maintenance is that it requires extra investment in the initial stage for the additional instrumentation and it needs a robust software for fault diagnosis and prognosis. Today, software with robust fault-detection algorithms is available, but for the initial runs, one must use them with the help of trained personnel so that no untested fault scenarios are misdiagnosed, leading to a disastrous consequence.

Selecting the predictive maintenance

Prepared by: Dr. Essam Bahgat

essbahgat@gmail.com
The first step in selecting the predictive maintenance system that will be used in your plant is to develop a list of the specific features or capabilities the system must have to support your program. At a minimum, the total system must have the following capabilities:

- User-friendly software and hardware
- Automated data acquisition
- Automated data management and trending
- Flexibility
- Reliability
- Accuracy
- Training and technical support

**User-Friendly Software and Hardware**

The premise of predictive maintenance is that existing plant staff must be able to understand the operation of both the data logger and the software program. Because plant staff normally has little, if any, computer or microprocessor background, the system must use simple, straightforward operation of both the data acquisition instrument and software. Complex systems, even if they provide advanced diagnostic capabilities, may not be accepted by plant staff and therefore will not provide the basis for a long-term predictive maintenance program.

**Automated Data Acquisition**

The object of using microprocessor-based systems is to remove any potential for human error, reduce staffing, and automate as much as possible the acquisition of vibration, process, and other data that will provide a viable predictive maintenance database. Therefore, the system must be able to automatically select and set monitoring parameters without user input. The ideal system would limit user input to a single operation, but this is not totally possible with today’s technology.

**Automated Data Management and Trending**

The amount of data required to support a total-plant predictive maintenance program is massive and will continue to increase over the life of the program. The system must be able to store, trend, and recall the data in multiple formats that will enable the user to monitor, trend, and analyze the condition of all plant equipment included in the program. The system should be able to provide long-term trend data for the life of the program. Some of the microprocessor-based systems limit trends to a maximum of 26 data sets and will severely limit the decision-making capabilities of the predictive maintenance staff. Limiting trend data to a finite number of data sets eliminated the ability to determine the most cost-effective point to replace a machine rather than let it continue in operation.
Flexibility

Not all machines or plant equipment are the same, and neither are the best methods of monitoring their condition equal. Therefore, the selected system must be able to support as many of the different techniques as possible. At a minimum, the system should be capable of obtaining, storing, and presenting data acquired from all vibration and process transducers and provide an accurate interpretation of the measured values in user-friendly terms. The minimum requirement for vibration-monitoring systems must include the ability to acquire filter broadband, select narrowband, time traces, and high-resolution signature data using any commercially available transducer.

Systems that are limited to broadband monitoring or to a single type of transducer cannot support the minimum requirements of a predictive maintenance program. The added capability of calculating unknown values based on measured inputs will greatly enhance the system’s capabilities. For example, neither fouling factor nor efficiency of a heat exchanger can be directly measured. A predictive maintenance system that can automatically calculate these values based on the measured flow, pressure, and temperature data would enable the program to automatically trend, log, and alarm deviations in these unknown, critical parameters.

Reliability

The selected hardware and software must be proven in actual field use to ensure their reliability. The introduction of microprocessor-based predictive maintenance systems is still relatively new, and it is important that you evaluate the field history of a system before purchase. Ask for a list of users and talk to the people who are already using the systems. This is a sure way to evaluate the strengths and weaknesses of a particular system before you make a capital investment.

Accuracy

Decisions on machine-train or plant system condition will be made based on the data acquired and reported by the predictive maintenance system. It must be accurate and repeatable. Errors can be input by the microprocessor and software as well as by the operators. The accuracy of commercially available predictive maintenance systems varies. Although most will provide at least minimum acceptable accuracy, some are well below the acceptable level.

It is extremely difficult for the typical plant user to determine the level of accuracy of the various instruments that are available for predictive maintenance. Vendor literature and salespeople will attempt to assure the potential user that their system is the best, most accurate, and so on. The best way to separate fact from fiction is to compare the various systems in your
plant. Most vendors will provide a system on consignment for up to 30 days. This will provide sufficient time for your staff to evaluate each of the potential systems before purchase.

**Training and Technical Support**

Training and technical support are critical to the success of your predictive maintenance program. Regardless of the techniques or systems selected, your staff will have to be trained. This training will take two forms: system users’ training and application knowledge for the specific techniques included in your program. Few, if any, of the existing staff will have the knowledge base required to implement the various predictive maintenance techniques discussed in the preceding chapters. None will understand the operation of the systems that are purchased to support your program.

**DATABASE DEVELOPMENT**

Each of the predictive maintenance technologies requires a logical method of acquiring, storing, valuating, and trending massive amounts of data over an extended period. Therefore, a comprehensive database that is based on the actual requirements of critical plant systems must be developed for the predictive maintenance program. At a minimum, these databases should include the following capabilities:

- Establishing data acquisition frequency
- Setting up analysis parameters
- Setting boundaries for signature analysis
- Defining alert and alarm limits
- Selecting transducers

**1.2.4 PROACTIVE-CENTERED MAINTENANCE (PCM):**

Proactive Maintenance differs from the other three maintenance modes because it addresses much more systemic elements of a maintenance program, rather than examining the machine itself. This approach is much more diligent and looks to control the problems that can lead to machine wear and tear as opposed to the deterioration itself. Machinery Lubrication reported that a Proactive Maintenance strategy will look to monitor and correct the root causes of failures, like contamination. Though this requires some due diligence, it is a less expensive maintenance strategy than the former three. There is a variety of techniques that are used to enact a proactive maintenance solution. For instance, making sure to train workers in the best practices for machine operation or using a chain of reliable suppliers for machine elements like lubrication and parts are great ways to help improve maintenance systems.
1.3 FAILURE TYPES, INVESTIGATION AND OCCURRENCES

Failure is defined in the BS Maintenance Glossary as 'The termination of the ability of an item to perform its required function.' This can involve such failure categories as follows:

1. Catastrophic failures which result in an immediate inability of a system to achieve its function;
2. Performance failures associated with a reducing performance of the equipment.
3. When the operator deliberately takes the equipment out of service, even though it is producing, at that time, its specified output.

1.3.1 FAILURE CLASSIFICATIONS

The terms 'failure' or 'fault' may be viewed from different angles according to the effect which the lack of performance has on the overall functional capability. Such aspects as economic viability, safety, engineering complexity, speed, causal influences all provide classifications leading to a description of failure.

Engineering Failure Classifications

There are two distinct classes of failure:

(1) Intermittent failure: failures which result in a lack of some function of the component only for a very short period of time, the component reverting to its full operational standard immediately after failure;

(2) Permanent failure: failures which result in a lack of some function which will continue until some part of the component is replaced.

Degree of Failure Classification

Permanent failures may be further subdivided into the following two types:
(1) Complete failure: failure which causes the complete lack of a required function.
(2) Partial failure: failure which leads to a lack of some function but not such as to cause a complete lack of the required function.

**Speed of failure classification**
Both complete and partial permanent failure may be further classified according to the suddenness with which the failure occurs:
(1) Sudden failure: failure which could not be forecast by prior testing or examination;
(2) Gradual failure: failure which could have been forecast by testing or examination.

**Degree and Speed Of Failure Classification**
Both failure forms can be combined to give the following further classifications:
(1) Catastrophic failures: failures which are both sudden and complete;
(2) Degradation failures: failures which are both partial and gradual.

**Cause-Of Failure Classification**
According to the manner by which failure develops, so it be further classified:
(1) Wear-out failures: failure attributable to the normal processes of wear as expected when the device was designed;
(2) Misuse failure: failure attributable to the application of stresses beyond the item's stated capabilities;
(3) Inherent weakness failure: failure attributable to a lack of suitability in the design or construction of the system or component itself when subjected to stresses within its stated capabilities.

**Hazard Classification**
Possible faults (major or minor failures) may be divided into two broad hazard groups, namely dangerous-failures or safe-failures.
(1) Dangerous faults: (a) protection system - failure to protect when needed, (b) machine tool - failure causing damage to work and/or operator, (c) traction system - failure to brake;
(2) Safe faults: (a) protection system - failure to operate when not needed, (b) machine tool - failure to start, (c) traction system - failure of brakes to apply when not needed.

**1.3.2 CAUSES OF FAILURE**
Defects may be broadly classified into those which result in a fracture and those for which fractures do not occur. These may in turn be sub-classified according to the causes due to
chemical, thermal and mechanical influences. Further sub-classification results in the identification of creep, corrosion, fatigue and mechanical fracture with a considerable further sub-classification of such causes of failure.

**Service failures**

The failures experienced most frequently in machinery are fracture, excessive deformation and surface failure, particularly corrosion deterioration. Typical review of failures and their causes are given in the following:

- Corrosion
- Excessive deformation
- Contamination
- Fatigue
- Overheating
- Overstressing
- Seizure
- Wear

**Failures That Can Be Prevented**

Simplified Failure Modes and Effects Analysis (SFMEA) provide a method for determining which failures can be prevented. Necessary inputs are the frequency of occurrence for each problem and cause combination and what happens if a failure occurs. Criticality of the failure is considered for establishing priority of effort. SFMEA is a top-down approach that looks at major components in the equipment and asks, ‘‘Will it fail?’’ And if so, how and why? Preventive maintenance investigators are, of course, interested in how a component will fail so that the mechanism for failure can be reduced or eliminated. For example, heat is the most common cause of failure for electrical and mechanical components. Friction causes heat in assemblies moving relative to each other, often accompanied by material wear, and leads to many failures. Any moving component is likely to fail at a relatively high rate and is a fine candidate for preventive maintenance.

The following are familiar causes of failure:

- Abrasion
- Abuse
- Age deterioration
- Bond separation
- Friction
- Operator negligence
- Puncture
- Shock

---

**Prepared by: Dr. Essam Bahgat**

essbahgat@gmail.com
Consumable depletion
Contamination
Corrosion
Dirt
Fatigue

Stress
Temperature extremes
Vibration
Wear.

**Maintenance to Prevent Failures**

Cleanliness is the watchword of preventive maintenance. Metal filings, fluids in the wrong places, ozone and other gases that deteriorate rubber components—all are capable of damaging equipment and causing it to fail. A machine shop, for example, that contains many electro-mechanical lathes, mills, grinders, and boring machines should have established procedures for ensuring that the equipment is frequently cleaned and properly lubricated. In most plants, the best tactic is to assign responsibility for cleaning and lubrication to the machine’s operator.
1.3.3 BATH TUB CURVE

The typical machine failure rate versus time plot is given in Fig. (1.5). The plot has three distinct zones: the infant mortality zone, the useful period, and the wear-out zone. The infant mortality zone with high failure rates occurs in the early stages of the machine. There could be several reasons for such high failure rates; some of them are faulty installation at the site, ignorance and unfamiliarity of the machine operator, improper electrical power supply, no availability of a user or training manual, improper specifications, and choice of the machine. Once the above reasons are sorted out, the machine’s failure rate reduces significantly; this state of the machine continues for a considerable time, which is known as the useful life of the machine. Finally, toward the end of the useful period, the failure rate of the machine again increases, which can be due to excessive wear and tear on the machine and fatigue failure of the machine component. Though by maintenance the failure rates can be controlled and reduced, a time comes when the cost of maintenance or upkeep is so high that it is better to completely replace the machine with a new one. The shape of the curve in Fig. (1.3), is in the form of a bath tub, hence the name.

The availability of the machine is defined as the ratio of the useful period (also known as uptime) to the total lifespan of the machine. Maintenance engineers strive to increase the availability of a machine by decreasing the machine’s downtime. The total lifespan of a machine is the summation of the uptime and downtime of the machine.

![Bath Tub Curve](image)

*Fig.1: 3: Bath tub curve.*
1.3.4 FAILURE ASPECTS

FAILURE PROCESS

- Wear
- Corrosion
- Fatigue
- Distortion

HUMAN PERFORMANCE

- Knowledge
- Skill
- Experience
- Motivation

ENVIRONMENTAL HOSTILITY VIBRATION

- Vibration
- Shock
- Humidity
- Temperature
- Dust / salt fog

ASSEMBLY

- Packaging
- Preserving

MATERIAL

- Metallurgy
- Heat Treatment
- Surface Coatings
- Surface Finishes
CHAPTER 2
MAINTENANCE RELIABILITY AND MANAGEMENT
2.1 INTRODUCTION

In real-world applications, a best practice is a very useful concept. Despite the need to improve on processes as times change and things evolve, a best practice is considered by some simply to be a business buzzword used to describe the process of developing and following a standard way of doing activities that any organization can use or implement to get better results. Implementing best practices in the area of maintenance and reliability can help an organization to:

• Increase output with the same set of assets
• Reduce the need for capital replacements
• Reduce maintenance cost per unit
• Reduce total cost per unit of output
• Improve performance cost, productivity, and safety
• Increase competitiveness
• Increase market share

Asset

An electronic or mechanical hardware component or device, a software product, or a manufacturing system or process.

Best Practices

Technique, methods, or processes that are more effective at delivering a desired outcome than any other techniques, methods, or processes. These are usually documented practices used by the most respected, competitive, and profitable organizations.

Maintenance

The act of maintaining, or the work of keeping an asset in proper operating condition.

Reliability

The probability that an asset or item will perform its intended functions for a specific period of time under stated conditions. It is usually expressed as a percentage and calculated using Mean Time Between Failures (MTBF).

2.2 IMPORTANCE OF RELIABILITY

Asset reliability is an important attribute for several reasons, including:
• **Customer Satisfaction.** Reliable assets will perform to meet the customer’s needs on time and every time. An unreliable asset will negatively affect the customer’s satisfaction severely. Thus, high reliability is a mandatory requirement for customer satisfaction.

• **Reputation.** An organization’s reputation is very closely related to the reliability of their services. The more reliable that plant assets are, the more likely the organization is to have a favorable reputation.

• **Repeat Business.** Reliable assets and plant will insure that customer’s needs are being met in a timely manner. Customer satisfaction will bring repeat business and also have a positive impact on future business.

• **Competitive Advantage.** Many leading and visionary companies have begun achieving high reliability / availability of their plants and assets. As a result of their greater emphasis on plant reliability improvement programs, they gain an advantage over their competition. As shown in Figure (2.1), we need to find the right level of reliability required to give us the optimum total cost. This graph illustrates the production or use cost, which is operations and downtime cost verses the reliability (and maintenance) cost.

![Fig.2: Reliability/ Availability Economics](image)

**Availability (A)**

The probability that an asset is capable of performing its intended function satisfactorily, when needed, in a stated environment. Availability is a function of reliability and maintainability.

**Failure**

---

*Prepared by: Dr. Essam Bahgat*

essbahgat@gmail.com
Failure is the inability of an asset / component to meet its expected performance. It does not require the asset to be inoperable. The failure could also mean reduced speed, or not meeting operational or quality requirements.

**Failure Rate**

The number of failures of an asset over a period of time. Failure rate is considered constant over the useful life of an asset. It is normally expressed as the number of failures per unit time. Denoted by Lambda (\( \lambda \)), failure rate is the inverse of Mean Time Between Failure (MTBF).

**Maintainability (M)**

The ease and speed with which a maintenance activity can be carried out on an asset. Maintainability is a function of equipment design and usually is measured by MTTR.

**Mean Time Between Failures (MTBF)**

MTBF is a basic measure of asset reliability. It is calculated by dividing total operating time of the asset by the number of failures over some period of time. MTBF is the inverse of failure rate (\( \lambda \)).

**Mean Time to Repair (MTTR)**

MTTR is the average time needed to restore an asset to its full operational condition upon a failure. It is calculated by dividing total repair time of the asset by the number of failures over some period of time. It is a basic measure of maintainability.

**Reliability (R)**

The probability that an asset or item will perform its intended functions for a specific period of time under stated conditions. It is usually expressed as a percentage and measured by the mean time between failures (MTBF).

**Reliability Centered Maintenance (RCM)**

A systematic and structured process to develop an efficient and effective maintenance plan for an asset to minimize the probability of failures. The process insures safety and mission compliance.

**Calculating Reliability and Availability**

Mean Time Between Failures

\[
\text{MTBF} = \frac{\text{Operating time}}{\text{Number of failures}}
\]

Failure rate = \( \frac{1}{\text{MTBF}} \)

\[
R(t) = e^{-\lambda t}
\]
Estimating reliability, \( R(t) \), is

\[
R(t) = e^{-\lambda t}
\]

Where, \( \lambda \) (lambda) = Failure rate = \( 1/MTBF \)

\( t \) = mission time, in cycles, hours, miles, etc.

(Note: \( e \) is base of the natural logarithm = 2.71828)

Calculating MTTR and Repair Rate

\[
MTTR = \frac{Total \, Repair \, Time}{\# \, of \, Failures}
\]

\[
Repair \, Rate \, (\mu) = \frac{1}{MTTR}
\]

\[
Availability = \frac{MTBF}{MTBF + MTTR}
\]

**Example 2-1**

A hydraulic system, which supports a machining center, has operated 3600 hours in the last two years. The plant’s CMMS system indicated that there were 12 failures during this period. What is the reliability of this hydraulic system if it is required to operate for 20 hours or for 100 hours?

**Solution**

MTBF = operating time / # of failures = 3600 / 12 = 300 hours

Failure rate = \( 1/MTBF = 1/300 = 0.003334 \) failures / hour

Reliability for 20 hours of operations,

\[
R(20) = e^{- (0.003334)(20)} = 93.55\%
\]

Reliability for 100 hours of operations,

\[
R(100) = e^{- (0.003334)(100)} = 71.65\%
\]

**Effectiveness** is defined by an equation as a figure-of-merit judging the opportunity for producing the intended results.

\[
Effectiveness = availability \ast reliability \ast maintainability \ast capability
\]

The effectiveness equation is the product of:

- The chance the equipment or system will be available to perform its duty,
- It will operate for a given time without failure.
It is repaired without excessive lost maintenance time and

- It can perform its intended production activity according to the standard.

Each element of the effectiveness equation requires a firm datum which changes with name plate ratings for a true value that lies between 0 and 1.

Consider these elements of the effectiveness equation for refineries and chemical plants. In many continuous process industries, availability is high (~85 to 98%), reliability is low (~0.001 to 10%) when measured against turnaround intervals, and maintainability is high (~50 to 90%) when measured against the allowed time for repairs, and productivity is high (~60 to 90%). So what does the effectiveness equation tell about these conditions? The one element destroying effectiveness is the reliability component (Barringer 1996b)—so it tells where to look for making improvements. Can the effectiveness equation be used to benchmark one business to another? In theory yes, but in practice no. The practical problem lies in normalizing effectiveness data across companies and across business lines. For example, one plant may have an acceptable mission time for their equipment of one year, whereas a second plant may require a five year mission time because of their turnarounds. Similarly, one plant may set a repair time for a specific pump as 8 hours elapsed time for a two man crew and the second plant may allow 12 hours elapsed time for a two man crew. At best, the effectiveness equation is applicable within a company where similar rules are applied across operating plants and the cost structure is similar.

**Availability** deals with the duration of up-time for operations and is a measure of how often the system is alive and well. It is often expressed as (up-time)/(up-time + downtime) with many different variants. Up-time and downtime refer to dichotomized conditions. Up-time refers to a capability to perform the task and downtime refers to not being able to perform the task, i.e., uptime ≠ not downtime. Also availability may be the product of many different terms such as:

\[ A = A_{\text{hardware}} \times A_{\text{software}} \times A_{\text{humans}} \times A_{\text{interfaces}} \times A_{\text{process}} \]

and similar configurations. Availability issues deal with at least three main factors

1) Increasing time to failure,
2) Decreasing downtime due to repairs or scheduled maintenance, and
3) Accomplishing items 1 and 2 in a cost effective manner.

As availability grows, the capacity for making money increases because the equipment is in service a larger percent of time.

Three frequently used availability terms (Ireson 1996) are explained below.
Inherent availability, as seen by maintenance personnel, (excludes preventive maintenance outages, supply delays, and administrative delays) is defined as:

\[ A_i = \frac{MTBF}{MTBF + MTTR} \]

Achieved availability, as seen by the maintenance department, (includes both corrective and preventive maintenance but does not include supply delays and administrative delays) is defined as:

\[ A_a = \frac{MTBM}{MTBM + MAMT} \]

Where \( MTBM \) is mean time between corrective and preventive maintenance actions and \( MAMT \) is the mean active maintenance time.

Operational availability, as seen by the user, is defined as:

\[ A_o = \frac{MTBM}{MTBM + MDT} \]

Where \( MDT \) is mean down time.

Reliability deals with reducing the frequency of failures over a time interval and is a measure of the probability for failure-free operation during a given interval, i.e., it is a measure of success for a failure free operation. It is often expressed as

\[ R(t) = \exp(-t/MTBF) = \exp(-\lambda t) \]

where \( \lambda \) is constant failure rate and MTBF is mean time between failure. MTBF measures the time between system failures and is easier to understand than a probability number. For exponentially distributed failure modes, MTBF is a basic figure-of-merit for reliability (failure rate, \( \lambda \), is the reciprocal of MTBF). For a given mission time, to achieve high reliability, a long MTBF is required. Also reliability may be the product of many different reliability terms such as

\[ R = R_{utilities} \times R_{feed-plant} \times R_{processing} \times R_{packaging} \times R_{shipping} \]

To the user of a product, reliability is measured by a long, failure free, operation. Long periods of failure free interruptions results in increased productive capability while requiring fewer spare parts and less manpower for maintenance activities which results in lower costs. To the supplier of a product, reliability is measured by completing a failure free warranty period under specified operating conditions with few failures during the design life of the product.

Improving reliability occurs at an increased capital cost but brings with it the expectation for ability, decreasing downtime and smaller maintenance costs, improved secondary failure costs, and results in better chances for making money because the equipment is free from failures for longer periods of time. While general calculations of reliability pertain to constant failure rates, detailed calculations of reliability are based on consideration of the failure mode which may be infant mortality (decreasing failure rates with time), chance failure (constant failure rates with time), or wear-out (increasing failure rates with time).
Maintainability

Deals with duration of maintenance outages or how long it takes to achieve (ease and speed) the maintenance actions compared to a datum. The datum includes maintenance (all actions necessary for retaining an item in, or restoring an item to, a specified, good condition) is performed by personnel having specified skill levels, using prescribed procedures and resources, at each prescribed level of maintenance. Maintainability characteristics are usually determined by equipment design which set maintenance procedures and determine the length of repair times.

The key figure of merit for maintainability is often the mean time to repair (MTTR) and a limit for the maximum repair time. Qualitatively it refers to the ease with which hardware or software is restored to a functioning state. Quantitatively it has probabilities and is measured based on the total down time for maintenance including all time for: diagnosis, trouble shooting, tear-down, removal/replacement, active repair time, verification testing that the repair is adequate, delays for logistic movements, and administrative maintenance delays. It is often expressed as

\[ M(t) = 1 - \exp(-t/\text{MTTR}) = 1 - \exp(-\mu t) \]

Where \(\mu\) is constant maintenance rate and MTTR is mean time to repair. MTTR is an arithmetic average of how fast the system is repaired and is easier to visualize than the probability value. Note the simple, easy to use criteria shown above, is frequently expressed in exponential repair times. A better and more accurate formula requires use of a different equation for the very cumbersome log-normal distributions of repair times describing maintenance times which are skewed to the right.

The maintainability issue is to achieve short repair times for keeping availability high so that downtime of productive equipment is minimized for cost control when availability is critical.

Capability

Deals with productive output compared to inherent productive output which is a measure of how well the production activity is performed compared to the datum. This index measures the systems capability to perform the intended function on a system basis. Often the term is the synonymous with productivity which is the product of efficiency multiplied by utilization. Efficiency measures the productive work output versus the work input. Utilization is the ratio of time spent on productive efforts to the total time consumed.

For example, suppose efficiency is 80% because of wasted labor/scrap generated, and utilization is 82.19% because the operation is operated 300 days per year out of 365 days. The capability is 0.8*0.8219 = 65.75%. These numbers are frequently generated by accounting departments for production departments as a key index of how they are doing. Thus these calculations need few explanations.
System effectiveness:

System effectiveness equations (Effectiveness/LCC) are helpful for understanding benchmarks, past, present, and future status as shown in Figure 3 for understanding trade-off information.

\[
\text{Effectiveness} = \frac{\text{Effectiveness}}{\text{LCC}}
\]

Life cycle costs, (LCC)

Life-Cycle Costing is a methodology where costs of a given asset are considered throughout its life-cycle.

2.3 Maintenance Performance Indicator (MPI)

Maintenance performance indicators (MPIs) are used for evaluating the effectiveness of maintenance carried out (Wireman, 1998). An indicator is a product of several metrics (measures). A performance indicator is a measure capable of generating a quantified value to
indicate the level of performance, taking into account single or multiple aspects. The selection of MPIs depends on the way in which the MPM is developed. MPIs could be used for financial reports, for monitoring the performance of employees, customer satisfaction, the health, safety and environmental (HSE) rating, and overall equipment effectiveness (OEE), as well as many other applications. When developing MPIs, it is important to relate them to both the process inputs and the process outputs.

2.4 Key Performance Indicators

Key Performance Indicators (KPI) is defined as “a performance indicator with a strategic significance, which is perceived as critical under given business circumstances and preferably selected from a pool of performance indicators”. Indicators can also be defined as a collection of one or more measures focusing on a predefined situation, and if several indicators are used at the same time and combined to each other with the purpose to get objective information the resulting measure is now also called KPI. These measures can be used to assess critical parameters or processes. From maintenance point of view the developed KPI must yield objective information regarding maintenance, e.g. preventive maintenance, spare parts logistics, planning and follow up. In order to compare and evaluate the different KPI, they must reflect the outcomes in monetary terms or other comparable measures such as reliability of plant assets.

2.5 MPI Standards and MPIs as in Use in Different Industries

The greatest challenge for measuring maintenance performance is the implementation of the MPM system for validation of the MPIs under a real and industrial set up. Implementation first involves executing the plan and deploying the system developed in place of the previously existing or planned system. Second, it means operating with the selected measures and validating the assurance that the defined maintenance measurement system works on a day-to-day basis. Without any formal measures of performance, it is difficult to plan control and improve the maintenance process. This is motivating senior business managers and asset owners to enhance the effectiveness of maintenance system. Also, with this, the focus is shifting to measure the performance of maintenance. Maintenance performance needs to be measured to evaluate, control and improve the maintenance activities for ensuring achievement of organizational goals and objectives. Different MPM frameworks and indicators to monitor, control and evaluate various performances are in use by different industries. More and more industries are working towards developing a specific MPM framework for their organization and identifying the indicators best suited to their industry. Organizations like International
Atomic Energy Agency (IAEA) has already developed and published safety indicators during 2000 for nuclear power plants, and Society for Maintenance and Reliability Professionals (SMRP) and European Federation of National Maintenance Societies (EFNMS) have started organizing working groups and workshops to identify and select MPIs for the industries. They have already defined and standardised some of the MPIs to be followed by their associates and members. Besides, a number of industries have initiated research projects in collaboration with universities to identify suitable MPIs as applicable to their specific industry. MPIs are measures of efficiency, effectiveness, quality, timeliness, safety, and productivity amongst others. Some of the industries where MPM framework has been tried out are in the nuclear, oil and gas (O & G), railway, process industry and energy sectors amongst others. A different approach is used for developing the MPM framework and indicators for different industries, as per the stakeholders’ requirements. Each organization under a specified industry is unique and as such the MPIs and the MPM framework is required to be modified or developed specifically to meet its unique organizational and operational needs. Some of the MPM approaches, frameworks and MPIs, as in use or under development by different societies, organizations and industries are discussed as under.

2.5.1 Nuclear Industry

The importance of the nuclear industry for energy generation as an alternate source is growing worldwide. International agencies like the International Atomic Energy Agency (IAEA) has been actively involved and sponsoring the development work in the area of indicators to monitor nuclear power plant (NPP) operational safety performance, from early 1990. The safe operation of nuclear power plants is the accepted goal for the management of the nuclear industry. A high level of safety results from the integration of the good design, operational safety and human performance. In order to be effective, a holistic and integrative approach is required to be adopted for providing a performance measurement framework and identifying the with desired safety attributes for the operation of the nuclear plant. Specific indicator trends over a period of time can provide an early warning to the management for investigating the causes of the observed change and comparing with the set target figure. Each plant needs to determine the indicators best suited to their individual needs, depending on the designed performance and, cost and benefit of operation/maintenance. The NPP performance parameters includes both the safety and economic performance indicators, with overriding safety aspects. To assess the operational safety of NPP, a set of tools like the plant safety aspect (PSA), regulating inspection, quality assurance and self assessment are used. Two categories of

Prepared by: Dr. Essam Bahgat

essbahgat@gmail.com
indicators of commonly applied are; risk based indicators and safety culture indicators. Operational Safety Performance Indicators Indicator development starts attributes usage and the operational safety performance indicators are identified. Under each attribute, overall indicators are established for providing overall evaluation of relevant aspects of safety performance and, under each overall indicator, strategic indicators are identified. The strategic indicators are meant for bridging the gap between the overall and specific indicators. Finally, a set of specific indicators are identified/developed for each strategic indicators to cover all the relevant safety aspects of NPP. Specific indicators are used to measure the performance and identify the declining performance, so that management can take corrective decisions. Some of the indicators used in plants are given in Table 2.1 (IAEA 2000).

<table>
<thead>
<tr>
<th>Table 2.1. Operational safety performance indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall indicators</td>
</tr>
<tr>
<td>1. Operating performance</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>2. State of structures, systems and components</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

2.5.2 Maintenance Indicators by EFNMS
Since, 2004, European Federation of National Maintenance Societies (EFNMS) has conducted a number of workshops by forming a working group from amongst the member National Maintenance Societies of Europe resulting in identifying maintenance indicators for different industries for the national societies and branches. These workshops collected data for the

Prepared by: Dr. Essam Bahgat

essbahgat@gmail.com
maintenance indicators from industries and also trained the participants in the use of the
dicators. The Croatian maintenance society (HDO) hosted the first workshop on maintenance
indicators for the food and pharmaceutical business. The workshop was organised to train the
maintenance managers in the use of maintenance indicators or Key Performance Indicators
(KPIs) and to create an understanding of how to interpret the performance measured by the
indicators. The participating maintenance managers were from the food and pharmaceutical
industries. A number of workshops are organized in the same sector of industries to compare
the results of the industry with the average maintenance performance in the sector. One of the
important objectives of these workshops, besides the calculation of the indicators, is to increase
the competence of the maintenance manager, who gets an understanding of the mechanism
behind the indicators.

2.5.3 SMRP Metrics

The SMRP best practices committee has been charted to identify and standardize maintenance
and reliability metrics and terminology since 2004. They followed a six step process for the
development of the metrics. The SMRP best practice metrics are published by the SMRP under
the “Body of knowledge”, available for viewing at www.smrp.org . The numbering system for
the metrics is explained on the web-page. Each metric has two files to describe the metric and
feedback from the review of the metric. There are 45 metrics under development by different
authors as of Feb 2006. A template is developed to provide a consistent method of describing
each metric. The basic elements of each metric are:
A number of metrics are published at the SMRP web-site, which can be easily accessed. These
metrics are explained in a clear and concise manner, which can be used by the personnel at
different hierarchical level without much difficulty. An example of the SMRP best practice
metrics is given below:

2.5.3.1 Oil and Gas Industry

The cost of maintenance and its influence on the total system effectiveness of oil and gas
industry is too high to ignore (Kumar and Ellingsen, 2000). The oil and gas industry uses MPIs
and MPM framework extensively due to its ever growing and competitive nature of business,
besides the productivity, safety and environmental issues. The safe operations of oil and gas
production units are the accepted goal for the management of the industry. A high level of safety is essential from the integration of good design, operational safety and human performance. To be effective, an integrative approach is required to be adopted for providing an MPM framework and identifying the MPIs with desired safety attributes for the operation of the oil and gas production unit. Specific indicator trends over a period of time can provide an early warning to management to investigate the causes of the observed change and comparing with the set target figure. Each production unit needs to determine the indicators best suited to their individual needs, depending on the designed performance and cost and benefit of operation/maintenance. Some of the MPIs reported from plant level to result unit level to the result area for the Norwegian oil and gas industry are grouped into different categories as follows:

**Production**
- Production Produced volume oil (Sm3).
- Planned oil-production (Sm3).
- Produced volume gas (Sm3).
- Planned gas-production (Sm3).
- Produced volume condensate (Sm3).
- Planned condensate-production (Sm3).

**Technical integrity**
- Backlog preventive maintenance (man-hours).
- Backlog corrective maintenance (man-hours).
- Bumber of corrective work orders.

**Maintenance parameters**
- Maintenance man-hours safety system.
- Maintenance man-hours system.
- Maintenance man-hours other systems.
- Maintenance man-hours total.

**Deferred production**
- Due to maintenance (Sm3).
- Due to operation (Sm3).
- Due to drilling/well operations (Sm3).
- Weather and other causes (Sm3).
2.5.3.2 Railway Industry
Railway operation and maintenance is meant for providing a satisfying service to the users, while meeting the regulating authorities’ requirements. Today, one of the requirements for the infrastructure managers is to achieve cost effective maintenance activities, a punctual and cost-effective railway transport system. As a result of a research project for the Swedish railway transport system, the identified maintenance performance indicators are:

- Capacity utilization of infrastructure;
- Capacity restriction of infrastructure;
- Hours of train delays due to infrastructure;
- Number of delayed freight trains due to infrastructure;
- Number of disruptions due to infrastructure;
- Degree of track standard;
- Markdown in current standard;
- Maintenance cost per track-kilometer;
- Traffic volume;
- Number of accidents involving railway vehicles;
- Number of accidents at level crossings;
- Energy consumption per area;
- Use of environmental hazardous material;
- Use of non-renewable materials;
- Total number of functional disruptions; and
- Total number of urgent inspection remarks.

2.5.3.3 Process Industry
Measuring maintenance performance has drawn considerable interest in the utility, manufacturing and process industry in the last decade. Organizations are keen to know the return on investment made in maintenance spending, while meeting the business objectives and strategy. Under challenges of increasingly technological changes, implementing an appropriate performance measurement system in an organization ensures that actions are aligned to strategies and objectives of the organization. Balanced, holistic and integrated multi-criteria hierarchical maintenance performance measurement (MPM) models developed with seven criteria and specific modification for the industry were tried out for implementation and achieving the total maintenance effectiveness for a pillarization plant and an energy producing service industry of Sweden. The MPIs for the process industry are:
1. Downtime (hours);  
2. Change over time;  
3. Planned maintenance tasks;  
4. Unplanned tasks;  
5. Number of new ideas generated;  
6. Skill and improvement training;  
7. Quality returned;  
8. Employee complaints; and  
9. Maintenance cost per ton.

In addition, MPIs identified for the multi-criterion hierarchical MPM framework, which are in existence and in use at LKAB (iron ore process company), are OEE, production cost per ton, planned maintenance tasks, quality complaints number, number of accidents, HSE complaints, and impact of quality.

### 2.6 Maintenance Engineering Versus The Maintainability Engineering.

Maintenance and maintainability are closely interrelated, and many people find it difficult to make a clear distinction between them. Maintenance refers to the measures taken by the users of a product to keep it in operable condition or repair it to operable condition. Maintainability refers to the measures taken during the design and development of a product to include features that will increase ease of maintenance and will ensure that when used in the field the product will have minimum downtime and life-cycle support costs.

*Maintenance* is the act of repairing or servicing equipment.  

*Maintainability* is a design parameter intended to minimize repair time.

#### 2.6.1 Maintainability Engineer

This person plays an instrumental role in product design, and has often served before as a design engineer. Some of a maintainability engineer's tasks are maintainability prediction, maintainability allocation, developing maintainability demonstration documents, performing analysis of maintainability feedback data, reviewing product design with respect to maintainability features, participating in the development of maintainability design criteria, and preparing maintainability design documents.

#### 2.6.2 Reliability Engineer
As is the case with the maintainability engineer, the reliability engineer is often a former design engineer. This person assists management in defining, evaluating, and containing risks. Responsibilities include reliability allocation, reliability prediction, analyzing customer requirements, developing reliability growth monitoring procedures and reliability evaluation models, monitoring subcontractor reliability programs, participating in design reviews, developing reliability test and demonstration procedures, and assessing the effect of environment on product reliability

2.6.3 Quality Control Engineer
While this individual does not necessarily belong to one of the engineering disciplines, he or she plays an important role during the product life cycle, especially during the design and manufacture phases. The quality control engineer develops and applies appropriate inspection plans, evaluates the quality of procured parts, develops quality related standards, analyzes quality associated defects, uses statistical quality control methods, participates in quality related meetings, and audits the quality control system at appropriate times.

2.6.4 Human Factors Engineer
The human factors engineer determines the tasks required to make full use of the product; the most appropriate way of grouping the various tasks required into individual jobs; the best way of displaying specified information, and of arranging visual displays to ensure optimum use, especially when the user must split his or her attention between two or more displays; how to label control devices most clearly; the arrangement of control devices that generates optimum use; what level of information flow or required decision making overburdens operators; how human decision-making and adaptive abilities can be put to best use; the environmental conditions that affect individual performance and physical well-being; and other similar matters.

2.6.5 Safety engineer
The role played by this professional during the design phase is very critical, as the newly designed product must be safe to operate and maintain. Some of the safety engineer's tasks include analyzing historical data on product hazards, failures, and accidents; analyzing new designs from the standpoint of safety; keeping management continually informed about safety program performance; monitoring subcontractors' safety efforts; providing safety-related information to those concerned; establishing criteria for analyzing any accident associated with a product manufactured by the company; and developing safety warning devices.

2.7 Maintainability Tools
### 2.7.1 Fault Tree Logic and Event Symbols

There are many logic and event symbols used to construct fault trees. Figure 2.2 presents the two most commonly used logic symbols: the OR gate and AND gate. The OR gate symbol signifies that an output fault event occurs if one or more of the m input fault events occur. The AND gate symbol denotes that the output fault event only occurs if all of the m input fault events occur.

![Figure 2.2: Two commonly used fault tree logic symbols: (a) OR gate, (b) AND gate.](image)

#### Example 2-2

Assume that a workshop repairs failed electric motors. Develop a fault tree for the following undesired event an electric motor will not be repaired by a given point in time—using the fault tree symbols described earlier. Consider only these four factors:

- Motor is too damaged to repair
- Skilled manpower is unavailable
- Spare parts are unavailable
- Repair tools or facilities are unavailable

![Fault Tree for Electric Motor Repair](image)

### 2.8 WORLD-CLASS MAINTENANCE

**What is world-class maintenance?**

To keep production in high gear—and to survive—manufacturers are increasingly obliged to move from a breakdown maintenance mindset toward a concept of proactive maintenance
organized around a well-trained staff, within a carefully defined plan, and with meaningful participation of employees outside of what is normally thought of as traditional maintenance. It’s a move toward a total team approach of effective preventive maintenance and total quality management (TQM).

At the core of world-class maintenance is a new partnership among the manufacturing or production people, maintenance, engineering, and technical services to improve what is called overall equipment effectiveness (OEE). It is a program of zero breakdowns and zero defects aimed at improving or eliminating the six crippling shop-floor losses:

1) Equipment breakdowns
2) Setup and adjustment slowdowns
3) Idling and short-term stoppages
4) Reduced capacity
5) Quality-related losses
6) Startup/restart losses

**Focus On Maintenance**

How does this path to maturity relate to the path to maturity for asset or equipment maintenance? Next figure compares the two. In stage 1 of the path to world-class performance, the equipment is not maintained or repaired unless the customer (i.e., operations, production, or facilities) complains that it is broken. Only then will the maintenance organization work (or in some cases be allowed to work) on the equipment. In other words, “if it isn't broke, don’t fix it.”
Overall Equipment Effectiveness

Overall equipment effectiveness (OEE) is the benchmark used for world-class maintenance programs. The OEE benchmark is established by measuring equipment performance. Measuring equipment effectiveness must go beyond just the availability or machine uptime. It must factor in all issues related to equipment performance. The formula for equipment effectiveness must look at the availability, the rate of performance, and the quality rate. This allows all departments to be involved in determining equipment effectiveness. The formula could be expressed as:

\[ \text{Availability} \times \text{Performance Rate} \times \text{Quality Rate} = \text{OEE} \]

The availability is the required availability minus the downtime, divided by the required availability. Expressed as a formula, this would be:

\[ \frac{\text{Required Availability} - \text{Downtime}}{\text{Required Availability}} \times 100 = \text{Availability} \]

The required availability is the time production is to operate the equipment, minus the miscellaneous planned downtime, such as breaks, scheduled lapses, meetings, and so on. The downtime is the actual time the equipment is down for repairs or changeover. This is also sometimes called breakdown downtime. The calculation gives the true availability of the equipment. This number should be used in the effectiveness formula. The goal for most companies is greater than 90 percent. The performance rate is the ideal or design cycle time to produce the product multiplied by the output and divided by the operating time. This will give a performance rate percentage. The formula is:

\[ \frac{\text{Design Cycle Time} \times \text{Output}}{\text{Operating Time}} \times 100 = \text{Performance Rate} \]

The design cycle time or production output will be in a unit of production, such as parts per hour. The output will be the total output for the given period. The operating time will be the availability value from the previous formula. The result will be a percentage of performance. This formula is useful for spotting capacity reduction breakdowns. The goal for world-class companies is greater than 95 percent. The quality rate is the production input into the process or equipment minus the volume or number of quality defects divided by the production input. The formula is:

\[ \frac{\text{Production Input} - \text{Quality Defects}}{\text{Production Input}} \times 100 = \text{Quality Rate} \]
The product input is the unit of product being fed into the process or production cycle. The quality defects are the amount of product that is below quality standards (not rejected; there is a difference) after the process or production cycle is finished. The formula is useful in spotting production-quality problems, even when the customer accepts the poor-quality product. The goal for world-class companies is higher than 99 percent. Combining the total for these goals, it is seen that:

$$90\% \times 95\% \times 99\% = 85\%$$

**Twelve-Point Data Collection Process Facilitates Analysis of OEE Information**

<table>
<thead>
<tr>
<th>Overall Equipment Effectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Planned Time Available</td>
</tr>
<tr>
<td>2. Planned Downtime</td>
</tr>
<tr>
<td>3. Net Available Run Time</td>
</tr>
<tr>
<td>4. Downtime Losses</td>
</tr>
<tr>
<td>5. Actual Operating Time</td>
</tr>
<tr>
<td>6. Equipment Availability</td>
</tr>
<tr>
<td>7. Total Output for Operating Time</td>
</tr>
<tr>
<td>8. Design Cycle Time</td>
</tr>
<tr>
<td>9. Operational Efficiency</td>
</tr>
<tr>
<td>10. Rejects During Turn (Shift)</td>
</tr>
<tr>
<td>11. Rate of Product Quality</td>
</tr>
<tr>
<td>12. OEE</td>
</tr>
</tbody>
</table>

To be able to compete for the national total productive maintenance (TPM) prize in Japan, the equipment effectiveness must be greater than 85 percent. Unfortunately, the equipment effectiveness in most U.S. companies barely breaks 50 percent. It is little wonder that there is so much room for improvement in typical equipment maintenance management programs.

A plastic injection molding plant had a press with the following statistics:

Prepared by: Dr. Essam Bahgat

essbahgat@gmail.com
• The press was scheduled to operate 15 eight-hour shifts per week.
• This gave a total possibility of 7,200 minutes of run time per week.
• Planned downtime for breaks, lunches, and meetings totaled 250 minutes.
• The press was down for 500 minutes for maintenance for the week.
• The changeover time was 4,140 minutes for the week.
• The total output was 15,906 pieces.
• The design cycle time was 9.2 pieces per minute.
• There were 558 rejected pieces for the week.
• What is the OEE for the press for the week in question?

**Overlaying World-Class Standard on the Baseline Data**

<table>
<thead>
<tr>
<th>Overall Equipment Effectiveness</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Planned Time Available</td>
<td>7,200 minutes</td>
</tr>
<tr>
<td>8 hours × 60 minutes = 480 minutes × 15 turns or shifts</td>
<td></td>
</tr>
<tr>
<td>2. Planned Downtime</td>
<td>250 minutes</td>
</tr>
<tr>
<td>For preventive maintenance, lunch, breaks, etc.</td>
<td></td>
</tr>
<tr>
<td>3. Net Available Run Time</td>
<td>6,950 minutes</td>
</tr>
<tr>
<td>Item 1 – Item 2</td>
<td></td>
</tr>
<tr>
<td>4. Downtime Losses</td>
<td>695 minutes</td>
</tr>
<tr>
<td>Breakdowns, setups, adjustments</td>
<td></td>
</tr>
<tr>
<td>5. Actual Operating Time</td>
<td>6,255 minutes</td>
</tr>
<tr>
<td>Item 3 – Item 4</td>
<td></td>
</tr>
<tr>
<td>6. Equipment Availability</td>
<td>90%</td>
</tr>
<tr>
<td>Item 5 ×100 Item 3</td>
<td></td>
</tr>
<tr>
<td>7. Total Output for Operating Time</td>
<td>54,516 units</td>
</tr>
<tr>
<td>Total produced in units, pieces, tons, etc.</td>
<td></td>
</tr>
<tr>
<td>8. Design Cycle Time</td>
<td>0.109 minutes/unit</td>
</tr>
<tr>
<td>9. Operational Efficiency</td>
<td>95%</td>
</tr>
<tr>
<td>Item 8 × Item 7 ×100 Item 5</td>
<td></td>
</tr>
<tr>
<td>10. Rejects During Turn (Shift)</td>
<td>545 units</td>
</tr>
<tr>
<td>11. Rate of Product Quality</td>
<td>99%</td>
</tr>
<tr>
<td>Item 7 – Item 10 × Item 7</td>
<td></td>
</tr>
<tr>
<td>12. OEE</td>
<td>85%</td>
</tr>
</tbody>
</table>

The gross time available for the press is entered in line 1. The planned downtime, which involves activities that management sets a priority on and cannot be eliminated, is entered in line 2 (the 250 minutes for the week). The net available time for operation is entered in line 3 (this is actually line 1 minus line 2). The downtime losses, which are all unplanned delays, are entered in line 4. This would include maintenance delays, changeovers (which can be
minimized), setups, adjustments, and so on. The actual time the press operated is entered on line 5 (this is the difference between lines 3 and 4). The equipment availability (line 6) is line 5 divided by line 3 times 100 percent.

The OEE is calculated in the next section. The total output for the operating time is entered in line 7. The actual design cycle time (this number must be accurate) is entered on line 8. The operational efficiency is calculated and entered on line 9. The operational efficiency is line 7 (the total output) times line 8 (design cycle time) divided by line 5 (the actual operating time) times 100 percent. This number should be evaluated carefully to ensure that the correct design capacity was used. If the percentage is high or exceeds 100 percent, then the wrong design capacity was probably used.

The quality rate is determined by the total output for the operating time (line 7) minus the number of rejects for the measured period (line 10) divided by the total output (line 7) times 100 percent. In the sample, the availability is 33 percent; the operational efficiency is 75 percent; and the quality rate is 96.8 percent. The OEE for the press for the week is 23.96 percent.

What do these conditions mean? What do the indicators show the typical manufacturer?

The answers are evident when a second model using the same press is examined. In Table 18–2, the parameters are set at world-class standards to give an OEE of 85 percent. As can be quickly observed, the major improvement is in the total output for the operating time (line 7). The press now will make 54,516 parts, compared to 15,348 with the 23.96 percent OEE. Because the resources to make the parts (labor and press time) are the same, it makes the company more products and ultimately more profits. With the press operating at an OEE of 85 percent, the same productivity results as if 3.5 presses were running at the 23.96 percent OEE. The potential for increased profitability and ultimate competitiveness is staggering.

Proactive maintenance can have a positive impact on any company’s productivity and profitability, as long as the entire organization is willing to change its culture and the way in which day-to-day business is conducted.

**OEE Calculation**

*Availability* considers **Down Time Loss**.

\[ A = \frac{\text{Operating Time}}{\text{Planned Production Time}} \]
Example: A conveyor system is scheduled to run for two 8-hour (960 minute) shifts, with a 30-minute break per 8 hours.

**Operating time** = 900 min scheduled - 90 min unscheduled downtime = 810 minutes

**Calculation:** **Availability** = Operating time / Planned Production Time

Availability 810 minutes / 900 = 90%

**Performance** considers **Speed Loss**

\[ P = \frac{\text{Ideal Cycle Time}}{(\text{Operating Time} / \text{Total Pieces})} \]

**Ideal Cycle Time** is the minimum cycle time that your process can be expected to achieve in optimal circumstances. It is sometimes called *Design Cycle Time, Theoretical Cycle Time* or *Nameplate Capacity.*

Example: A conveyor system is scheduled to run for two 8-hour (960 minute) shifts, with a 30-minute break per 8 hours.

Operating time = 900 min scheduled - 90 min unscheduled downtime = 810 minutes

Parts are produced at the Standard Rate of 40 units/hour or 1.5 minutes per unit

The conveyor system produces 525 units during the total day. This is Total Units, not Good Units. Quality is not measured at this point.

**Calculation:** **Performance** = Ideal Cycle Time / (Operating Time of Total Pieces)

Time to produce parts = 525 * 1.5 minutes/unit = 787 minutes (Ideal Cycle Time)

Performance (Productivity) = 787 / 810 = 97%

**Quality** considers **Quality Loss**, and is calculated as:

\[ Q = \frac{\text{Good Pieces}}{\text{Total Pieces}} \]

Example: 525 units are produced. 56 are defective.

(525 units produced - 39 defective units) = 469 units

469 good units / 525 total units produced = 89%

**OEE** considers all three **OEE Factors**, and is calculated as:

\[ \text{OEE} = \text{Availability} \times \text{Performance} \times \text{Quality} \]

**OEE** = 90% X 97% X 89%

OEE = 77.7%
CHAPTER 3
MACHINERY CONDITION MONITORING TECHNIQUES
Machine condition monitoring is the process of monitoring the condition of a machine with the intent to predict mechanical wear and failure. Vibration, noise, and temperature measurements are often used as key indicators of the state of the machine. Trends in the data provide health information about the machine and help detect machine faults early, which prevents unexpected failure and costly repair.

### 3.1 UNDERSTANDING MACHINE CONDITION MONITORING

Machine condition monitoring is important because it provides information about the health of a machine. You can use this information to detect warning signs early and help your organization stop unscheduled outages, optimize machine performance, and reduce repair time and maintenance costs. Figure (3.1) shows a typical machine failure example and the warning signs.

![Fig.3: The warning signs of machine failure](image)

**Fig.3: 1 The warning signs of machine failure:** Using a machine condition monitoring system, you can detect failure signs months before repair is required, allowing for proper maintenance scheduling and shutdown.

### 3.2 TYPES OF MACHINE CONDITION MONITORING

Each of the five main varieties of machine condition monitoring serves a different role. **Route-Based Monitoring.**

---

**Prepared by: Dr. Essam Bahgat**

essbahgat@gmail.com
Route-based monitoring involves a technician recording data intermittently with a handheld instrument. This data is then used for trending to determine if more advanced analysis is needed.

**Portable Machine Diagnostics.**

Portable machine diagnostics is the process of using portable equipment to monitor the health of machinery. Sensors are typically permanently attached to a machine and portable data acquisition equipment is used to read the data.

**Factory Assurance Test.**

Factory assurance test is used to verify that a finished product meets its design specifications and to determine possible failure modes of the device.

**Online Machine Monitoring.**

Online machine monitoring is the process of monitoring equipment as it runs. Data is acquired by an embedded device and transmitted to a main server for data analysis and maintenance scheduling.

**Online Machine Protection.**

Online machine protection is the process of actively monitoring equipment as it runs. Data is acquired and analyzed by an embedded device. Limit settings can then be used to control turning on and off machinery.

**3.3 Monitoring Of Machinery Faults**

**3.3.1 Corrosion monitoring**

Since corrosion proceeds as an electro-chemical process the techniques of monitoring are directed to the identification of the electrical currents set up specifically by the corrosive activity.

**CORROSOMETER**

This is the proprietary name for an instrument in which the electrical resistance monitoring technique uses a balanced bridge technique to measure the change in resistance of a probe as it thins away under corrosion.
Corrator

A device or probe employed to measure current flow in a process flow. It consists of two identical electrodes, to which a small current is applied and measured, from which corrosion rates can be calculated. Corrosion rate measurements are made using the linear polarization resistance technique. The instrument measures the current required to polarize the electrodes of a probe to a known potential. From the polarization potential and the measured current, polarization resistance can be calculated. Then, using Faraday’s law, the instantaneous corrosion rate can be calculated from polarization resistance.

3.3.2 Gas leakage monitoring

Two basic methods can be applied to test for leaks:

**Sniffing:** in which the leak tester responds to the gas. With sealed, unpressurized units a tracer gas filling is used. With small pressurized components, they are placed in a ‘bomb’ with a pressurized tracer gas and after a ‘soaking’ removed and sniffed for tracer leakage.

**Hood method:** Components are put in a hood and air is pumped round. Gas leakage in the purged gas indicates a leak in the test component.

Auer detector tubes/gas testers

These testers are used by drawing air through the detector tubes which contain chemical reagents sensitive to selective gases.

---

*Fig.3: 2 handheld instrument provides quick metal loss measurements of ER corrosion probes.*

*Fig.3: 3 Handheld Corrosion Data Logger*
Gas thermal conductivity testers

Changes in the thermal conductivity when gases leak into a small volume provide the basis for leak sensing which responds to such gases as hydrogen, helium, methane hydrocarbons, refrigerant gases, carbon dioxide, argon etc. In fact, this method is suitable for any gas with a different thermal conductivity from air.

3.3.3 Ultrasonic

This predictive maintenance technique uses principles similar to vibration analysis. Both monitor the noise generated by plant machinery or systems to determine their actual operating condition. Unlike vibration monitoring, however, ultrasonics monitors the higher frequencies (i.e., ultrasound) produced by unique dynamics in process systems or machines. The normal monitoring range for vibration analysis is from less than 1 Hz to 30,000 Hz. Ultrasonics techniques monitor the frequency range between 20,000 Hz and 100 kHz.

ULTRASONIC APPLICATIONS

As part of a predictive maintenance program, ultrasonic instruments are used for three primary applications: airborne noise analysis, leak detection, or material testing.

Airborne Noise Analysis all plants are required by Occupational Safety and Health Administration (OSHA) regulations to meet ambient noise levels throughout their facilities. These mandates have forced these plants to routinely monitor the noise levels within each area of the plant and to provide hearing protection in those areas where the ambient noise level is above acceptable levels.

Ultrasound Applications

- Steam Traps
- Heat Exchangers
- Pneumatic Systems
- Pressure / Vacuum Leaks
- Motors / Gearboxes
- Electric Arc / Corona / Tracking
- Bearing / Valves

---

Prepared by: Dr. Essam Bahgat

essbahgat@gmail.com
Leak Detection

The principal application for ultrasonic monitoring is in leak detection. The turbulent flow of liquids and gases through a restricted orifice (i.e., leak) will produce a high frequency signature that can easily be identified using ultrasonic techniques. Therefore, this technique is ideal for detecting leaks in valves, steam traps, piping, and other process systems.

Ultrasonic leak testing

Leaks in automobile pneumatic braking systems can be accurately located by means of an ultrasonic detector which gives both an audible and visual indication of the leak. It is a simple hand-held device no bigger than a large torch designed to provide on-site location of air or gas leaks in both pressure and vacuum pipe systems and vessels. Ultrasonic Leak Surveying.

Ultrasonic Leak Detection Benefits:

- The completion of a service whose purpose is to improve energy conservation
- A report which details photos and location of all items identified for repair
- A calculation of the energy cost for each leak
- € Value Cost Avoidance Calculation

Materials Testing

Ultrasonics has been, and continues to be, a primary test methodology for materials testing. Typical test frequencies start at 250 kilo Hertz (kHz), or 250,000 cycles per second (cps), up to 25 MegaHertz (MHz), or 25 million cps. Testing materials generally consist of introducing an energy source into the material to be tested and recording the response characteristics using...
ultrasonic instruments. These tests may be as simple as striking the material with a hammer and recording the results with an accelerometer and ultrasonic meter. Ultrasonic testing relies on the measurement of time and amplitude or strength of a signal between emission and reception. Because of a mismatch of acoustic properties between materials, the sound will partly reflect at interfaces. The quality of reflected energy depends on the acoustic impedance ratio between two materials. For example, found transmitted through steel reaching a steel/air boundary will cause 99.9 percent internal reflection, whereas a steel/water boundary would reflect only 88 percent within the material and transmit 12 percent into the water. If impedance ratios are widely different, such as an open crack with a steel/air interface, then adequate reflection will occur and permit detection of the flaw. Conversely, a small crack in a compressive stress field that does not have oxidized faces will yield a steel/steel boundary and cannot be detected using this method.

**Ultrasonic Thickness Gauge**

Ultrasonic thickness measurement (UTM) is a method of performing non-destructive measurement (gauging) of the local thickness of a solid element (typically made of metal, if using ultrasound testing for industrial purposes) basing on the time taken by the ultrasound wave to return to the surface. This type of measurement is typically performed with an ultrasonic thickness gauge, using this simple formula:

\[ l_m = \frac{ct}{2} \]

where

\( l_m \) is the thickness of the sample
\( c \) is the celerity of sound in the given sample
\( t \) is the traverse time

Fig.3: 7 hand held ultrasonic thickness gauge

**Sheet Metal Thickness Laser Measurement**

Automotive metal stamping requires a specific thickness to ensure the parts are equal, and to ensure the die is not damaged during this process. As metal
sheets move down the conveyor, *laser displacement sensors* measure the sheets on a very precise scale to ensure they are the correct thickness.

### 3.3.4 Temperature monitoring

Temperature measurement (i.e., sensors, thermography) helps detect potential failures related to a temperature change in equipment. Measured temperature changes can indicate problems such as excessive mechanical friction (i.e., faulty bearings, inadequate lubrication), degraded heat transfer (i.e., fouling in a heat exchanger), or poor electrical connections (i.e., loose, corroded, or oxidized connections). Two methods are used in the temperature monitoring:

- Point Temperature Devices
- Infrared Thermography

**Thermocouple**

A thermocouple consists of two wires of dissimilar metals joined near the measurement point (junction), a reference junction, and a measuring device. The output is a small voltage measured between the two wires which is converted to a temperature readout by an instrument.

**Infrared thermometer**

Is a thermometer which infers temperature from a portion of the thermal radiation sometimes called blackbody radiation emitted by the object being measured. They are sometimes called laser thermometers as a laser is used to help aim the thermometer.

**Infrared Thermography**

This non-contact technique uses infrared cameras to measure the temperature of heat-radiating surfaces within the line of sight of the camera. The camera measures temperature variations on the surface of the object being monitored and converts the temperature data into video or audio.
signals that can be displayed or recorded in a wide variety of formats for future analysis. Thermography is a predictive maintenance technique that can be used to monitor the condition of plant machinery, structures, and systems. It uses instrumentation designed to monitor the emission of infrared energy (i.e., temperature) to determine operating condition. By detecting thermal anomalies (i.e., areas that are hotter or colder than they should be), an experienced surveyor can locate and define incipient problems within the plant.

**Infrared Basics**

Infrared technology is predicated on the fact that all objects with a temperature above absolute zero emit energy or radiation. Infrared radiation is one form of this emitted energy. Infrared emissions, or below red, are the shortest wavelengths of all radiated energy and are invisible without special instrumentation. The intensity of infrared radiation from an object is a function of its surface temperature; however, temperature measurement using infrared methods is complicated because three sources of thermal energy can be detected from any object. Only the emitted energy is important in a predictive maintenance program. Reflected and transmitted energies will distort raw infrared data. Therefore, the reflected and transmitted energies must be filtered out of acquired data before a meaningful analysis can be completed. The surface of an object influences the amount of emitted or reflected energy. A perfect emitting surface, Figure 3-11, is called a “blackbody” and has an emissivity equal to 1.0. These surfaces do not reflect. Instead, they absorb all external energy and re-emit it as infrared energy.

![Infrared Basics Diagram](image)

**Fig.3: 11 Blackbody emissions. A perfect or blackbody absorbs all infrared energy. A = Absorbed energy. R =Reflected energy. T = Transmitted energy. E = Emitted energy.**

**Types of infrared problems**

There are three basic types of thermal problems:

- Mechanical looseness
- Load problems
• Component failure

3.5 Vibration Monitoring
Dynamic monitoring (i.e., spectrum analysis, ultrasonic analysis) involves measuring and analyzing energy emitted from mechanical equipment in the form of waves, such as vibration, pulses, and acoustic effects, over a period of time. Measured changes in the vibration...
characteristics from equipment over time can indicate problems such as wear, imbalance, misalignment, and damage. This topics will be discuss in details later in chapter 7.

Vibration Monitoring used to monitor the health of rotational equipment. This involves the trending and analysis of certain machine performance parameters to detect and identify developing problems such as:

- Imbalance
- Miss-alignment
- Bearing Faults
- Gear Tooth Mesh
- Under/Over Lubrication
- Before failure occurs

**How Vibration Monitoring works?**

- Vibration measurements are taken on specific machine locations using piezo-electric transducers (Accelerometers) and stored in a Data Collector
- Vibration Data is uploaded to Vibration Analysis Software and Analysed
- A vibration Report outlining Findings and Recommendations is prepared

**Benefits to you from Vibration Monitoring**

- Determine machine health status
- Predict failure mode
- No surprise downtime
- No unexpected failures
- No secondary damage
- Planned Maintenance program
- Vibration Analysis

Once vibration monitoring detects a problem with a piece of machinery, Taylor Condition Monitoring will pinpoint specific machinery problems by revealing their unique vibration characteristics. These vibration characteristics are then broken down into individual vibration frequencies, which can be matched to known defect frequencies that a range of different problems normally appear at. These frequencies will largely depend on the type of machine, speed and the type of bearings installed. With hundreds of specific operational and mechanical problems that can cause a machine to exhibit excessive vibration, knowing the frequency of vibration is paramount. The forces that cause vibration are usually generated through the
rotating motion of the machine's parts, and because these forces change in direction or amplitude according to the rotational speed of the machine components, it follows that most vibration problems will have frequencies that are directly related to rotational speeds. Some of these problems have their own unique characteristics in the vibration spectrum, while other problems have similar characteristics.

3.3.6 Oil Analysis monitoring

Oil analysis (i.e., Ferrography, particle counter testing) can be performed on different types of oils such as lubrication, hydraulic, or electrical insulation oils. It can indicate problems such as machine degradation (i.e., wear), oil contamination, improper oil consistency (i.e., incorrect or
improper amount of additives), and oil deterioration. Oil analysis is most often based on regularly scheduled sampling. Regular sampling allows monitoring of the rate of change of the measured parameters. This topic will be discussed in more detail later in chapter 8.

**Aspects of Oil Analysis**

The tests described in chapter 8 monitor all or some of these three aspects of oil:

- Lubricant condition
- Contaminants
- Machine wear

**Oil Analysis Condition-monitoring Techniques**

**Atomic Emissions Spectroscopy**

Atomic emissions spectroscopy identifies problems with contaminants and additives, and to a limited extent, corrosion and wear metals in lubrication and hydraulic oil samples, by measuring the characteristic radiation emitted when samples are subjected to high energy and temperature conditions. The test results are in parts per million (ppm) for a wide variety of elements of interest, including iron, aluminum, chromium, copper, lead, tin, nickel, and silver, and components of oil additives such as boron, zinc, phosphorus, and calcium. This technique can be used to analyze oil used in diesel and gasoline engines, compressors, transmissions, gearboxes, and hydraulic systems.

**Ferrography**

Ferrography is a technique that identifies the density and size ratio of particles in oil or grease caused by problems such as wear, fatigue, corrosion or combination. A representative sample is diluted with a fixer solvent and then passed over an inclined glass slide that is subjected to a magnetic field. The magnetic field provides separation of the ferrous particles (ferrous particles align with the magnetic field lines) and distributes them along the length of the slide (nonmagnetic and nonmetallic particles are distributed randomly along the slide).

---

Prepared by: Dr. Essam Bahgat

essbahgat@gmail.com
3.3.7 Speed monitoring

Speed measuring instruments vary with the job. Whether measuring traveling speed, wind speed, acceleration or engine speed, there is a proper instrument for calculating that form of speed. Police use RADAR and LIDAR to measure traffic speed. Cars have speedometers for travel speed and tachometers for engine speed. Accelerometers are also used with car speed measurements.

**Tachometers for Engine or Motor Speed**

Tachometers measure the engine speed in revolutions per minute (rpm). This instrument determines the rotational speed. The classification of Tachometer is different in different books. Somewhere there is just broad classification whereas at other places classification is one the basis of principal, types of uses or construction.

**Tachometers classification**

- Analog Tachometers
- Digital Tachometers
- Time & Frequency measuring Tachometers
- Contact and Non contact Tachometers

**Speedometers for Travelling Speed**

---

*Fig.3: 13 Ferrography analyzer*

*Fig.3: 14 Tachometer types*
Speedometers measure the traveling speed of land vehicles. They are used to help drivers determine their driving speed and maintain it at save and realistic levels. These meters use magnets and a set of rotating cables connected to the transmission to determine the rate of travel, which is shown on an analog display on the vehicle's dashboard.

Fig.3: 15 speedometer
CHAPTER 4
REVIEW OF VIBRATION FUNDAMENTALS
4.1 INTRODUCTION

Machine vibration is one of the most convenient signals for the maintenance engineer to measure and analyze and thus get to know about the machine’s health condition. In this context the various basic concepts of machine vibration will be introduced in the chapter. A basic understanding of machine vibration will enable one to diagnose machinery faults and control machinery vibration.

Vibration is a representation of the motion of a body, is essentially oscillatory about a mean position, and can be periodic or aperiodic. This motion of the body can be described in any direction. However, the degree of freedom is the number of independent coordinates that can be used to describe the motions of a rigid body. For example, a rigid body in space has 6 degrees of freedom, which consist of three translatory motions and three rotational motions about these axes. Usually, for a large machinery supported at the four corners on a shop floor, the vibration motion in three directions is measured and analyzed. However, before discussing in details of the motion of bodies in more than one direction, the motion of a rigid mass in one direction is considered in order to get familiar with the associated terminology.

4.2 Single Degree-of-Freedom Motion

A rigid body of mass $m$ is supported in the axial direction by a linear spring of stiffness $k$ as shown in Figure 4.1. The body is only allowed to have motion in one direction. The equation of the motion of such a body is given as in Equation (4.1).

\[ m \frac{d^2x}{dt^2} + kx = 0 \]  

However, no system in real life is undamped, and the oscillator can be damped as shown in Figure 4.2. The
The equation of motion of such a damped harmonic oscillator is given in Equation (4.2).

\[
m\frac{d^2 x}{dt^2} + c\frac{dx}{dt} + kx = 0 \tag{4.2}
\]

The natural frequency of such a system is given as \( \omega_n = \sqrt{\frac{k}{m}} \)

The damping coefficient \( C \) is known as viscous damping, where the damping force is proportional to the velocity of the body. The damping factor is described as \( \xi = \frac{c}{2\sqrt{km}} \).

The response to such a damped oscillator is given in Equation (4.3)

\[
x(t) = A^{-\xi\omega t} \sin((\omega_d t + \phi)) \tag{4.3}
\]

where, \( \omega_d = \omega_n\sqrt{1 - \xi^2} \)

**4.3 Forced Vibration Response**

In actual rotating machines, there are many forms of excitation, like forces due to unbalanced rotors, forces at couplings due to shaft misalignment, dynamic forces at bearing locations due to loose components, and so on. The vibration response of such machines is quite different and their study is essential for understanding the cause behind such responses. The dynamic forces or *torques* that excite the machine system can be represented by harmonic functions. In fact, any type of excitation can be mathematically represented as a sum of harmonic functions, as is done by a Fourier series expansion (more in Chapter 5). The equation of motion of a damped single degree-of-freedom system subjected to an external force is given by Equation (4.4)

\[
m\frac{d^2 x}{dt^2} + c\frac{dx}{dt} + kx = F(t) \tag{4.4}
\]

For a harmonic force, the force can be represented as Equation (4.5).

\[
F(t) = F_0 \cos(\omega_f t) \tag{4.5}
\]

The excitation force is at a frequency \( \omega_f \). The system’s response is very much dependent on the frequency ratio, \( r \), which is given as \( r = \frac{\omega_f}{\omega_n} \)

The excitation frequency is sometimes known as *input frequency*, *excitation frequency*, or *forcing frequency*. In rotating machines, the rotational speed of the machine corresponds to this
frequency. The response of a damped harmonic oscillator to a harmonic force of a kind represented by Equation (3.5) is given by Equation (3.6). The first expression of Equation (3.6) is known as the transient response and the second expression is known as the steady-state response.

\[ x(t) = Ae^{-\alpha t} \sin(\omega_d t + \theta) + A_0 \cos(\omega_d t - \phi) \]  

[4.6]

where \( \frac{A_0k}{F_0} = \frac{1}{\sqrt{(1 - r^2)^2 + (2\zeta r)^2}} \) and \( \phi = \tan^{-1} \frac{2\zeta r}{1 - r^2} \).

The normalized response of the damped harmonic oscillator is shown in Figure (4.3). At resonance, when \( r = 1 \), the forcing frequency is equal to the natural frequency of the system, and the ratio of the amplitudes decreases with an increase of the damping ratio.

4.3.1 Base Excitation

There are instances when machine bases that are supported on some stiffness are excited by a harmonic motion, as indicated in Figure (4.4). The ratio of the steady-state response of the body \( X \) to external motion at the base \( Y \) can be represented by the expressions given in Equation (4.7). The ratio of the forces transmitted due to the base motion is given in Equation (4.8).

\[ \frac{X}{Y} = \sqrt{\frac{1 + (2\zeta r^2)}{(1 - r^2)^2 + (2\zeta r)^2}} \]  

[4.7]

\[ \frac{F_T}{F_y} = r^2 \sqrt{\frac{1 + (2\zeta r)^2}{(1 - r^2)^2 + (2\zeta r)^2}} \]  

[4.8]

Fig.4: 3 Magnitude response of a harmonically forced damped oscillator

Fig.4: 4 Body subjected to base excitation.

essbahgat@gmail.com
Transmitted in the vertical direction to the base or foundation of the machine. The ratio of the force transmitted to the applied force is known as the *force transmissibility* and is given by Equation (4.9).

\[
\frac{F_t}{F_0} = \frac{1 + (2\zeta r)^2}{\sqrt{(1 - r^2)^2 + (2\zeta r)^2}}
\]  

[4.9]

### 4.4 VIBRATION ISOLATORS

Vibration isolators are available on the market that are selected based on two important parameters, one of which is the maximum deflection that allows the base and the damping to support the payload, as given by Equation (3.10). Depending on the configuration of the isolator arrangement (usually for parallel arrangements of the isolators at the base), the effective spring stiffness is given by Equation (3.11).

\[
K_{\text{eff}} \delta_{\text{max}} = W
\]

[4.10]

\[
K_{\text{eff}} = N k_{\text{spring}}
\]

[4.11]

For vibration and force isolation, usually helical springs, wire rope springs, pneumatic vibration isolators and elastomeric pads are used in industry. A typical wire rope spring used as a machinery mount for vibration isolation is shown in Figure (4.5). These types of springs have some inherent damping present in them, which can be anywhere from 5% to 15%.
The second important parameter in the selection of isolators is that of stiffness, so that the effective natural frequency of the system in the direction of the force transmission is less than $\frac{\omega_f}{\sqrt{2}}$

![Wire rope vibration isolator.](image1)

![Elastomeric pads](image2)

![Coil spring vibration isolator](image3)

![Pneumatic Vibration Isolators](image4)

**Fig.4:5 Different types of vibration isolators**

### 4.5 Tuned Vibration Absorber

Tuned mass dampers (TMDs) work by fastening a mass block to a structural component (such as a floor) via a spring. This system is set up so that, when the floor vibrates at a resonant frequency, it induces analogous movement of the mass block and spring. By the conservation of energy, the TMD motion in turn reduces the amplitude of the system vibration. A damping device (dashpot) is usually connected in parallel with the spring between the mass-block and system, increasing the TMD's effectiveness over a range of frequencies and taking a small amount of mechanical energy out of the system as heat. The vibration response of a system is predominantly at the forcing frequency. However, there are instances where the forcing frequency may be equal to the natural frequency of the system; then the system will undergo...
resonance and undergo large motions. A designer usually ensures that the operating frequency or the forcing frequency of excitation of the machine is not close to its natural frequency. However, in instances where such a coincidence of the frequencies cannot be avoided, another secondary spring mass system is usually attached to the primary machine, so that the natural frequency of the primary system is shifted and the condition of resonance is avoided. Such a secondary system is known as a tuned dynamic absorber. A tuned absorber attached to a primary mass is shown in Figure (4.6). The secondary system is selected so that the natural frequency of both the primary and secondary systems are the same, as shown in Equation (4.12).

\[
\omega_n = \sqrt{\frac{k_p}{m_p}} = \sqrt{\frac{k_s}{m_s}} \tag{4.12}
\]

Fig.4:6 Moose damper under resonance test on slip table with dynamic model representation

**4.6 Unbalanced Response**

A common machinery defect is rotating unbalance. If a rotating machine of mass \( M \) has an unbalance mass, \( m_0 \), at a radius of \( e \) from the center of rotation, and the machine is only allowed to have vertical motion the amplitude of vibration response of the unbalanced machine is given by Equation (4.13). The response of such an unbalanced system is shown in Figure (4.7). In order to reduce this response at operating speeds, care is taken to reduce the unbalance mass, \( m \).

\[
X = \frac{m_0 e}{M} = \frac{r^2}{\sqrt{(1 - r^2)^2 + (2\zeta r)^2}} \tag{4.13}
\]
4.7 Characteristics of Vibrating Systems

In machinery condition monitoring, the response of a system is measured, analyzed, and an understanding of the system is achieved. However, the vibrating system’s characteristics can be known by measuring its transfer function or frequency response function. The transfer function is usually represented in the Laplace domain and the frequency response function (FRF) is represented in the frequency domain. The FRF is the ratio of the response of the system to its excitation. There is a phase difference between the response and the excitation and FRF is a complex quantity.

4.7.1 Experimental Modal Analysis

In many instances, the vibration at the measurement points in machines may increase due to an unavoidable resonance condition. This resonance happens because the external forcing frequency has matched that of one of the resonant frequencies of the machine. The source of the external forcing frequency is usually due to the operating speed of the machinery, and many times it is observed that by shifting the operating speed, the resonant condition or the maximum vibration suddenly decreases. Of course, the machines are designed so that the system’s natural or resonance frequencies are not close to the machine’s operating speed. However, in many processing plants for example, if there is a stirrer mounted on top of a storage tank, it has been observed that sudden violent vibrations occur on the tank at a certain filling position. This is due to the fact that, at a certain minimum level of the tank, the mass of the storage tank changes, which increases the corresponding natural frequency. While troubleshooting such excessive vibration phenomena, it is desirable to do an in situ estimation of the natural frequency of the system.

Experimental modal analysis is an experimental technique to determine the natural frequencies, associated damping at the natural frequencies, and the mode shapes of the few major modes of vibration.
the machinery. To determine the frequency response of the system, the system has to be excited and the response measured at the desired points. Depending upon the number of excitation points and the response points, the system can be called as a single-input single-output (SISO) or a multiple-input multiple-output (MIMO) system. Depending on the response and excitation type, the FRFs are obtained. The two most common types of excitation system used for modal analysis are:

**The impact type**, using an instrumented hammer with a force transducer at the tip of the hammer.

**The random excitation type**, using an electromagnetic exciter driven by a random noise. At the tip of the stinger that is attached to the electromagnetic exciter a force transducer is attached. The response of the structure is usually measured by any contact or noncontact vibration transducer.

**Impact Testing**

One of the common excitation mechanism in modal testing is an impact device. Although it is a relatively simple technique to implement, it’s difficult to obtain consistent results. The convenience of this technique is attractive because it requires very little hardware and provides shorter measurement times. The method of applying the impulse, shown in Figure (4.8) for both case in free structure and fixed one, includes a hammer, fixed or a suspended mass and accelerometer. Since the force is an impulse, the amplitude level of the energy applied to the structure is a function of the mass and the velocity of the hammer. This is due to the concept of linear momentum, which is defined as mass time’s velocity. The linear impulse is equal to the incremental change in the linear momentum. It is difficult though to control the velocity of the hammer, so the force level is usually controlled by varying the mass. Impact hammers are available in weights varying from a few ounces to several pounds. Also, mass can be added to or removed from most hammers, making them useful for testing objects of varying sizes and weights. The frequency content of the energy applied to the structure is a function of the stiffness of the contacting surfaces and, to a lesser extent, the mass of the hammer. The stiffness of the contacting surfaces affects the shape of the force pulse, which in turn determines the frequency content.
Shaker Testing

The most useful shakers for modal testing with the electromagnetic shaker force is generated by an alternating current that drives a magnetic coil. To begin with, the shaker is physically mounted to the structure via the force transducer, thus creating the possibility of altering the dynamics of the structure. With lightweight structures, the mechanism used to mount the load cell may add appreciable mass to the structure. This causes the force measured by the load cell to be greater than the force actually applied to the structure. Figure (4.9) describes how this mass loading alters the input force. Since the extra mass is between the load cell and the structure the load cell senses this extra mass as part of the structure.

Frequency Response Measurements

The type of modal testing known as the frequency response function method, which measures the input excitation and output response simultaneously, as shown in the block diagram in

---

Prepared by: Dr. Essam E.
Figure (4.10), is examined. With a dynamic signal analyzer, which is a Fourier transform-based instrument, many types of excitation sources can be implemented to measure a structure’s frequency response function. In fact, virtually any physically realizable signal can be input or measured. The selection and implementation of the more common and useful types of signals for modal testing are discussed. Transducer selection and mounting methods for measuring these signals along with system calibration methods, are also included. Techniques for improving the quality and accuracy of measurements are then explored. These include processes such as averaging, windowing and zooming, all of which reduce measurement errors. Finally, a section on measurement interpretation is included to aid in understanding the complete measurement process.

Structural Dynamics of a Single Degree of Freedom (SDOF) System

The natural frequency, $\omega$, is in units of radians per second (rad/s). The typical units displayed on a digital signal analyzer, however, are in Hertz (Hz). The damping factor can also be represented as a percent of critical damping – the damping level at which the system experiences no oscillation. This is the more common understanding of modal damping. Although there are three distinct damping cases, only the underdamped case ($\zeta < 1$) is important for structural dynamics applications. When there is no excitation, the roots of the equation are as shown in Equation (4.14). Each root has two parts: the real part or decay rate, which defines damping in the system and the imaginary part, or oscillatory rate, which defines the damped natural frequency, $\Omega_d$ this free vibration response is illustrated in Figure (4.11).

$$S_{1,2} = -S + j\omega_d$$  \[4.14\]

**Prepared by: Dr. Essam Bahgat**

essbahgat@gmail.com
Determination of Natural Frequency and Damping Ratio

Mechanical Vibration
Mechanical Vibration is defined as the motion of a system (a particle or a body) which oscillates about its stable equilibrium position. Mechanical Vibration generally results when a system is displaced from a position of stable equilibrium. The system tends to return to its equilibrium position by virtue of restoring forces. However, the system generally reaches its original position with certain acquired velocity that carries it beyond that position. Ideally, this motion can repeat indefinitely.

Free Vibration
When the vibration motion is maintained by the restoring forces only, the vibration is termed as free vibration.

Natural frequency
Natural frequency is defined as the lowest inherent rate (cycles per second or radians per second) of free vibration of a vibrating system. Its unit is Hz or rad s\(^{-1}\) and it is designated by \(\omega_n\).

Damping
Damping is dissipation of energy in an oscillating system. It limits amplitude at resonance. All vibrating systems are damped to some degree by friction forces. These forces can be caused by dry friction or Coulomb friction, between rigid bodies, by fluid friction when a rigid body moves in a fluid, or by internal friction between the molecules of a seemingly elastic body.

Viscous damping and Coefficient of viscous damping
Viscous damping is caused by fluid friction at low and moderate speeds. It is characterized by the fact that the friction force is directly proportional and opposite to the velocity of the moving body.

The magnitude of the friction force exerted on the plunger by the surrounding fluid is equal to \(cx'\). Where \(c\) is known as the coefficient of viscous damping expressed in N s/m. It depends on the physical properties of the fluid and depends on the construction of the dashpot.

Critical damping coefficient
Assuming that the motion of the system is defined by the following differential equation: \(mx'' + cx' + kx = 0\)
The motion is termed as critically damped when the coefficient of viscous damping equals 2 m
ω₀ and it is designated by cc.

Damping ratio

Damping ratio is defined as the ratio of the coefficient of viscous damping to critical damping
coefficient. It is designated by ζ.

Measurement of damping ratio experimentally - Logarithmic Decrement

A convenient way to measure the amount of damping present in a system is to measure the rate
of decay of free oscillations. The larger the damping, the greater is the rate of decay.

Rate of decay of the oscillation
Considering a damped vibration expressed by
the general equation:

\[ x = X e^{-\zeta \omega_n t} \sin(\sqrt{1-\zeta^2} \omega_n t + \varphi) \]

Logarithmic decrement can be defined as the
natural logarithm of the ration of any two
successive amplitudes.

\[ \delta = \ln \frac{x_{n-1}}{x_n} = \frac{1}{n} \ln \frac{x_0}{x_n} \]

\[ \delta = \zeta \omega_n \tau_d \approx 2\pi \zeta \]

If the time period of the oscillations is T_d, then the damped natural frequency is given

\[ \omega_d = \frac{2\pi}{T_d} \]

From the damped natural frequency and damping ratio, the undamped natural frequency can be
calculated using

\[ \zeta = \frac{1}{\sqrt{1 + \left( \frac{2\pi}{\delta} \right)^2}} \]

\[ \omega_n = \frac{\omega_d}{\sqrt{1 - \zeta^2}} \]

Examples 1

Prepared by: Dr. Essam Bahgat
essbahgat@gmail.com
The free response of the damped single degree of freedom with mass of 2 kg is recorded to be of the form given in figure below. A static deflection test is performed and the stiffness is determined to be $1.5 \times 10^3$ N/m. The displacements at time $t_1$ and $t_2$ are measured to be 9 and 1 mm, respectively. Calculate the damping coefficient.

\[ \delta = \ln \left( \frac{x(t_1)}{x(t_2)} \right) = \ln \left( \frac{9 \text{ mm}}{1 \text{ mm}} \right) = 2.1972 \]

\[ \zeta = \frac{2.1972}{\sqrt{4\pi^2 + 2.1972^2}} = 0.33 \quad \text{or} \quad 33\% \]

\[ c_{cr} = 2\sqrt{km} = 2\sqrt{(1.5 \times 10^3 \text{ N/m})(2 \text{ kg})} = 1.095 \times 10^2 \text{ kg/s} \]

\[ c = c_{cr}\zeta = (1.095 \times 10^2)(0.33) = 36.15 \text{ kg/s} \]

**Example 2**

Determine the natural frequency, damping ratio and damped natural frequency

---

*Prepared by: Dr. Essam Bahgat*

essbahgat@gmail.com
Solution

\[ T_d = \frac{2\pi}{\omega_d} \]

\[ \omega_d = \frac{\pi}{2} = 1.5708 \]
\[ x_1 = 0.74 \]
\[ x_2 = 0.21 \]
\[ x_3 = 0.6 \]

\[
\delta = \ln \frac{x_1}{x_2} = \ln \frac{0.74}{0.21} = 1.2595
\]
\[
\delta = \ln \frac{x_2}{x_3} = \ln \frac{0.21}{0.06} = 1.2528
\]
\[
\delta = \frac{1}{2} \ln \frac{x_1}{x_3} = \frac{1}{2} \ln \frac{0.74}{0.06} = 1.2562
\]

\[
\xi = \frac{\delta}{\sqrt{(2\pi)^2 + \delta^2}} = 0.19604
\]
Structural Dynamics for a Multiple Degree of Freedom (MDOF) System

The physical system is simply comprised of an interconnection of idealized SDOF models, and is described by the matrix equations of motion as illustrated in Figure (4.12), and is described by the matrix equations of motion as illustrated in Equation (4.15). The equations of motion for the forced vibration case also lead to frequency response of the system. The solution of the equation with no excitation again leads to the modal parameters (roots of the equation) of the system. For the MDOF case, however, a unique displacement vector called the mode shape exists for each distinct frequency and damping. The free vibration response is illustrated in Figure (4.13).

Fig. 4:12 MDOF discrete parameter model
\[ [m]\{x\} + [c]\{x\} + [k]\{x\} = \{f(t)\} \]

\[ \{f\} r, r = 1, n \text{ modes} \]  

It can be written as a weighted summation of SDOF systems shown in Figure (4.14).

The weighting, often called the modal participation factor, is a function of excitation and mode shape coefficients at the input and output degrees of freedom. The participation factor identifies the amount each mode contributes to the total response at a particular point. An example with 3 degrees of freedom showing the individual modal contributions is shown in Figure (4.15).
4.7.2 Mode Shapes and Operational Deflection Shapes

Modes are used as a simple and efficient means of characterizing resonant vibration. The majority of structures can be made to resonate. That is, under the proper conditions, a structure can be made to vibrate with *excessive, sustained, oscillatory motion*. Resonant vibration is caused by an interaction between the *inertial* and *elastic* properties of the materials within a structure. Resonant vibration is often the cause of, or at least a contributing factor to many of the vibration related problems that occur in structures and operating machinery.

**Mode shapes**

Mode shapes are the loci of points on the structure that are displaced when the structure is vibrating at one of its natural frequencies. On the structure, there are points where there is no displacement at a particular natural frequency and points at which there is maximum displacement; these are known as the antinode points.
Easier To Understand In Terms Of Modes

Figure (4.17) points out another reason why vibration is easier to understand in terms of modes of vibration. It is a plot of the Log Magnitude of an FRF measurement (the solid curve), but several resonance curves are also plotted as dotted lines below the FRF magnitude. Each of these resonance curves is the structural response due to a single mode of vibration. The overall structural response (the solid curve) is in fact, the summation of resonance curves. In other words, the overall response of a structure at any frequency is a summation of responses due to each of its modes.

4.7.3 Operating Deflection Shape

Traditionally, an ODS has been defined as the deflection of a structure at a particular frequency. However, an ODS can be defined more generally as any forced motion of two or more points on a structure. Specifying the motion of two or more points defines a shape.

Why Measure ODS’s?

Measuring ODS’s can help answer the following vibration related questions,

- How much is a machine moving?
- Where is it moving the most, and in what direction?
- What is the motion of one point relative to another (Operating Deflection Shape)?
- Is a resonance being excited? What does its mode shape look like?
- Is there structure-born noise?
- Do corrective actions reduce noise or vibration levels?

4.8 VIBRATION TERMINOLOGY

Prepared by: Dr. Essam Bahgat
Vibration means the state of an object moving repetitively back/forward, right/ left or up/down and is generally expressed by Frequency, Displacement, Velocity, and Acceleration. These 4 elements are generally denoted as F, D, V, A. This is illustrated simply as a spring and mass. When the mass is pulled down from the start position and released, the mass moves just like the vibration waveform shown in the graph on the right.

To understand well the vibration analysis, you should know the common terms used in the vibration analysis. As discussed before there are three main factors are control the vibration amplitude (mass, spring and damper). Amplitude is an indicator of a vibration level or severity. Amplitude can be expressed as one of the three engineering units:

**Displacement**

**Velocity**

**Acceleration**

Before explaining the previous terms you should understand the harmonic motion.

### 4.8.1 Simple Harmonic Motion

Simple harmonic motion is a type of periodic motion or oscillation motion where the restoring force is directly proportional to the displacement and acts in the direction opposite to that of displacement as shown in figure (4.17).

$$ y = A \sin \omega t = A \sin \sqrt{\frac{k}{m}} t $$

Fig.4:17 Simple harmonic motion

**Frequency**

Frequencies the rate of mechanical oscillation in a period of time. Frequency can be expressed in one of the following units: RPM - Revolutions per Minute CPM - Cycles per Minute CPS - Cycles per Second Hz - Hertz, 1 Hz - 1 Cycle per Second (to convert from Hz to RPM or CPM, apply the following formula: Hz * 60 = RPM.

**Displacement**
Displacement means the amplitude (distance) between the peaks of vibration. Unit: μm, mm.

The difference between the final and initial position of a point (for instance, the center of mass of a moving object). It is measured by meter or any length units [mm, m, in, ft, etc…].

**Velocity**

Velocity means the changing rate of displacement (D) to time. Unit: mm/s, cm/s.

The **velocity of an object is the rate of change of its position with respect to** a frame of reference we can calculate the velocity from the reference circle representation in Equation (4:16). Linear Velocity measured by m/s, and angular velocity measured by rad/s.

\[
v_T = r\omega = A\omega \\
-v_T \sin \omega t = -A\omega \sin \omega t \\
v = -A\omega \sin \omega t \quad [4.16]
\]

**Acceleration**

Acceleration means the changing rate of velocity(V) to time. Unit: m/s², g.

Acceleration, is the rate of change of velocity of an object with respect to time. For harmonic motion the acceleration can be expressed in the Equation (4.17). Acceleration measured by m/s².

\[
a_C = r\omega^2 = A\omega^2 \\
a = -a_C \cos \omega t \\
a = -A\omega^2 \cos \omega t \quad [4.17]
\]

Amplitude units can be converted from one to another unit in the conversion format in figure (4.18).
4.8.2 Measurement functions

These are the means to express the vibration in absolute value basing upon the waveforms measured by the modes shown above. Since vibration is transmitted as an AC signal, there are four Unit Modifiers that may be used to condition the signal. These modifiers have a direct impact on the measurement value. If the wrong modifier is used, the measurement could be either too high, or too low, thus causing possible maintenance action to be, or not to be, accomplished erroneously.

**Unit Modifiers:**

- **Peak to Peak** - the distance from the top of the positive peak to bottom of the negative peak.
- **Peak** - the measurement from the zero line to the top of the positive peak.
- **Average (AVG)** - .637 of peak.
- **Root Mean Square (RMS)** - .707 of peak.

---

**PEAK (Peak amplitude)**

---

*Prepared by: Dr. Essam Bahgat*

essbahgat@gmail.com
Peak value in a certain time duration. It is used to measure shocks or waves which are rather stable.

\[ V_{\text{PEAK}} = |v(t)|_{\text{max}} \]

**rms (Root Mean Square value)**

Root mean square value of the instantaneous values in a certain time duration. It relates to the power of the wave. The rms value of velocity is one of the important factors for machinery status diagnosis.

\[ V_{\text{rms}} = \sqrt{\frac{1}{T} \int_{t1}^{t2} v(t)^2 \, dt} \]

**4.8.3 Types of vibration**

Vibration can be divided into three types based on human perceptions: "slow motion and visible" "invisible but sensible by touching" and "insensible by touching, but audible as abnormal noise".
4.8.4 Types of Vibration signal

Vibration can be classified into one or more of the following categories:

- Periodic
- Random
- Resonant
- Harmonic

**Periodic**

Repeats itself once every time period as shown in figure (4.19), Result of a mass imbalance in a component or disc.

![Fig. 4.19 Periodic signal](image)

---

**Random**

---

*Prepared by: Dr. Essam Bahgat*

essbahgat@gmail.com
Do not repeat themselves, not related to a fundamental frequency. An example - the shock that is felt as a result of driving down the road and hitting a pothole.

Harmonic

Spectrum containing only frequency components whose frequencies are whole number multiples of the fundamental frequency or exact multiples of a fundamental frequency. Classified in terms as 1\textsuperscript{st}, 2\textsuperscript{nd}, 3\textsuperscript{rd} order as shown in figure (4.20)

Resonant

The natural frequency at which an airframe or mechanical system is inclined to vibrate. All things have one or more resonant frequencies. Resonant vibrations are the result of a response in a mechanical system to a periodic driving force. Figure (4.21) show the resonant frequency range for an accelerometer.
Fig. 4.21 The typical frequency response of a piezoelectric accelerometer
CHAPTER 5
VIBRATION SENSOR AND TRANSDUCERS
Measurement is an important subsystem of a mechatronics system. Its main function is to collect the information on system status and to feed it to the micro-processor(s) for controlling the whole system. Measurement system comprises of sensors, transducers and signal processing devices. Today a wide variety of these elements and devices are available in the market. For a mechatronics system designer it is quite difficult to choose suitable sensors/transducers for the desired application(s). It is therefore essential to learn the principle of working of commonly used sensors/transducers.

**Transducer**

It is defined as an element which converts a primary form of energy into a corresponding signal with a different energy form Primary Energy Forms: mechanical, thermal, electromagnetic, optical, chemical, etc...

**Sensor**

Is a device that detects a change in a physical stimulus and turns it into a signal which can be measured or recorded. or " A device which provides a usable output in response to a specified measured."

**What is the difference between a sensor and a transducer?**

*The main difference between a sensor and a transducer is their use. A transducer is any device that can be used to convert energy from one form to another.*

*An example of a transducer is an antenna, which can be used to convert electricity to electromagnetic waves and vice versa. A sensor also converts one form of energy to another, mainly to electric signals, which can then be used for processing.*

**5.1 PERFORMANCE AND TERMINOLOGY**

The desirable features of sensors are:

1. Range / span
2. Errors and accuracy
3. Nonlinearity
4. Hysteresis
5. Dead band and Saturation
6. Output impedance
7. Repeatability
8. Reliability
9. Sensitivity

---

*Prepared by: Dr. Essam Bahgat*

essbahgat@gmail.com
10. Resolution
11. Frequency Response
12. Response time
13. Calibration

**Range and Span**

**Range:** lowest and highest values of the stimulus. For example, a thermocouple for the measurement of temperature might have a range of 220 – 20 °C.

**Span:** the arithmetic difference between the highest and lowest values of the input that being sensed. The span is difference between the maximum and minimum values of the input. Thus, the above-mentioned thermocouple will have a span of 200°C.

**Error and accuracy**

**Error** is the difference between the result of the measurement and the true value of the quantity being measured. A sensor might give a displacement reading of 29.8 mm when the actual displacement had been 30 mm, then the error is – 0.2 mm.

**Accuracy** defines the closeness of the agreement between the actual measurement result and a true value of the measured. It is often expressed as a percentage of the full range output or full-scale deflection.

**Nonlinearity**

The nonlinearity indicates the maximum deviation of the actual measured curve of a sensor from the ideal curve.

**Hysteresis**

Is the deviation of the sensor’s output at any given point when approached from two different directions. If temperature is measured, at a rated temperature of 50°C, the output might be 4.95V when temperature increases but 5.05V when temperature decreases. This is an error of ±0.5% (for an output full scale of 10V in this idealized example).

**Dead band**

The lack of response or insensitivity of a device over a specific range of the input. In this range which may be small, the output remains constant. A device should not operate in this range unless this insensitivity is acceptable.
Repeatability

Also called *reproducibility*: failure of the sensor to represent the same value under identical conditions when measured at different times. Usually associated with calibration given as percentage of input full scale of the maximum difference between two readings taken at different times under identical input conditions.

Repeatability = \((\text{maximum} – \text{minimum values given}) \times 100 / \text{full range}\).

Reliability

A statistical measure of quality of a device which indicates the ability of the device to perform its stated function, under normal operating conditions without failure for a stated period of time or number of cycles.

Sensitivity

*Sensitivity* of a sensor is defined as the change in output for a given change in input, usually a unit change in input. Sensitivity represents the slope of the transfer function.

For example, a general purpose thermocouple may have a sensitivity of \(41\mu\text{V}^\circ/\text{C}\).

Resolution

Resolution is the smallest detectable incremental change of input parameter that can be detected in the output signal. Resolution can be expressed either as a proportion of the full-scale reading or in absolute terms. For example, if a LVDT sensor measures a displacement up to 20 mm and it provides an output as a number between 1 and 100 then the resolution of the sensor device is 0.2 mm.

Frequency response

The ability of the device to respond to a harmonic (sinusoidal) input. Indicates the range of the stimulus in which the device is usable (sensors and actuators) as shown in figure (5.1), also provides important design parameters.

*Bandwidth*, is the difference between the upper and lower frequencies in a continuous set of frequencies. It is typically measured in hertz, also referred to as Frequency Response, is an indication of a displacement sensor's ability to respond to changes in the measured displacement.
Response time
Indicates the time needed for the output to reach steady state for a step change in input.
• Typically the response time will be given as the time needed to reach 90% of steady state output upon exposure to a unit step change in input as shown in figure (5.2).
• The response time of the device is due to the inertia of the device (both “mechanical” and “electrical”).
• Fast response time is usually desirable
• Slow response times tend to average readings

Calibration
The experimental determination of the transfer function of a sensor or actuator.
• Typically, needed when the transfer function is not known or, • When the device must be operated at tolerances below those specified by the manufacturer.
• Example, use a thermistor with a 5% tolerance on a full scale from 0 to 100°C to measure temperature with accuracy of, say, ±0.5°C.

**5.2 VIBRATION SENSORS**

The key to proper machine monitoring however is the proper choice of sensor for the particular installation. Without the proper sensor, the best instrumentation and software available will not provide the definitive information on which to make a “sound engineering determination” regarding the mechanical operating condition or deficiencies of the machine.

**Vibration Sensors types**

The three parameters representing motion detected by vibration monitors are displacement, velocity, and acceleration. These parameters can be measured by a variety of motion sensors and are mathematically related (displacement is the first derivative of velocity and velocity is the first derivative of acceleration). Selection of a sensor proportional to displacement, velocity or acceleration depends on the frequencies of interest and the signal levels involved

**5.2.1 Displacement and position sensors**

Displacement sensors are basically used for the measurement of movement of an object. Position sensors are employed to determine the position of an object in relation to some reference point.

**Eddy current proximity sensors**

Eddy current proximity sensors are used to detect non-magnetic but conductive materials. They comprise of a coil, an oscillator, a detector and a triggering circuit as shown in figure (5.3). The Eddy Currents generate their own magnetic field these fields have interaction with the coil through mutual inductance this leads to a measurable result as shown in figure (5.4).

---

**Prepared by: Dr. Essam Bangat**

essbahgat@gmail.com
Displacement Probes applications

- Sleeve bearings
- Most turbine and large motor
- driven machinery
- Main turbine generators
- Boiler feed water pumps
- Shaft centerline from DC gap data
- Orbits from vibration data

Sensor mounting methods

Only vertical and horizontal position as shown in figure (5.5)

Types of displacement transducers
<table>
<thead>
<tr>
<th>Potentiometer</th>
<th>Strain-gauged element</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Potentiometer" /></td>
<td><img src="image2" alt="Strain-gauged element" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Capacitive element</th>
<th>linear variable differential transformer (LVDT)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image3" alt="Capacitive element" /></td>
<td><img src="image4" alt="LVDT" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inductive proximity switch</th>
<th>Optical encoders</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image5" alt="Inductive proximity switch" /></td>
<td><img src="image6" alt="Optical encoders" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pneumatic sensors</th>
<th>Hall effect sensors</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image7" alt="Pneumatic sensors" /></td>
<td><img src="image8" alt="Hall effect sensors" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Eddy current proximity sensors</th>
<th>Proximity switches (magnetic)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image9" alt="Eddy current proximity sensors" /></td>
<td><img src="image10" alt="Proximity switches" /></td>
</tr>
</tbody>
</table>

### 5.2.2 Velocity transducer

Prepared by: Dr. Essam Bahgat

essbahgat@gmail.com
**Velocity Pickup**

When a coil of wire is moved through a magnetic field, a voltage is induced across the end wires of the coil. The induced voltage is caused by the transferring of energy from the flux field of the magnet to the wire coil. As the coil is forced through the magnetic field by vibratory motion, a voltage signal representing the vibration is produced.

![Diagram of velocity pickup](image)

*Fig 5: 6 Velocity Pickups*

Use the principle of electromagnetic induction: linear and angular velocity transducer as shown in figure (5.7)

![Diagram of linear and angular velocity measurement](image)

*Fig 5: 7 Linear velocity measurement and Angular velocity measurement*

*There are two primary types of velocity transducers:*

---

*Prepared by: Dr. Essam Bahgat*

essbahgat@gmail.com
• **Moving coil type**

Larger than other types of transducers. It is not reliable for low frequency measurements without the use of a correction factor because the coil tends to follow the magnet at low frequencies. Balancing at lower frequencies ok since the relative amplitudes recorded will be consistent.

• **Piezoelectric type.**

Piezoelectric Effect

A piezoelectric material produces an electric charge when it's subject to a force or pressure. The piezoelectric materials such as quartz or polycrystalline barium titanate, contain molecules with asymmetrical charge distribution. Therefore, under pressure, the crystal deforms and there is a relative displacement of the positive and negative charges within the crystal.

Based on the piezoelectric effect theory. The charge is created by the piezoelectric disks responding to the stress of whatever applied forces there may be. Generally require a built-in amplifier due to the extremely small signal generated. There are thermal sensitivities and ambient temperature limitations involved. Typical and illustrated piezoelectric for velocity transducer shown in figure (5.9)

5.2.3 **Accelerometers**

---

*Prepared by: Dr. Essam Bahgat*

essbahgat@gmail.com
Accelerometers are used to detect the motion of an object or point in inertial space. Most accelerometers use the mass-spring-damper system as shown in figure (5.10), under a steady acceleration, the mass will move, stretching or compressing the spring until the force exerted by spring balance the force by the force due to acceleration.

**Common Accelerometer Types**

- Resistive
  - Strain Gauge
  - Piezoresistive
  - Micromachined
  - Thin-Film
- Capacitive
- Fiber Optic
- Servo or Force Balance
- Vibrating Quartz
- Piezoelectric

Magnetic methods and electrostatic (capacitive) methods are quite commonly used. The distance between the mass and a fixed surface, which depends on acceleration can be made into a capacitor. Capacitance increases (or decreases) with acceleration.

A magnetic sensor can be used by measuring the field of a magnetic mass. The higher the acceleration, the closer (or farther) the magnet from a fixed surface and hence the larger or lower the magnetic field.

**Resistive Operating Principle**
Voltage output of resistor bridge changes proportionally with applied acceleration as shown in figure (5.11).

Fig 5: 11 Pizoresistive accelerometer types

**Capacitive Operating Principle**
Utilizes frequency modulation technique through varying capacitor bridge as shown in figure (5.12).

Fig 5: 12 Capacitive accelerometer type

**Fiber Optic Operating Principle**
Amount of light gathered by receivers is proportional to applied acceleration as shown in figure (5.13).

Fig 5: 13 Fiber Optic accelerometer

**Vibrating Quartz**
Resonant frequency difference between elements is proportional to applied acceleration as shown in figure (5.14).
Piezoelectric

Force on self-generating crystal provides charge output proportional to acceleration. Figure (5.15) illustrates the Operating Principle of piezoelectric accelerometer. Some typical types of accelerometer are shown in figure (5.16).

Fig 5.14 Quartz Accelerometer

Fig 5.15 Schematic drawing for piezoelectric accelerometer

Fig 5.16 Typical types based on piezoelectric principle

5.3 SENSOR SELECTION
Selecting the best accelerometer for a specific predictive maintenance application can be a daunting task unless you determine what you need from the sensor to detect and the factors should be considered during measurement process.

- Frequency Range
- Minimum vibration amplitude
- Maximum vibration amplitude
- Operating temperature range
- Environment (fluids, gases, chemicals)
- Mounting method

Frequency Range Considerations
What is the minimum and maximum frequency to be measured? Some examples for estimating maximum frequency to be measured.

- Roller Element Bearings - 20 to 40 times the shaft RPM
- Journal Bearings - 10 to 20 times the shaft RPM
- Gear Boxes - 3.5 times the gear mesh
- Electric Motors - 3.5 times the rotor bar frequency, or 3.5 times the line current frequency
- Pumps & Fans - 3.5 times the blade pass frequency

Minimum Vibration Amplitude
- This is rarely a consideration for industrial applications except for very low frequency measurements.

The minimum amplitude
vibration should generate a signal at least 5 times the amplifier noise present at that frequency.

- Maximum possible vibration must not “overload” the sensor amplifier
- The sensor sensitivity should provide sufficient voltage output without creating distortion due to amplifier overload.

Operating Temperature Range
- The storage temperature is the same as the specified operating temperature.
- Higher input power current adversely affects maximum operating temperature (4mA maximum is recommended for high temperature)
- Low temperatures for industrial applications rarely go below -50°C.

Environmental Considerations
The environment in which the transducer will operate needs to be closely examined. This includes not only temperature and humidity, etc.… but also ingress protection. Another consideration hermetically sealed sensors and splash proof cables, or sensors with integral cables rated to IP68 should be used in “wet” environments.

**Mounting Considerations & Physical Constraints**

- How will the sensor be mounted? Stud, epoxy, quick link, magnet
- Is there sufficient room to mount the sensor?
- Side exit or top exit connector required?

### 5.4 ACCELEROMETER MOUNTING CONSIDERATIONS

For best accuracy, it is important that the mounting surface of the accelerometer be tightly coupled to the test surface to ensure the duplication of motion, especially at higher frequencies. Since various mounting methods may adversely affect accuracy, it is important to understand the mechanics and types of mounting the accelerometer for best results.

#### 5.4.1 Frequency range for accelerometer

Accuracy of frequency response is one of the most important factors when deciding which mounting technique to use. The accelerometer operating frequency range is determined by a back- to- back calibration method where the accelerometer under test is securely mounted to a reference standard accelerometer. The sensor under test should be stud mounted whenever possible for highest accuracy. Directly stud mounting an accelerometer to a very smooth surface yields the highest mounted resonant frequency allowing the broadest usable frequency range. The usable measuring range can be considered from the accelerometer frequency range chart as shown in figure (5.17).

![Accelerometer frequency range chart](image_url)

*Fig 5:17 Accelerometer frequency range chart*
5.4.2 Mounting methods

To obtain useful measurement information, an accelerometer must be coupled so that complete event information is transferred. Mounting methods may vary, with some transferring event information more effectively than others. A high performance accelerometer will behave like a low performance accelerometer if the mounting method is inadequate. The mounting method of an accelerometer should not only be considered by the ease of mounting criteria, but also by the corresponding upper frequency response for the chosen mounting method. Each mounting solution acts as a spring; the more flexible the coupling, the lower the frequencies that have to be produced in order to induce resonance phenomena as shown in figure (5.18). The most rigid connection is possible with a stud mount, while the least rigid can be obtained with magnetic mounting. The wax mounting and adhesive mounting methods fall somewhere in between, with layers as thin and hard as possible while still attaining optimal performance.

![Mounting Methods Diagram](image)

**Fig 5:18 Mounting types and bandwidth reduction from different mounting methods**
**Stud bolts**

Figure (5.19) show the stud mounting for best performance. For permanent mounting. Tapped hole in test object required.

![Stud mounting](image1)

**Adhesive Mount or wax**

Adhesive mounting methods as shown in figure (5.20) such as when mounting the accelerometer to sheet metal or to other surfaces where drilling a hole is not allowed.

![Adhesively mounted accelerometer](image2)

**Adhesive Mount pad or double side**

For quick mounting of light sensors at room temperature and low acceleration.

**Magnetic Mounting**

The accelerometer is attached to the magnet (usually by stud mount) as shown in figure (5.21), and the assembly is then applied to the surface.

![Magnet mounted accelerometer](image3)

**Hand-held probe**

A hand-held probe with the accelerometer mounted on top is very convenient for quick-look survey work, but can give gross measuring errors because of the low overall stiffness.
CHAPTER 6
SIGNAL PROCESSING
6.1 INTRODUCTION

As computers got more powerful, they started being used more frequently as diagnostic tools. However, when using computers, the Machinery Diagnostic Engineer has to consider an important fact: they work in a digital world. For this reason, the various vibration signals coming from the rotating equipment, which are “analog” in nature, need to be transformed into the digital domain. The term “analog” refers to the continuous nature of the signal; it contains an infinite number of amplitude levels, separated by infinitesimal time intervals. A digital signal, on the other hand, is not continuous, as it is formed by a finite number of amplitude levels, separated by finite time intervals as shown in figure (6.1). Even though a great amount of published information can be found on this subject, it tends to be too theoretical for the machinery diagnostic Engineer, making it difficult to evaluate the consequence of choosing the wrong parameters in the quality of the data, which in turn drives the effectiveness of the diagnosis.
The process through which an analog signal is transformed into digital is usually referred to as “sampling”. In Figure (6.2), the analog signal goes through a low pass filter, known as “Antialiasing filter” (AAF), which attenuates frequency components above the specified frequency. Let’s take a look at each part of the process with a little more detail. We start with the ADC and shall cover the antialiasing filters later.

**6.2 ANALOG TO DIGITAL CONVERTER**

An “Analog-to-digital” converter or ADC is a device incorporated in all modern data acquisition instruments, and is used to convert the input analog signal into a digital or discrete signal, so that it can be further processed. Analog-to-Digital converters (ADC) translate analog signals, real world signals like temperature, pressure, voltage, current, distance, or light intensity, into a digital representation of that signal. This digital representation can then be processed, manipulated, computed, transmitted or stored.

**Basic Operation**

An ADC samples an analog waveform at uniform time intervals and assigns a digital value to each sample. The digital value appears on the converter’s output in a binary coded format. The value is obtained by dividing the sampled analog input voltage by the reference voltage and then multiplying by the number of digital codes. The resolution of converter is set by the number of binary bits in the output code.
6.2.1 Advantage of Digital Signal Processing

- More flexible.
- Often easier system upgrade.
- Data easily stored.
- Better control over accuracy requirements.
- Reproducibility.

6.2.2 Bandwidth

The bandwidth of a signal composed of components of various frequencies (complex signal) is the difference between its highest and lowest frequency components, and is expressed in Hertz (Hz) - the same as frequency.

For example, a square wave may be constructed by adding sine waves of various frequencies:

**Width of signal spectrum:** \( B = F_{\text{highest}} - F_{\text{lowest}} \)
 Upper and lower frequency limits of the survey being acquired - either hardware set (with the use of an external band pass filter) or software controlled by the analyzer.  
Setting the frequency bandwidth is a way of eliminating vibration data or noise that is of no interest for your particular application. In the survey above, the frequency bandwidth is 0 CPM to 3000 CPM

6.3 FOURIER SERIES

When the French mathematician Joseph Fourier (1768–1830) was trying to solve a problem in heat conduction, he needed to express a function as an infinite series of sine and cosine functions:

\[ f(x) = a_0 + \sum_{n=1}^{\infty} \left( a_n \cos nx + b_n \sin nx \right) \]

\[ = a_0 + a_1 \cos x + a_2 \cos 2x + a_3 \cos 3x + \ldots + b_1 \sin x + b_2 \sin 2x + b_3 \sin 3x + \ldots \]

Many problems in physics involve vibrations and oscillations. Often the oscillatory motion is simple (e.g. weights on springs, pendulums, harmonic waves etc.) and can be represented as single sine or cosine functions. However, in many cases, (electromagnetism, heat conduction, quantum theory, etc.) the wave forms are not simple and, unlike sines and cosines, can be difficult to treat analytically. Fourier methods give us a set of powerful tools for representing any periodic function as a sum of sines and cosines as shown in figure (6.3).
This problems you can found it when you design a system like Mobile Communication Systems that you see it today; As Telecommunication engineer you need infinite Bandwidth (BW), if you send Square Pulse Wave as shown in figure (6.4), and that not possible because we can't make a transfer media can hold infinity number of data...so we escape to Fourier series to represent this square wave as cosine and sine wave with contain same data [0,1].
Example

1. Find the Fourier series for the function for which the graph is given by figure (6.5).

\[ f(t) = \begin{cases} 
-3 & -\pi \leq t < 0 \\
3 & 0 \leq t < \pi 
\end{cases} \]

We can see from the graph that it is periodic, with period \(2\pi\).

So \(f(t) = f(t + 2\pi)\).

Also, \(L = \pi\).

We can also see that it is an odd function, so we know \(a_0 = 0\) and \(a_n = 0\). So we will only need to find \(b_n\).

Since \(L = \pi\), the necessary formulae become:

\[ b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin nt \, dt \]

\[ f(t) = \sum_{n=1}^{\infty} b_n \sin nt \]

Now
\[ b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin nt \, dt \]
\[ = \frac{1}{\pi} \left( \int_{-\pi}^{0} -3 \sin nt \, dt + \int_{0}^{\pi} 3 \sin nt \, dt \right) \]
\[ = \frac{3}{\pi} \left( \left[ \cos nt \right]_{-\pi}^{0} + \left[ -\cos nt \right]_{0}^{\pi} \right) \]
\[ = \frac{3}{\pi n} (\cos 0 - \cos (-\pi n) - \cos \pi n + \cos 0) \]
\[ = \frac{3}{\pi n} (2 - \cos \pi n - \cos \pi n) \]
\[ = \frac{6}{\pi n} (1 - \cos \pi n) \]
\[ = \frac{12}{\pi n} \text{ (n odd)} \quad \text{or} \quad \]
\[ = 0 \quad \text{(n even)} \]

We could write this as: \[ b_n = \frac{12}{\pi (2n - 1)} \quad n = 1, 2, 3 \ldots \] (Substitute \( n = 1, 2, 3 \ldots \) to see how this works.)

So the Fourier series for our odd function is given by:

\[ f(t) = \sum_{n=1}^{\infty} b_n \sin nt \]
\[ = \sum_{n=1}^{\infty} \frac{12}{\pi (2n - 1)} \sin (2n - 1)t \]
\[ = \frac{12}{\pi} \sum_{n=1}^{\infty} \frac{\sin (2n - 1)t}{(2n - 1)} \]

**NOTE:** Since \( b_n \) is non-zero for \( n \) odd, we must also have odd multiples of \( t \) within the sine expression (the even ones are multiplied by 0, so will be 0).

Checking, we graph the first 5 terms. It should closely resemble the square wave we started with.

\[ \frac{12}{\pi} \sum_{n=1}^{5} \frac{\sin (2n - 1)t}{(2n - 1)} \]
\[ = \frac{12}{\pi} \left( \sin t + \frac{1}{3} \sin 3t + \frac{1}{5} \sin 5t + \frac{1}{7} \sin 7t + \frac{1}{9} \sin 9t \right) \]

Graph of \( f(t) \), the Fourier series approximation of a square wave.
Example 2

Given \( f(t) = |t|, \ t \in (-\pi, \pi) \) find its Fourier series equivalent?

Solution

The function must first be rewritten as

\[
f(t) = \begin{cases} 
-t, & -\pi < t < 0; \\
0, & 0 \leq t < \pi.
\end{cases}
\]

\[
a_n = \frac{1}{\pi} \int_{-\pi}^{0} -t \cos(nt) \, dt + \frac{1}{\pi} \int_{0}^{\pi} t \cos(nt) \, dt
\]

\[
= \frac{1}{\pi} \left[ \left( \cos(nt) \right) \frac{t}{n} \right]_{-\pi}^{0} + \frac{1}{\pi} \left[ \left( \cos(nt) \right) \frac{t}{n} \right]_{0}^{\pi}
\]

\[
= \frac{\cos(n\pi) - 1}{\pi n^2} + \frac{\cos(n\pi) - 1}{\pi n^2} = \frac{2(\cos(n\pi) - 1)}{\pi n^2}
\]

\[
a_n = \frac{2((-1)^n - 1)}{\pi n^2}
\]

and

\[
a_0 = \frac{1}{\pi} \int_{-\pi}^{0} -t \, dt + \frac{1}{\pi} \int_{0}^{\pi} t \, dt = \frac{\pi}{2} + \frac{\pi}{2} = \pi
\]

Similarly,

\[
b_n = \frac{1}{\pi} \int_{-\pi}^{0} -t \sin(nt) \, dt + \frac{1}{\pi} \int_{0}^{\pi} t \sin(nt) \, dt
\]

\[
= \frac{1}{\pi} \left[ \left( \sin(nt) \right) \frac{t}{n} \right]_{-\pi}^{0} + \frac{1}{\pi} \left[ \left( \sin(nt) \right) \frac{t}{n} \right]_{0}^{\pi}
\]

\[
= \frac{\sin(n\pi)}{n} - \frac{\sin(n\pi)}{n} = 0
\]

finally,

\[
f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left( a_n \cos(nt) \right) = \frac{\pi}{2} + \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^n - 1}{n^2} \cos(nt)
\]
6.4 FOURIER TRANSFORM

The Fourier Transform is a mathematical technique that transforms a function of time, $x(t)$, to a function of frequency, $X(\omega)$. It is closely related to the Fourier series. There are some terms you should be understand.

Aperiodic-Continuous
This includes, for example, decaying exponentials and the Gaussian curve. These signals extend to both positive and negative infinity without repeating in a periodic pattern. The Fourier Transform for this type of signal is simply called the Fourier Transform.

Periodic-Continuous
Here the examples include: sine waves, square waves, and any waveform that repeats itself in a regular pattern from negative to positive infinity. This version of the Fourier transform is called the Fourier Series.

Aperiodic-Discrete
These signals are only defined at discrete points between positive and negative infinity, and do not repeat themselves in a periodic fashion. This type of Fourier transform is called the Discrete Time Fourier Transform.

Periodic-Discrete
These are discrete signals that repeat themselves in a periodic fashion from negative to positive infinity. This class of Fourier Transform is sometimes called the Discrete Fourier Series, but is most often called the **Discrete Fourier Transform**.

### 6.4.1 Discrete Fourier Transform (DFT)

The Discrete Fourier Transform (DFT) is the equivalent of the continuous Fourier Transform for signals known only at instants separated by sample times (i.e. a finite sequence of data). The DFT equation:

\[
F(m) = \sum_{k=0}^{N-1} x(k) e^{-j2\pi km/N}
\]

x(k) is the time wave that is converted to a frequency spectrum by the DFT. Here are key concepts required to understand a DFT:

- The "sampling rate", [sr]. The sampling rate is the number of samples taken over a time period. For simplicity we will make the time interval between samples equal. This is the "sample interval", [si].
- The **fundamental period**, [T], is the **period** of all the samples taken. This is also called the "window".

---

Prepared by: Dr. Essam Bahgat

essbahgat@gmail.com
The "fundamental frequency" is \( f_0 \), which is \( 1/T \). \( f_0 \) is the first harmonic, the second harmonic is \( 2*f_0 \), the third is \( 3*f_0 \), etc.

The number of samples is \( N \).

The "Nyquist Frequency", \( f_c \), is half the sampling rate. The Nyquist frequency is the maximum frequency that can be detected for a given sampling rate. This is because in order to measure a wave you need at least two sample points to identify it (trough and peak).

"Euler's formula"  
\[ e^{ix} = \cos x + i \sin x \]

For simplicity, we will sample a sine wave with a small number of points, \( N \), and perform a DFT on it, then we will employ each of the concepts above. Note, the sine wave is a time wave, and could be any wave in nature, for example a sound wave. The horizontal axis is time. The vertical axis is amplitude.

Notice how in figure (6.6), we are sampling four points. The fundamental period, \( T \), of the wave sampled is set to \( 2*\pi \). This applies to any wave we want to sample. The interval between samples is \( 2*\pi/N \), so in this case it is \( 2*\pi/4 \). Thus, the interval between samples is \( \pi/2 \) in this case.

**Fig 6: 6 Time domain capture for sine wave at 0.1 s time.**
The time wave is thus, \( x(k) = \sin(\pi/2 \cdot k) \) for \( k = 0 \) to \( N - 1 \). The last point sampled is always the point just before \( 2\pi \), because the wave is considered to be a repeating pattern and wraps around back to the value at \( k = 0 \), so you aren't missing any information.

We also need to know the time taken to sample the wave, so that we can tie it to a frequency. In our example, the time taken for the fundamental period, \( T \), is 0.1 seconds (this value is measured when the wave is captured). That means the sine wave is a 10 Hz wave. Hertz = cycles per second. Also, the sampling interval, \( s_i \), is the fundamental period time divided by the number of samples. So, \( s_i = T/N = (0.1)/4 \) seconds, or 0.025 seconds. The sampling rate, or frequency, \( s_r \), = \( 1/s_i = 40 \) Hz, or 40 samples per second.

Evaluating the output data. Each \( F(n) \) value outputs a phase at a particular frequency. The frequency of the point is determined by the fundamental frequency multiplied by \( n \), i.e. \( f = f_0 \cdot n \), where \( f_0 = 1/T = 10\) Hz. The output values are the phase of the frequencies, which are represented by a real part and an imaginary part thus:

Real + j*Imaginary. The fundamental frequency, first harmonic, is 10 Hz as calculated above. The magnitude at a frequency is calculated thus:

\[
\text{sqrt}(\text{real}^2 + \text{imaginary}^2)
\]

The frequency plot is in the "frequency domain" as shown in figure (6.7). The magnitudes are plotted in Diagram 2. The spike at 10 Hz shows that the DFT pulled out one of the frequencies.

Fig 6: 7 Frequency domain representation

Prepared by: Dr. Essam Bahgat

essbahgat@gmail.com
that is in the sine wave. In fact, the sine wave is a 10 Hz sine wave, so that makes sense. However, the spike at 30 Hz should not be there, because there is no 30 Hz wave in the sine wave. So what accounts for that spike? Well, this is where the **Nyquist Frequency**, \( f_c \), mentioned above comes in. The Nyquist frequency is the cut off point above which the data from the DFT is no longer valid. The sampling rate is 40 Hz, and \( f_c \) is half the sampling frequency, which means that any frequency above 20 Hz will not be valid in this case. So, the 30 Hz frequency is a spurious signal.

### 6.4.2 Fast Fourier Transform (FFT)

The Fast Fourier Transform (FFT) is one of the most important algorithms in signal processing and data analysis. I've used it for years, but having no formal computer science background, it occurred to me this week that I've never thought to ask **how** the FFT computes the discrete Fourier transform so quickly.

**What is FFT (Fast Fourier Transform)?**

FFT is one method of analysis, based on vibration waveform. Generally, waveforms are complicated and difficult to analyze. In FFT, we break waveforms down into a series of discrete sin waves, (left chart) and evaluate each individually. (Right chart)

**Why use FFT?**

The only reason to learn the FFT is for speed. An FFT is a very efficient DFT calculating algorithm.
In practice, actual performance on modern computers is usually dominated by factors other than the speed of arithmetic operations and the analysis is a complicated subject, but the overall improvement from \( N^2 \) to \( N \log_2(N) \) remains. On the Figure (6. 8) you can see an original data of a signal in time domain (units in seconds [s]), and data after Fast Fourier transformation in frequency domain (units in hertz [Hz]).

**Fig 6: 8 an original data of a signal in time domain and FFT processing in frequency domain**

**Making use of FFT spectrum analysis for vibration analysis**

When a machine functions abnormally, for example due to imbalance or bearing damage, it will produce different vibrations that can be detected using FFT.
6.5 SAMPLING THEOREM

The sampling process, illustrated in Figure (6.9), where the analog signal $x(t)$ is periodically measured every $T$ seconds. Thus, time is discretized in units of the sampling interval $T$:

Time resolution is basically controlled by how fast the instrument collects the data (sampling frequency).

*Fig 6: 9  Ideal sampler.*

In contrast with the amplitude resolution, time resolution can be modified when configuring the data acquisition device up to some maximum limit. Because of this, the Machinery Diagnostic Engineer is required to have some knowledge on its effect on data quality. Let’s examine two
basic examples shown in Figure (6.10), one with low time resolution and the other with a high time resolution. In this example, the black line corresponds to the analog signal being sampled. For case a), by using a good sampling frequency, the reconstructed digital waveform (connecting the red dots) results in a closer representation of the original signal. However, when we look at case b), in which the sampling frequency is three times slower, the reconstructed signal becomes a poor representation of the analog one.

![Comparison between good time resolution (case a) and poor time resolution (case b).]

A more quantitative criterion is provided by the sampling theorem which states that for accurate representation of a signal $x(t)$ by its time samples $x(nT)$, two conditions must be met:

1. The signal $x(t)$ must be bandlimited, that is, its frequency spectrum must be limited to contain frequencies up to some maximum frequency, say $f_{\text{max}}$, and no frequencies beyond that. A typical bandlimited spectrum is shown in Equation (6.1).

2. The sampling rate $f_s$ must be chosen to be at least twice the maximum frequency $f_{\text{max}}$, that is,

$$f_s \geq 2f_{\text{max}}$$  \[6.1\]

or, in terms of the sampling time interval: $T \leq \frac{1}{2f_{\text{max}}}$

Prepared by: Dr. Essam Bahgat

essbahgat@gmail.com
The minimum sampling rate allowed by the sampling theorem, that is, \( f_s = 2f_{\max} \), is called the Nyquist rate. For arbitrary values of \( f_s \), the quantity \( f_s/2 \) is called the Nyquist frequency or folding frequency. It defines the endpoints of the Nyquist frequency interval:

\[
\left[-\frac{f_s}{2}, \frac{f_s}{2}\right] = \text{Nyquist Interval}
\]

### 6.4.1 Nyquist sampling theorem

The Nyquist Sampling Theorem states:

If a time-varying signal is periodically sampled at a rate of at least twice the frequency of the highest-frequency sinusoidal component contained within the signal, then the original time-varying signal can be exactly recovered from the periodic samples. The sampling frequency should be at least twice the highest frequency contained in the signal. Consider a sine wave. How often do we need to sample it in order to figure out its frequency.

If we sample with a frequency which corresponds to once per wavelength, we would think the sine wave is constant as shown in figure (6.11). If we sample 1.5 times the real frequency, we’d think the function is a sine with a longer wavelength (the red curve) see figure (6.12). If we sample at twice the real frequency, we get the right answer as shown in figure (6.13).
6.4.2 Aliased Signals

Aliasing is a signal processing term. Aliasing occurs when a system is measured at an insufficient sampling rate. Aliasing can occur if the sampling rate is less than the Nyquist sampling rate. In vibration applications, it is sometimes required to reconstruct the signal from its Fourier spectrum. Inverse DFT is used for this purpose and is particularly applicable in digital equalizers in vibration testing. Due to sampling in the frequency domain, the signal becomes distorted as shown in figure (6.14).

The aliasing error \( m - x(m\Delta T) \) is reduced by decreasing the sample period \( \Delta F \). It should be noted that no information regarding the signal for times greater than \( T = 1/\Delta F \) is obtained from the analysis. Aliasing can be explained practically an image processing in figure (6.15).
Example 6.1
Consider 1024 data points from a signal, sampled at 1 millisecond (ms) intervals.

Sample rate $f_s = 1/0.001$ samples/s $= 1000$ Hz $= 1$ kHz

Nyquist frequency $= 1000/2$ Hz $= 500$ Hz

Due to aliasing, approximately 20% of the spectrum (i.e., spectrum beyond 400 Hz) will be distorted. Here, one can use an anti-aliasing filter.

**How to avoid aliasing?**
Increasing sampling rate.
Use antialiasing filter.

6.5 SIGNAL FILTERING

Filters play a vital role in data acquisition systems to remove selected frequencies from an incoming signal and minimize artifacts (i.e. baseline wander, mains interference and noise).

**Important Terminology:**
- **Cut-Off Frequency** (fc): Also referred to as the corner frequency, this is the frequency or frequencies that define(s) the limits of the filter range(s). It is the desirable cut-off point for the filter.
- **Stop Band**: The range of frequencies that is filtered out.
- **Pass Band**: The range of frequencies which is let through and recorded.
- **Transition Band**: The range of frequencies between the passband and the stopband where the gain of the filter varies with frequency.

### 6.5.1 Low-Pass Filter

A low-pass filter allows signal frequencies below the low cut-off frequency to pass and stops frequencies above the cut-off frequency. It is commonly used to help reduce environmental noise and provide a smoother signal. A simple way to understand how a filter works is to plot signal frequency against signal gain see Figure (6.16).

**Prepared by: Dr. Essam Bahgat**
6.5.2 High-Pass Filters
A high-pass filter allows frequencies higher than the cut-off frequency to pass and removes any steady direct current (DC) component or slow fluctuations from the signal. Such filters are often used to stabilize the baseline of a signal (i.e. minimize baseline drift in an ECG signal). A useful comparison of the effects of a low pass filter in comparison to a high-pass filter is presented in the Figure (6.17).

**Band-Pass Filters:**
A band-pass filter may be used to pass a larger range of frequencies (i.e. 0 – 100 Hz EEG activity). Frequencies either side of this band are not passed.

**Band-Stop Filters:**

![Fig 6: 16 Effect of low-pass filter on signal gain](image1.png)

![Fig 6: 17 Effect of high-pass filter on signal gain](image2.png)
A band-stop filter blocks a certain range of frequencies and allows frequencies either side of this range to be passed (i.e. You may wish to block Beta [B1: 16 – 32 Hz] activity from an EEG recording but record all other frequencies between 0 – 15 Hz and 33 – 100 Hz).

**Cut-Off Frequency and Center Frequency**

For low-pass filters, the cut-off frequency is the frequency at which higher frequencies are blocked and lower frequencies are passed. For high-pass filters, the cut-off frequency is the frequency at which higher frequencies are passed and lower frequencies are blocked. Band-pass and band-stop filters have two cut-off frequencies (representing lower and upper limits).

For notch filters, the center frequency is the component frequency which is removed when the filter is applied to a waveform. For narrow band-pass filters, the center frequency is the frequency which is allowed through.

**Transition Width (Filter Sharpness or Roll-off )**

As mentioned earlier, all filters are imperfect. For example, a 200 Hz low-pass filter might leave frequencies up to 150 Hz untouched, reduce a 200 Hz signal to half its original amplitude and reduce the amplitude of higher frequencies more and more. However, it is not an “all-or-none process”.

### 6.6 SPECTRAL LEAKAGE

It is often said that the FFT implicitly assumes that the signal essentially repeats itself after the measured interval and hence the FFT assumes the signal to be continuous (conceptually, juxtapose the measured signal repetitively). This lead to glitches in the assumed signal (see Figure (6.18)). When the measurement time is purposefully made to be a non-integral multiple of the actual signal rate, these sharp discontinuities will spread out in the frequency domain leading to spectral leakage. This explanation for spectral leakage need to be carefully investigated.
Leakage results in the signal energy smearing out over a wide frequency range in the FFT when it should be in a narrow frequency range. Figure (6.19) illustrates the effect of leakage. The left-top graph shows a 10 Hz sine wave with amplitude 1.0 that is periodic in the time frame. The resulting FFT (bottom-left) shows a narrow peak at 10 Hz in the frequency axis with a height of 1.0 as expected. Note the dB scale is used to highlight the shape of the FFT at low levels. The right-top graph shows a sine wave that is not periodic in the time frame resulting in leakage in the FFT (bottom-right). The amplitude is less than the expected 1.0 value and the signal energy is more dispersed. The dispersed shape of the FFT makes it more difficult to identify the frequency content of the measured signal.

6.7 WINDOWING

Most digital signals are infinite, or sufficiently large that the dataset cannot be manipulated as a whole. Sufficiently large signals are also difficult to analyze statistically, because statistical calculations require all points to be available for analysis. In order to avoid these problems, engineers typically analyze small subsets of the total data, through a process called windowing. After the signal was digitized using an A/D converter, the next step in the process (before it can be subjected to the FFT algorithm) is called windowing. A ‘window’ must be applied to the data to minimize signal ‘leakage’ effects. Windowing is the equivalent of multiplying the signal sample by a window function of the same length. When an analog signal is captured, it is sampled with fixed time intervals. Sampling fixed time intervals can cause the actual waveform to get truncated at its start and end. The results obtained can vary with the location of the sample with respect to the waveform’s period. This results in discontinuities in the continuous waveform. Windowing fills the discontinuities in the data by forcing the sampled data to zero at
the beginning and at the end of the sampling period. Figure (6.20), shows the effects of windowing. Windows can be thought of as a way to fill in the discontinuities in the data by forcing the sampled data to zero at the beginning and end of the sampling period (or time window), thereby making the sampled period appear to be continuous. When the signal is not windowed and is discontinuous, a ‘leakage error’ occurs when the FFT algorithm is applied.

Fig 6: Window functions

There are many window functions as shown in Figure (6.21). Some used in vibration signal processing are:
1. Rectangular (basically no window)
2. Flat top
3. Hanning
4. Hamming
5. Kaiser Bessel
6. Blackman

Fig 6: Different types of windowing function in time domain.

Prepared by: Dr. Essam Bahgat

esson@gmail.com
To prevent the leakage error, *Hanning window* is used when measuring steady signals. The point is that the signal was suppressed at their ends to zero, thereby removing the signal discontinuities and creating a signal that is closer to the real periodic signal. The result of using Hanning window is shown in Fig. 6.22 below. In the case where the signal was periodic, the result got worse (instead of one spectral line there are three lines), but in case of non-periodic signal the result is substantially improved - there are just a few spectral lines.

6.8 Averaging

Another feature provided in analyzers/data collectors. The purpose is to obtain more repeatable results, and it also makes interpretation of complex and noisy signals significantly easier. There are various types of averaging:

- Linear averaging
- Peak hold
- Exponential
- Synchronous time averaging.
6.8.1 Linear averaging
Each FFT spectrum collected during a measurement is added to one another and then divided by the number of additions. This helps in obtaining repeatable data and tends to average out random noise. This is the most commonly used averaging technique. The spectra are typically averaged 2, 4, 8, 16 or 32 times, but any number could be used.

6.8.2 Synchronous time averaging
Synchronous time averaging is a technique using a speed trigger to a vibration data collector. This technique is very useful in isolating a single machine in areas where numerous machine trains are operating at near synchronous speeds. This technique is ideal for balancing and vibration analysis. Synchronous refers to data collector being triggered by the speed sensor. Time Averaging refers to a technique in the data collect, which takes many vibration readings over a time period and averages them together. This technique helps to isolate the vibration from the machine under test. In practice it is simple- place the vibration sensor on the machine to be diagnosed and use a remote optical sensor or stroboscope with a trigger output to accurately determine the exact speed of the machine (Figure(6.23). This external trigger will tell the vibration data collector when to collect vibration information and with the use of a high number of averages, will eliminate the “cross-talk” signals of near-by machines, providing precise vibration data on only the machine under test. You do not have to shutdown near-by machines. Using a stroboscope you do not even have to shutdown the machine to be tested!
6.9 WATERFALLS

The waterfall is a special display of the FFTs collected on the same position on a machine over a period of time. Each FFT is plotted one after the other giving an impression of a waterfall of FFTs. This kind of display makes it very easy to view variations in amplitudes of any frequency over the entire range. It considers another diagnostic tool for quick visual comparison of spectra is waterfall (cascade) diagrams. They are used in two ways:

- To compare spectra from steady operation during a longer period of time.
- To compare spectra during transient process - usually during a run-up or coast-down of the machine or during a significant change in load. Resonant frequencies of the machinery can be detected in this way (especially those of a rotor). Examples of waterfall plots are shown in Figure (6.24).

6.10 TRIGGERING

The key to implementing a successful triggering strategy is to establish a control signal that is capable of immediately starting data acquisition on one or more systems. The ability for a data acquisition system to start with a digital control signal is usually described as a Start Trigger Input. The ability for a data acquisition system to start other systems with a digital control signal...
is usually described a Start Trigger Output. In most cases, the specifications of these control signals are 5V digital pulses whose polarity and width will vary for different manufacturers.

6.11 BEATING

A phenomenon where a signal pulsates periodically because the signal comprises two signals of nearly the same frequency. The frequency of pulsation or beating is equal to the difference between the frequencies of the two signals. Beating can occur when there are identical machines operating at about the same speed, or when the frequency of the excitation force is close to the natural frequency.

If you add two waves of slightly different frequencies, the resulting amplitude will vary or oscillate at a rate that is the difference between the frequencies. That beat frequency will create a beat envelope around the original sine wave.
Since the frequencies of the two sounds are so close, you would hear a sound that is an average of the two. But you would also hear the modulation of the amplitude as a beat frequency, which is the difference between the initial frequencies:

$$f_b = |f_1 - f_2|$$

where

- $F_b$ is the beat frequency
- $f_1$ and $f_2$ are the two sound frequencies
- $|f_1 - f_2|$ is the absolute value or positive (+) value of the difference