Evaluation of Pavement Roughness and Vehicle Vibrations for Road Surface Profiling

Chinedum Anthony Onuorah

A thesis submitted to the University of Hertfordshire in partial fulfilment of the requirements for a degree of Doctor of Philosophy

November 2017

Abstract

The research explores aspects of road surface measurement and monitoring, targeting some of the main challenges in the field, including cost and portability of high-speed inertial profilers. These challenges are due to the complexities of modern profilers to integrate various sensors while using advanced algorithms and processes to analyse measured sensor data. Novel techniques were proposed to improve the accuracy of road surface longitudinal profiles using inertial profilers.

The thesis presents a Half-Wavelength Peak Matching (HWPM) model, designed for inertial profilers that integrate a laser displacement sensor and an accelerometer to evaluate surface irregularities. The model provides an alternative approach to drift correction in accelerometers, which is a major challenge when evaluating displacement from acceleration. The theory relies on using data from the laser displacement sensor to estimate a correction offset for the derived displacement.

The study also proposes an alternative technique to evaluating vibration velocity, which improves on computational factors when compared to commonly used methods. The aim is to explore a different dimension to road roughness evaluation, by investigating the effect of surface irregularities on vehicle vibration.

The measured samples show that the drift in the displacement calculated from the accelerometer increased as the vehicle speed at which the road measurement was taken increased. As such, the significance of the HWPM model is more apparent at higher vehicle speeds, where the results obtained show noticeable improvements to current techniques. All results and analysis carried out to validate the model are based on real-time data obtained from an inertial profiler that was designed and developed for the research. The profiler, which is designed for portability, scalability and accuracy, provides a Power Over Ethernet (POE) enabled solution to cope with the demand for high data transmission rates.

Acknowledgments

I would like to express my sincere gratitude and thanks to my supervisors for their guidance and continuous support of my PhD study. Johann Siau for being ever-present to bounce off ideas, and providing any resource I required to progress on the study. Professor Yichuang Sun for the attention to detail, and ensuring a high standard for my writing and presentation. Dr Sara Chaychian for all her help in reviewing my academic papers and thesis.

I must also thank my colleagues and friends at the smart systems research group Al Lalani (for the occasional distracting videos), Angus Hutton-Mckenzie, Chengqi Yang (for his unique personality), Emilio Mistretta, and Ivy Okike, for providing a very friendly environment during my time of study. Special mention goes to our vast discussion topics that spanned across various levels of appropriateness, relevance and intelligence.

I will also like to thank Chris Freeman at MATtest Southern for helping with the tests, and providing data samples, which was essential to this research.

Most importantly, I must thank my family for all their support and encouragement, because none of this would have been possible if it wasn't for them.

Contents

List of Figures

List of Tables

Abbreviations

AC	Alternating Current
ADC	Analog to Digital Converter
ASTM	American Society of Testing and Materials
AT	Attention
BMFLC	Band-Limited Multiple Fourier Linear Combiner
BSM	Basic Safety Message
BSS	Basic Service Set
C-ITS	Cooperative Intelligent Transport System
C/A code	Coarse Acquisition code
CALM	Communication Access in Land Mobile
CAN	Controller Area Network
CCW	Cooperative Collision Warning
CCWS	Cooperative Collision Warning Systems
CP	Common Peak
CRM	Customer Relationship Management
CWS	Collision Warning Systems
dA	Accelerometer Displacement
dB	Decibel
DC	Direct Current
DfT	Department of Transportation
DFT	Discrete Fourier Transform
DGP	Difference in Gradient Polarity
DHCP	Dynamic Host Configuration Protocol
dL	Laser Displacement
DMP	Digital Motion Processor
DSI	Dynamic State Information
DSM	Digital Single Market
DSRC	Dedicted Short Range Communication

EC	European Commission
FAA	Federal Aviation Administration
FFT	Fast Fourier Transform
GNSS	Global Navigation Satellite Systems
GPIO	General Purpose Input Output
GPS	Global Positioning System
GSM	Global System of Mobile communication
HWPM	Half-Wavelength Peak Matching
I/O	Input/Output
ICT	Information and Communications Technology
IDE	Integrated Development Environment
IEEE	Institute of Electrical and Electronics Engineers
IMU	Inertial Measurement Unit
IoT	Internet of Things
IP	Internet Protocol
IRI	International Roughtness Index
ISO	International Organization of Standardization
ITS	Intelligent Transport Systems
LAN	Local Area Network
LBRM	Laser Based Roughness Measurement
M2M	Machine-to-Machine
MAC	Media Access Control
MANET	Mobile Ad-hoc Network
MCU	Microcontroller Unit
MEMS	Microelectromechanical Systems
MT	Mean Threshold
NCHRP	National Cooperative Highway Research Program
OEM	Original Equipment Manufacturer
OSI	Open Systems Interconnection
P code	Precise code
PHY	Physical
PIARC	Permanent International Association of Road Congresses
POE	Power Over Ethernet
PRN	Pseudo Random Noise
PSI	Present Serviceability Index
RMS	Root Mean Square
RPM	Revolutions Per Minute

RSE	Rolling Straight Edge
RSS	Root Squared Sum
RSU	Roadside Unit
RTOS	Real Time Operating System
SAE	Society of Automotive Engineers
SDK	Software Development Kit
SPI	Serial Peripheral Interface
STA	Station
TC	Technical Committee
TCP	Transmission Control Protocol
TMR	Transparency Market Research
UART	Universal Asynchronous Receiver/Transmitter
UDP	User Datagram Protocol
UKDoT	UK Department of Transportation
UP	Uncommon Peak
USB	Universal Serial Bus
USDoT	US Department of Transportation
V2I	Vehicle-to-Infrastructure
V2V	Vehicle-to-Vehicle
V2X	Vehicle to Vehicle/Infrastructure
VANET	Vehicular Ad-hoc Network
VC	Vehicular Communication
WAVE	Wireless Access in Vehicular Environment
WFLC	Weighted-Frequency Fourier Linear Combiner

List of Publications

The publications listed below present the work done in the following chapters, and are either in review or already published.

- Onuorah C., Hutton-McKenzie A., Chaychian S., Sun Y. and Siau J. "Improving Displacement Measurement for Evaluating Longitudinal Road Profiles". In review for *IEEE Sensors Journal*. **2017**.

- Onuorah C., Chaychian S., Sun Y. and Sisau J. "Development of a Vibration Measurement Device based on a MEMS Accelerometer". In *Proceedings of the 3rd International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS 2017), Porto, Portugal,* **2017**.

CHAPTER 1

INTRODUCTION

1.1 Overview

This research focuses on road monitoring, advancing current techniques in evaluating road surface conditions and pavement analysis, with a view to enable smart vehicles to be aware of surface conditions along their path in real time, alerting drivers of potentially dangerous areas.

Road transport is vastly considered to be the most used mode of transport in the world, and this realisation stresses the need to ensure optimum road conditions. The drive to advance the road transport industry is universal because of its direct relationship to a country's economy, environment, transport safety etc. The UK's Department of Transport's (DfT) January 2016 report "*Transport infrastructure skills strategy: building sustainable skills*" [1] highlights that people are making twice as many journeys as they did in 1970. This increasing demand and challenges that comes with it has encouraged the government to invest £411 billion towards 564 road and rail projects and programmes, with a plan to modernise the country's transport infrastructure.

Road conditions are expected to meet certain design and structural requirements following standards established by internationally recognised organisations (for example ASTM International, Standards for Highways etc.). To build and maintain

road surfaces according to these standards, accurate measurement techniques are needed. Road quality assessment is critical to ensuring comfortable, efficient and safe transport, and advancement in technology has given rise to better and more efficient means of measuring and monitoring road conditions. Studies in [2] show that bad road conditions (i.e. damages and anomalies on road surfaces) has a negative effect on a vehicle's energy efficiency, causing increase in fuel consumption. Undesirable surfaces, bumps and/or potholes generates vibrations to the vehicle, causing a higher potential for wear and tear, and therefore increasing the cost of maintenance over time to the owner of the vehicle. Safety is also one of the major drives to road maintenance and monitoring, since good road conditions allows smoother driving experience and comfort to road users.

Legacy methods of evaluating road quality required hand held devices (like the rod and level, 3 meter rolling straight edge etc.) to be pushed, pulled, carried or placed along the road surface to record measurements. Most of these devices are still use because of the accuracy and consistency of measurement they provide, but their efficiency diminishes as the length of road to be measured increases due to the average time take to obtain the measurements. To efficiently maintain road conditions on a large scale, faster and accurate methods of monitoring/measuring road surfaces are necessary. This has given rise to the development of inertial profilers, which is mounted on a vehicle and driven along the road surface, to allow faster measurement. The profilers require inertial sensors to operate, and these sensors come with challenges that affect accuracy of the measured data taken when used in this application. These challenges are discussed in subsequent chapters and is one of the motivations for this research.

More advanced and expensive profilers integrate a combination of proximity (laser), visual (camera) and inertial sensors, running complex processing algorithms to evaluate the corresponding road surface. Due to the cost, size and processing requirements of these profilers, there is a need for alternative processing techniques to achieve smaller, portable and scalable devices.

1.2 Research Aims

The aims of this research are summarised below:

- To design a POE enabled hardware solution for a single point laser based road surface measurement system, allowing real time data streaming over a network.
- Investigate and propose new techniques to improve the accuracy of the laser based inertial profilers.

1.3 Contribution to Knowledge

The novel contributions to knowledge are summarised as follows:

- The first contribution proposes a novel Half-Wavelength Peak Matching (HWPM) model to improve the accuracy of the evaluated longitudinal road surface irregularity measured by inertial profilers. To verify this model an inertial profiler (referred to as Laser Based Roughness Measurement device in this literature) was designed and developed to acquire actual road measurement samples. In its simplest form, the HWPM model employs a peak matching principle to correct the drift present in the calculated displacement signal derived from accelerometers after integration.
- The second contribution proposes a novel approach for measuring a body's vibration velocity, with an aim to evaluate the effect of road roughness on vehicle vibration. Using an accelerometer to acquire the vibration samples, this approach evaluates the vibration velocity via the frequency domain of acceleration data. A Root Sum Squared (RSS) average with a fixed length is then applied to the calculated velocity frequency bins to consistently achieve accurate vibration measurements.

1.4 Summary

This thesis is organized as follows,

Chapter 1 concisely defines the problem domain, supporting the research drive to improve the current state of road pavement measurement and monitoring, and outlines the research objectives and contribution to knowledge.

Chapter 2 reviews the sensors, theory and existing methods of evaluating road roughness and vibration. It defines the profiling terminologies used throughout the thesis, describing the challenges faced by vehicles and road users as a side effect of bad road conditions. The chapter also explains common techniques used to improve measurement accuracy in vibration analysis, illustrating the importance of examining the frequency spectrum, while outlining applications and recent studies in the field.

Chapter 3 presents a detailed description of a Laser Based Roughness Measurement (LBRM) device, which was designed (for this research) with a focus on portability, accuracy and scalability of road surface monitoring. It starts out by outlining the requirements of the device in terms of power usage, measurement properties, and data

transmission, then proceeds to expand on the hardware features, comparing design decisions and stating any corresponding trade-offs. The drive for scalability encouraged a novel application of an existing technology in Power Over Ethernet (POE), to support sufficient power sourcing and fast data transmission rates from a single port. The accompanying software was thoroughly defined using flow charts (describing program flow), justifying the use of certain design styles and transmission protocols based on overall system performance. The device is operated via a custom Attention (AT) command set, over an established TCP connection. This allows remote control functionalities to power device, stream samples and upgrade device firmware over a network. All supported AT commands are listed in the chapter with a description of their respective functionalities, modes of operation, and expected response from the device on execution.

Chapter 4 introduces a novel correction technique called Half-Wavelength Peak Matching (HWPM), to improve the accuracy of longitudinal road profiles generated by high-speed profilers (like the LBRM device described in Chapter 4). The technique operates on a matching principle that estimates an error offset on a sensor, based on measurements from another. Details of the proposed correction technique is comprehensively described, including related mathematical models and program implementations in pseudo code. The chapter then discusses and compares the results of the HWPM model with commonly used correction methods. The analysis highlights the benefits and convenience of the HWPM technique, explaining the differences and improvements at different profiling conditions.

Chapter 5 proposes the theory of investigating road roughness via vibration analysis on the vehicle's unsprung mass. It describes a device developed to measure the vibration velocity of a body using a MEMS accelerometer. The operation of the device is based on a novel algorithm that consistently achieves accurate and precise vibration measurements comparable to piezoelectric meters. The purpose of this development is to realise cost effective and flexible means of vibration analysis, which is fueled by the increasing availability and reduced cost of MEMS accelerometers. The idea is to analyse a road surface based on its longitudinal profile and the vibration induced on the vehicle, to better evaluate the road surface condition.

Chapter 6 summarises, discussing possible areas for improvements and concludes.

CHAPTER 2

ROAD PROFILING AND VIBRATION ANALYSIS

2.1 Overview

The ability to predict and inform drivers of potentially hazardous situations while driving is the basis for intelligent transport safety applications. The purpose is to allow drivers more time to take necessary steps to react to these risks.

Bad road condition is a hazard, and one of the major factors that pose a threat to drivers. Drivers unaware of potholes or objects on the road can only rely on visual identification, which in optimal conditions should not pose a problem. Factors like driver distraction, weather conditions, or visual unawareness are the main concerns, and a vehicle's ability to detect and warn drivers of upcoming road defects autonomously can play a significant role in road safety.

Generally, the basic idea is to monitor vehicle dynamics using a sensor, to identify areas of sudden unusual changes. Considering vehicle are predictable to an extent in their movements, since they are expected to follow road paths, in normal conditions, the vehicle dynamics can also be predictable. With this theory, a sudden change in its manoeuvre is likely to be the effect of an unexpected event, some examples of these could be a result of a collision, traffic calming measures, or a neighbouring driver trying

to evade a hazardous situation. A repeated occurrence of such events by multiple vehicles at the same location suggests a point of interest.

This chapter is split into two main sections. Section 3.2 reviews the theory of road roughness and profiling, describing usual measurement techniques and terminologies. Section 3.3 reviews the theory of machine vibration, evaluating typical challenges involved in its measurement, and analysing its application in the transport environment.

2.2 Road Roughness and Profiling

The profile of a road is a continuous line along its lateral or longitudinal axis plotted against its height as shown in Figure 2-1. A lateral profile reveals the cross-sectional shape of the road, making it easy to identify road defects like rutting. A longitudinal profile is taken along the vehicle path, and shows the road texture and roughness.

Profiling is a vital aspect of pavement engineering, enabling monitoring of road networks, diagnosing failures and determining suitable solutions, evaluating quality of new or repaired roads, and aiding research.

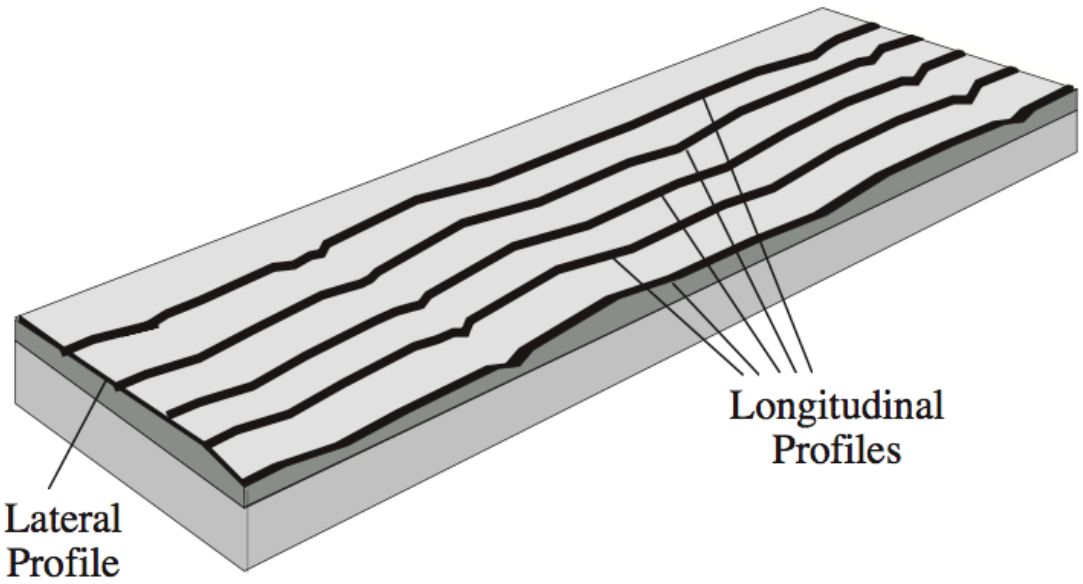

Figure 2-1 Longitudinal and Lateral Profile [3]

2.2.1 Profilers

Profilers are instruments used to generate a plot representing the profile of the road. Equipment like rod and level, 3 metre rolling straight edge (RSE), and dipstick are common methods of measuring road surfaces.

2.2.1.1 Rod and level

The rod and level is a common surveying tool used to measure elevation between two points on a pavement. The rod is essentially a rule (metric or imperial) which indicates the height of the measurement, while the level is an optical instrument (mounted on a tripod) used to verify a point in same horizontal plane. Measurement using this equipment requires the rod to be placed line of sight to the level, and the elevation is derived as the difference between the level's height from the ground and measurement value taken from the rod.

The requirements for road surface profiling are much stricter compared to surveying, for example, elevation measurements must be taken at intervals of a foot or less [3], [4]. Performing a road roughness analysis with a rod and level will ultimately be time consuming and impractical, hence is not ideal for the application.

2.2.1.2 Dipstick

The dipstick (which is patented and developed by the Face companies) is faster when compared to the rod and level for measuring road roughness. It consists of an on-board processor to automatically record and perform calculations required to generate a profile. The device contains an inclinometer to measure the slope between the supports, which is approximately 300mm apart [3]. To measure the road profile, the device needs to be positioned in parallel to a predefined line (for consistency) along the road. When the on-board computer detects stability, it automatically records the height difference between the support, and signals to the operator. At this point, the equipment is rotated 180° pivoting on the leading support leg, before another measurement is taken. This becomes more time consuming as the length of the road to be profiled increases, hence the need for even faster profiling techniques.

2.2.1.3 3m rolling straight edge (RSE)

The RSE typically consists of a 3m frame supported by wheels on both ends of the frame. Midway through the RSE is a wheel connected to a dial, with the capability to be displaced freely on the vertical axis. During operation, this center wheel measures its height displacement in relation to the 3m frame's reference, which is reflected on

the dial and is recorded by the operator. The difference between the RSE and previously mentioned equipment is fact that it can be rolled (via its wheels) along a road path, making measurement easier.

2.2.1.4 Inertial profiler

Inertial profilers in its simplest form consists of an accelerometer and a non-contact displacement sensor mounted on a vehicle, with an on-board computer to carry out data processing algorithms. The accelerometer keeps track of the vehicle's vertical displacement, while the displacement sensor (typically laser transducers) measures the precise height between the profiler's mount position and the road surface. Inertial profilers are high-speed profilers, and do not function accurately at low speeds. This is because at low speeds, the accelerometer records lower frequency displacement with respect to the road roughness, which is not ideal since accelerometers are prone to increased error rates at lower frequencies.

2.2.2 Roughness and texture

Road roughness is a term used to indicate road quality, and this is essential to both vehicle and road maintenance. According to the American Society of Testing and Materials (ASTM), road roughness is "the deviations of a pavement surface from a true planar surface with characteristic dimensions that affect vehicle dynamics, ride quality, dynamic loads, and drainage, for example, longitudinal profile, transverse profile, and cross slope" [5].

As a road user, roughness is felt as undesirable vibrations from the vehicle, which has a negative effect on the ride comfort and vehicle condition. Essentially, it is the vertical imbalance of a road whether laterally or longitudinally, as opposed to a smooth surface. The roughness property of a road can change overtime due to several factors, which can be due to changes in weather, wear and tear caused by constant road use, or simply bad road design and construction. Increased road deterioration demands more expensive repairs to the road, reduced driver safety, and a rise in the vehicle's operational costs (including fuel consumption and maintenance). Generally, road roughness monitoring is an expensive process, but it is an essential and critical aspect of road transportation.

Road texture is a complementary part of roughness as it defines the deviations from a smooth surface affecting vehicle/tyre interaction.

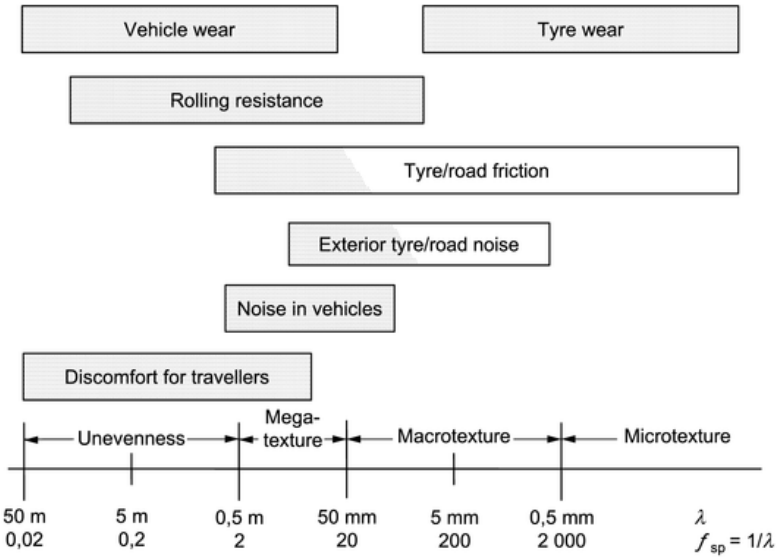

Figure 2-2 Road texture classifications [6]

Figure 2-2 show the different classifications of road textures, illustrating their respective effect on vehicles and road users. The image is taken from the ISO 13473-5 standard that characterises pavement textures by use of surface profiles, where λ and f_{sp} represent the texture wavelength and spatial frequency (cycles/m) respectively. Lighter and darker shades in the image indicate favourable and unfavourable (respectively) effect of texture over the stated range [6]. Figure 2-3 show a visual illustration of the different textures in the real world.

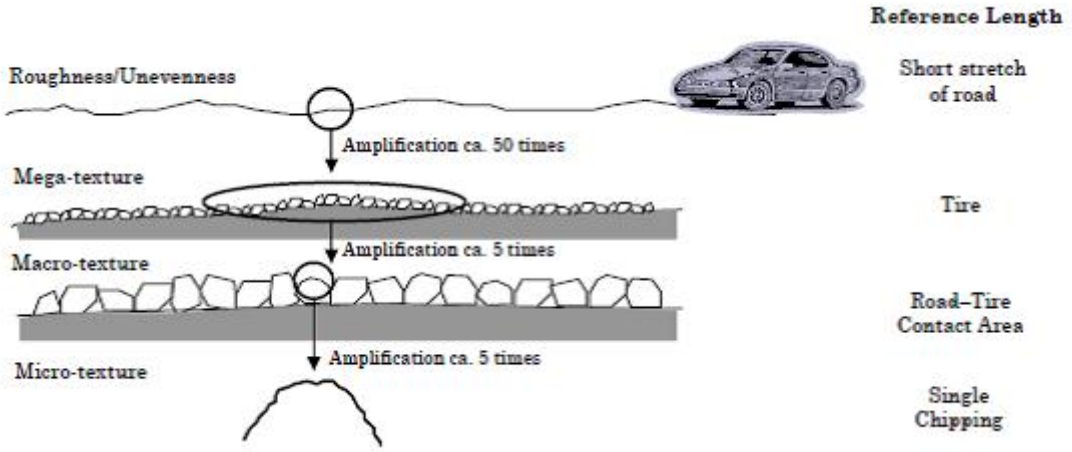

Figure 2-3 Road texture visualization [7]

Pavement texture deviations are grouped into three scales, defined by its component wavelength and peak-to-peak amplitude. As established by the Permanent International Association of Road Congresses (PIARC) [7][8], these ranges are:

2.2.2.1 Micro-texture

This is produced by the surface properties of the individual chippings or aggregate particles that make up the paving material (i.e. asphalt or concrete), with texture wavelength up to 0.5mm, and peak to peak amplitudes varying between 0.001mm to 0.5mm.

2.2.2.2 Macro-texture

This is produced by the mixing properties and surface finishing technique used on the paved surface. Their wavelengths range from 0.63mm to 50mm, with peak-to-peak amplitudes varying between 0.1mm to 20mm.

2.2.2.3 Mega-texture

This is usually a result of defects and distress on the pavement surface such as potholes or "waviness". Their wavelengths range from 63mm to 500mm, with peak-to-peak amplitudes varying between 0.1mm to 50mm.
Textures with wavelengths longer than mega-texture (500mm) are referred to as unevenness or roughness [7].
There are several factors that affect surface texture, which is beyond the scope of this thesis but can be found in [7]. In some cases (micro and macro textures) they are induced (within limits) to encourage friction between tire and road, which is essential for vehicle balance and control.

2.2.2.4 Roughness Index

The change in longitudinal profile over time is an indicator of road smoothness, and is used to evaluate the roughness of a road. Roughness index is a computed value used to identify the quality of a pavement for a specified distance. When road conditions are assessed with longitudinal profile measurements, the roughness index summarises and reduces the thousands of elevation values into a single value to represent particular road sections. The International Roughness Index (IRI) is the most widely used roughness index as a general pavement condition indicator, and its analysis is intended to achieve consistent profile index irrespective of the type of profiler being used. IRI was developed in research sponsored by the National Cooperative Highway Research Program (NCHRP) and the World Bank [9]. The index measures how many millimetres

per meter (mm/m) jumped by a vehicle along a pavement, where a lower IRI value indicates a smoother more level surface. It is essentially the accumulated suspension displacement divided by the travelled distance. The process of calculating the IRI of a pavement's longitudinal profile is described in the ASTM E1926 standard [10].

Alternative roughness indices created by highway agencies, standard organisations, and individuals do exists for evaluating the general condition of a road and its ride quality. Although they all produce a roughness index, not all devices do this by direct measurement of the longitudinal elevation profile. Some of these are typically cast onto a different type of scale, and usually used to evaluate different road features than the IRI. Many are also a transformation of the IRI, scaled with a conversion equation, for example the Present Serviceability Index (PSI), which is a scale from 0 (bad) to 5 (perfect) used rate a section of a highway based on visual observations [9]. Ultimately, regardless of which index calculated from a longitudinal profile, the accuracy of the index is only as good as the measured profile.

2.2.3 Related Research

Researchers in [11], [12] develop models to estimate road surface condition based on data obtained from smartphone sensors. Their goal is to explore easier and low cost methods in road monitoring, and with the increasing popularity and availability of smartphones, considering the presence of the various sensors bundled in them, they provide a viable option in this application. Their work show a linear relationship between the road surface roughness and the vehicle's vertical acceleration, and the results provided in [11] show that the vertical acceleration and evaluated roughness index is dependent on the speed of the vehicle. Other methods like [13] utilise a two dimensional laser displacement sensor to evaluate a three-dimensional road roughness. In [14], a combination of vision and laser sensors was used, proposing a laser line recognition method that only depends on grey value to determine the roughness, employing an anisotropic diffusion PM filter to smooth the pavement texture. In [15], the researchers propose a real time road profile estimation technique for the adjusting vehicle dynamics using conventional sensors. With this approach, an adaptive observer estimates the dynamic road signal, and then a Fourier analysis is performed on the signal to accurately determine the road roughness condition even with a varying vehicle velocity. [16] proposes an automated multi-sensor data acquisition system designed to inspect road pavement condition. The system consists of a laser sensor, camera, and proximity sensors, correlated by geolocation and distance travelled. Methods like [17]–

[19] combine image processing with inertia and laser sensors for road defect measurements.

One of the challenges with these systems is the ability to accurately calculate the vertical displacement from an accelerometer. This is because the double integration process of converting acceleration to displacement amplifies any slight error measured by the accelerometer. The presence of any low frequency or DC signals causes the evaluated displacement waveform to drift away from the expected result. This drift is due to the integration process of converting the acceleration samples to displacement, and the most common method of reducing this error is by passing the derived displacement through a high-pass filter, to eliminate low frequency signals present. The concept of utilizing Inertial Measurement Unit (IMU) sensors (which typically consists of an accelerometer and gyroscope) in motion applications to detect the movement of a body is common practice. Do et al proposes an inverted pendulum model [20] that evaluates the vertical displacement derived from an accelerometer (mounted at the upper torso), to detect a person's step, and estimating the stride length for infrastructure-less localization. The model converts that acceleration on the vertical axis to displacement by integration, and then passes the resulting displacement through a high pass filter with a cut-off frequency of 0.3Hz to remove the drift effect. Similarly, the researchers in [21], [22] integrate IMU sensors with Global Positioning System (GPS) signals to improve localization accuracy in areas where GPS signals are unavailable. Aside from integrating and high pass filtering the acceleration data to derive displacement, [22] employs a gyroscope and pressure sensor using a cascaded two-step (for orientation and position) Kalman filter to minimise the effect of drift in estimating the vertical position of the device.

The works carried out by the researchers in [23]–[29] confirm the negative effect of drift in applications that use IMU sensors to evaluate displacement, and various solutions were proposed in their respective studies to minimise this drift. For applications that require precise displacement measurements like in [22], [24], [29] where an IMU sensor is used to detect an object's physical positon in space, by computing the displacement from its initial location, Kalman filters are typically used. This technique works by estimating an error offset on the evaluated displacement, based on a mathematical model that integrates the measurements obtained from other related sensors. Results from [29] claim low error rates of approximately 5% (for applications with movements constrained along an axis) and 9% (for applications with free movement in space). Correction models in [23], [25] are targeted to applications where the motion of the object is periodic. The theory in [25] relies on pre-existing knowledge of the frequency

(or frequency bands) of motion, since the movement is expected to be periodic. The method proposed uses a Weighted-Frequency Fourier Linear Combiner (WFLC) to detect a single dominant frequency, and Band-Limited Multiple Fourier Linear Combiner (BMFLC) to detect multiple frequencies in a band. The estimated signal is then modelled using a series of sine and cosine components based on the determined frequencies.

2.3 Vibration analysis in a vehicular environment

Vibration is the movement of a body about its position of rest due to an applied external force, and can be classified as periodic (with repetitive characteristics) or random. According to the ISO 2041:2009 standard, vibration is defined as mechanical oscillations about an equilibrium point [30]. This oscillation may be forced or resonant, occurring at the natural frequency of the vibrating object, where an electric motor rotating at constant speed, and a vibrating guitar string are examples of forced and resonant vibrations respectively [31]. In automotive applications, the vibration oscillation is complex, consisting of multiple frequencies and amplitudes, because the machines are made up of other components, vibrating at their respective frequency and intensity.

2.3.1 Theory (units and relationships)

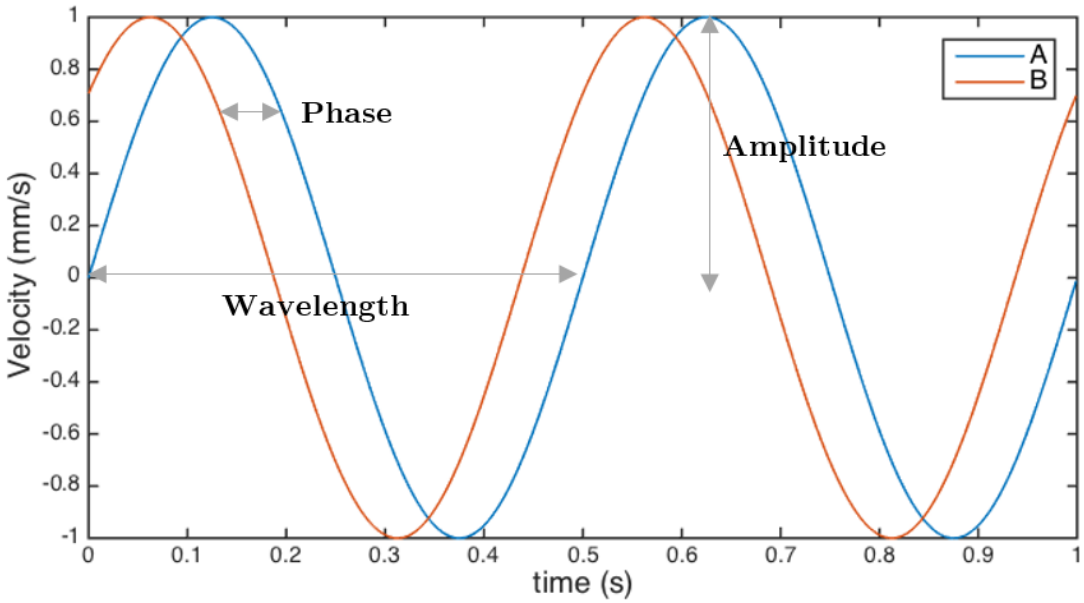

Figure 2-4 Properties of a signal

There are key terms used to describe and compare vibration waveforms as shown in Figure 2-4, and these include:

- **Wavelength**: the time occupied by a cycle at any given instance.

- **Frequency**: this is relevant in a periodic waveform, and it defines the number of times a cycle (repetitive characteristic) occurs in the signal for a period of one second.

- **Amplitude**: the amplitude of a wave gives a relative indication of the amount of energy the wave transmits [32]. For vibration signals, this is a unit or derivative of distance to indicate how far away the mass is displaced from its reference point.

- **Phase**: this is relevant when comparing signals to show the difference in time domain shift. Considering the signals A and B in Figure 2-4, the waveforms show identical characteristics in terms of amplitude and frequency, but there is an offset between them in the x (time) axis, this offset is referred to as the phase difference.

2.3.2 Measurement and instrumentation

Vibration is expressed in displacement or its derivatives i.e. velocity (rate of change of displacement) and acceleration (rate of change of velocity), which is used to represent the severity of vibration. The decision on selecting an acceptable unit of measurement is dependent on the application and nature of the vibration. Essentially, since acceleration is a derivative of velocity, and velocity, a derivative of displacement, the frequency component of a signal determines the choice of measurement unit. Low frequency signals ($<$ 10Hz) will appear more pronounced on a displacement scale compared to velocity or acceleration (where the equivalent vibration will be very little), hence such vibration signals are best analysed in displacement. On the other hand, high frequency signals ($>$ 1000Hz) are better represented in acceleration as they reflect more significant values than velocity and displacement, while velocity is preferred for vibrations between 10Hz and 1000Hz.

Transducers are used to measure vibration, they are devices that convert one form of energy to another, in this case, converting vibration to electrical voltages or current. Commonly used transducers include proximity probes (for displacement measurement), velocity pickup (to measure velocity), and accelerometers (to measure acceleration).

2.3.3 Measurement terminologies

2.3.3.1 Peak to Peak (pk-pk)

This is the entire distance travelled in one vibration cycle from maximum to minimum displacement.

2.3.3.2 Peak (pk)

The peak displacement is half the pk-pk, and represents the displacement from reference point.

2.3.3.3 Root Mean Squared (RMS)

RMS as the name suggests is the square root of the mean squared for all vibration measurement taken for a specified period. For sinusoidal signals, this is equivalent to $pk/\sqrt{2}$.

2.3.3.4 Crest factor

This is the ration of the peak value of a signal to its RMS value.

According to [32], RMS tends to provide the energy content in the vibration signal, whereas the peak correlates better with the intensity of vibration, and for machine monitoring applications, a higher RMS is generally more damaging than a similar peak magnitude.

2.3.4 Spectral analysis, windowing and representation

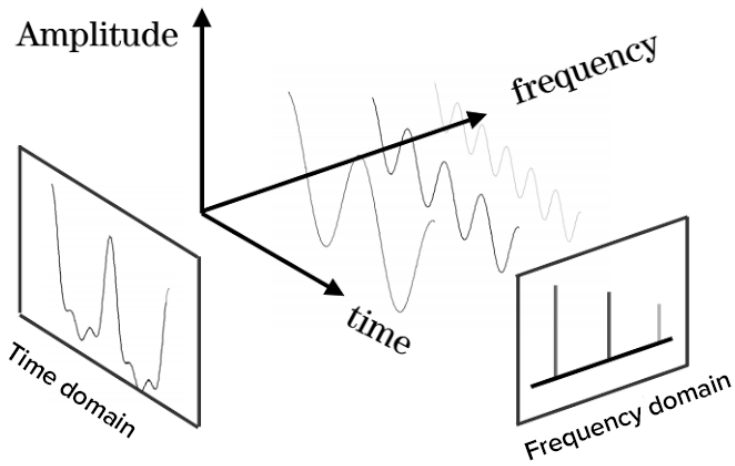

Figure 2-5 Time and Frequency domain

Complex vibration signals are made up of several signals with different frequencies and magnitudes. Measurement samples obtained from vibration transducers are a time representation of the vibration, which essentially is an accumulation or sum of all the sensed vibration at the measurement point. For applications that don't require identifying a specific vibration(s) this data is sufficient, otherwise there is a need to breakdown this complex waveform into its individual components, to determine all the signals (with their respective frequencies and amplitude) that make up the complex waveform as shown in Figure 2-5.

Spectral analysis is the evaluation of a vibration signal into a frequency spectrum, and a Discrete Fourier Transform (DFT) is a calculation on the time domain used to transform the signal to the frequency domain. In the frequency domain, each sine and cosine wave making up the signal is plotted according to their frequencies with their respective magnitudes, making it easier to visualize the various components of the signal. An important parameter that needs to be established to correctly translate the signal to its frequency domain is the sampling rate. The sampling rate is the time take to acquire a new measurement sample from the transducer. For accurate spectral analysis, this sampling rate needs to be consistent for each sample in the analysed waveform. A higher sampling rate is required for better signal representation, hence a better (more accurate) frequency domain representation after DFT.

DFT is a slow algorithm with an execution time directly proportional to the square of the number of samples to be processed. In big O notation [33] (describing the performance of an algorithm), DFT is expressed as $O(n^2)$, where '**n**' represents the data size (number of samples in this case). For example, with a sample length of 20, the DFT algorithm would require 400 (20^2) operations to convert the data to its frequency domain, while 500 samples would require 250000 operations. In modern computers with an abundance of memory and processing power, these sorts of operations would be executed in a fraction of a second, but considering the performance of the algorithm, if higher sampling lengths are required (e.g. > 10000), there will be a noticeable delay in execution. For MCUs with limited memory and processing capabilities performing a DFT algorithm would be impractical.

Fast Fourier Transform (FFT) are a collection of algorithms that achieve faster DFT computations. The notable advantage to FFT is in their computation time, where the number of operations required is O(n*log(n)). The more popular FFT algorithms like the Cooley-Tukey algorithm [34] require the sample size to be in the power of 2 (i.e. 16, 256, 1024 etc.) because of the nature of their computation, which rely on a divide

and conquer principle. In applications that use MCUs, FFT is a more practical option when Fourier transformation is required, since their computational performance is logarithmically proportional to the sample length, i.e. the number of operations that would be required for a 20/500 samples is approximately 26/1350 respectively.

The decision on selecting an appropriate sampling rate is dependent on the maximum signal frequency expected during measurement, and this is based on the Nyquist sampling theorem, stating that "if we are not to lose any information contained in a sampled signal, we must sample at a frequency rate of at least twice the highest frequency component of interest" [32]. The effect not following this theorem generates a false lower frequency wave because of under-sampling, and this is known as aliasing.

FFT requires a finite number of samples to translate the analog (time) data, and the choice of this number lies on the sampling rate and expected resolution on the frequency domain representation. The resolution is the frequency step difference between each discrete frequency on the x-axis of the frequency domain plot. This finite nature of evaluation brings about a concept known as windowing, and is required to minimise a phenomenon called spectral leakage.

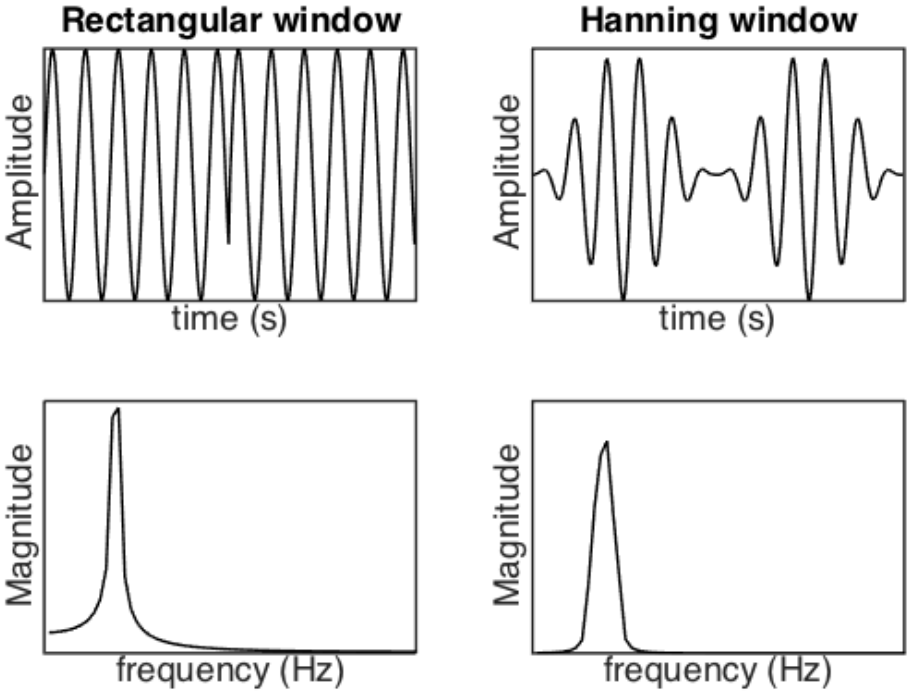

Figure 2-6 Effect of Hanning window on discontinuity

In theory, windowing is achieved by multiplying a signal by a window function of the same length. Analog signals acquired during measurement are only a section of the

infinite continuous signal, hence are truncated at the start and end based on the fixed time interval. Since it is not feasible to process an infinite analog signal, FFT assumes the actual continuous signal is an infinite recurrence of the processed samples. Except for a sample block that converges to zero on each end, there will be discontinuity of the signal as shown in Figure 2-6.

The purpose of windowing is to eliminate this discontinuity, by bringing the start and ends of the samples to zero based on a mathematical model, making the resultant waveform appear continuous. A rectangular window is the default window used in FFT, as it simply is the use of a finite series of consecutive samples extracted from the continuous signal as shown in Figure 2-6.

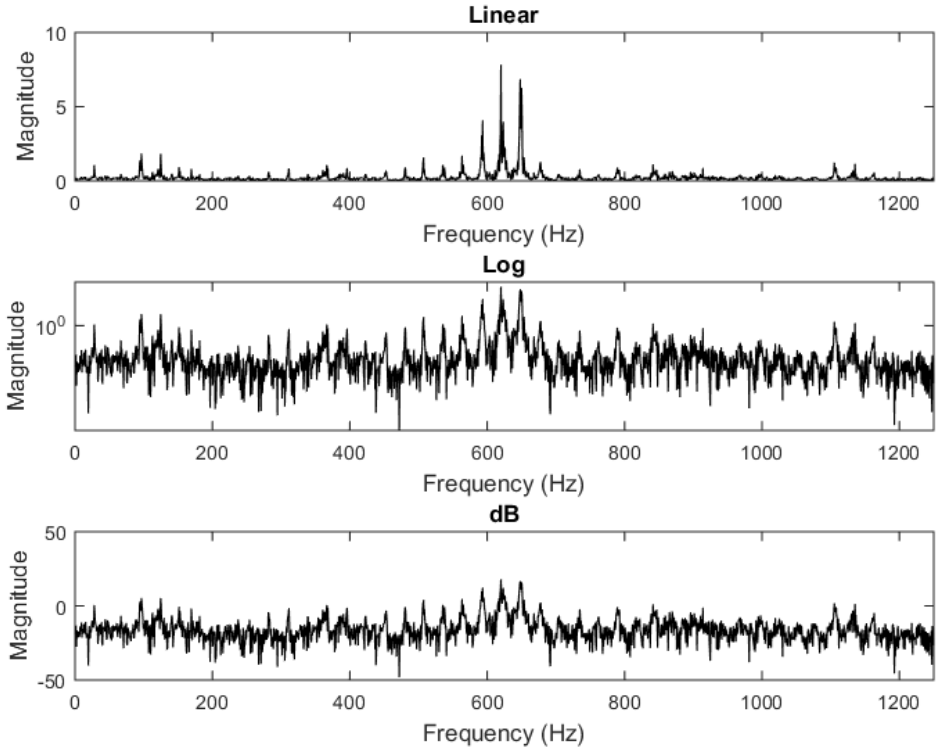

Figure 2-7 Difference between linear, log and dB scale

The rectangular window suffers from discontinuity as the window function is basically a series of 1, making no change to the analog signal, hence can be assumed as having no window. Some other examples of window techniques include Hanning (Figure 2-6), Flat top, Hamming, Blackman etc. Windows have a broadening effect to the signal peaks in the frequency domain, spanning across multiple frequencies. This means that in cases where two signals of very close frequencies are present in a signal, it becomes difficult to identify both signals, as they appear as one broader peak. The wideness of

the peak differs from window to window, and the selection of a window is based on the application. For example, in applications where the magnitude of the signal is more important than the frequency, a flat top window is the best solution, but their disadvantage is in the frequency inaccuracy. [35] describes some of the more popular window techniques, highlighting their pros and cons, and applications better suited for each window.

The magnitude scale is an important aspect of understanding characteristics of vibration signals, and this can be either linear, log or Decibels (dB) as shown in Figure 2-7. The linear scale is the simplest, and displays a true picture of the measured vibration. In most applications, this is sufficient if all significant magnitudes on the spectrum are of similar level. Log and dB scales are better suited for monitoring vibration spectrums where changes in small vibration magnitudes are very visible, allowing easier identification.

2.3.5 Applications

Several applications generally rely on vibration analysis in machine operation, typically in relation to diagnostics and fault detection. In the transport industry, more recent studies are being carried out to investigate other areas that can benefit from understanding vibration signatures. These studies are based on the vibrations sensed on the vehicles (due to road irregularities), or the pavements (caused by the vehicles in motion), and the focus on this chapter is primarily on applications that relate to the automotive and transport environment.

2.3.5.1 Monitoring and diagnostics

The vibration of a machine while in operation gives an insight to its condition, and this is usually the first indicator of mechanical problem such as loose or defective bearings, unbalance, misalignment, bent shafts etc. A vibration analysis allows an operator to evaluate the condition of the machine, diagnosing mechanical problems, and preventing any potential faults by monitoring vibration levels. Every complex machinery is allowed a certain magnitude of vibration to suggest acceptable operating conditions (described in the respective machine standard). Vibrations exceeding these magnitudes are more likely to cause an increasingly negative effect on performance, and if unattended, causes a faster deterioration rate, shortening its lifespan.

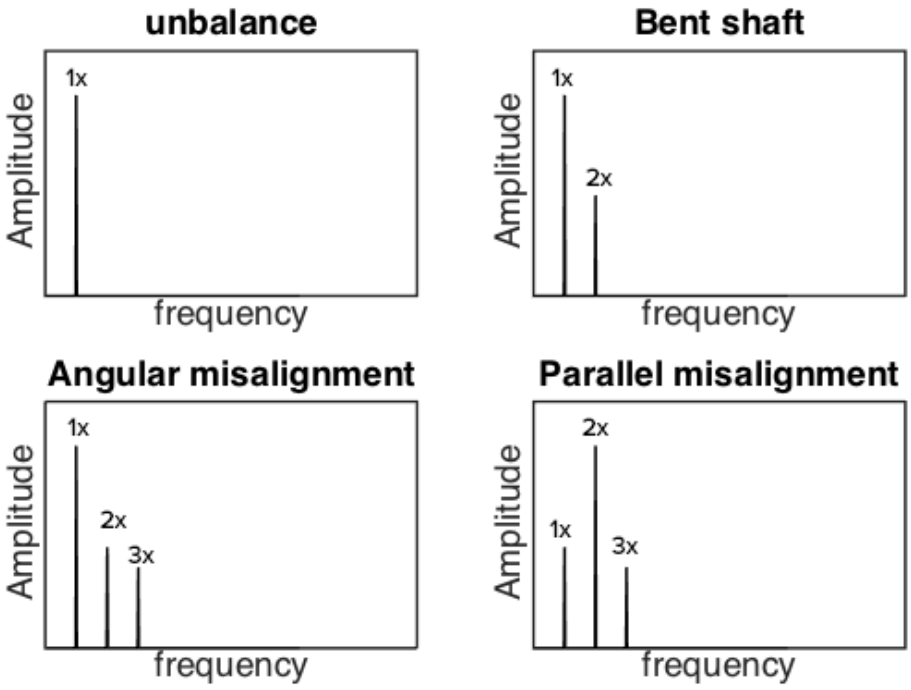

Figure 2-8 Machine defects in the frequency domain

The indication of a fault is reflected by excessive amplitude in both time and frequency domain of the vibration signal. The time domain is used to visualise the total vibration sensed at the point of measurement, which is made up of signals from various vibration sources. At best, the time domain signal indicates a fault, the same way an operator can hear a change in the vibration sound when a part comes undone, but this does nothing to identify the source of the fault, hence why the frequency domain plays a major role in machine diagnostics. Essentially, the time domain generally reveals that there is a problem, while the spectrum lets an operator know exactly what and where the problem is. Figure 2-8 show vibration signals in the frequency domain, representing different machine defects, and how the vibration differs from normal operation [32].

2.3.5.2 Unbalance

ISO defines unbalance as a condition which exists in a rotor when vibration force or motion is imparted to its bearings as a result of centrifugal forces [36]. This is a common defect in machinery, and reflects an increased magnitude at 1x rpm vibration frequency as shown in the plot.

2.3.5.3 Bent shaft

In the case of a bent shaft, the FFT will normally reflect peaks on both 1x and 2x of the vibration RPM. For instances where 1x magnitude is dominant, this indicates a

bend in the shaft closer to the center, while a dominant magnitude at 2x indicates a bend near the end of the shaft. With bent shafts, there is also a phase difference of 180^{o} between the axial and radial direction.

2.3.5.4 Misalignment

Misalignment can either be angular or parallel, in angular, the centerline of the two shafts meet at an angle, while in parallel, the centerline of the two machines is parallel to each other with an offset. In misalignment, the vibration occurs on either the axial or radial axis for angular and parallel respectively, typically having high magnitudes on 1x, 2x or both vibration frequencies, and in severe cases, higher amplitude peaks can be noticed at higher harmonics between 3x to 8x.
More detailed information on vibration analysis and techniques for detecting various machine defects can be found in [32], [37]–[39].
The ISO 13373, 10816, and 2954 are standards that describe the machine vibration diagnostics and condition monitoring techniques, evaluation of machine vibration, and requirements for measurement instruments respectively [40]–[42].

2.3.5.5 Localization

Vehicular localization is saturated with satellite based positioning systems like GPS and GLONASS, but their weakness in shadowing and multipath environment mean that they depend on other forms of localization for assistance. [43]–[45] study the idea of enabling localization based on vehicular vibrations cause by road surface roughness. The basic theory in [43] works by comparing live vehicle vibrations with pre-measured vibrations which have been stored in a database, with an error of about 2.5 meters for best case scenario. [44] uses this technique in an extended Kalman filter model to improve vehicle tracking based on its speed.

2.3.5.6 Speed and load detection

Speed detection is another application that can benefit from vibration analysis in a vehicular environment, [46] proposes a method that estimates the velocity of a vehicle by monitoring a frequency component of an accelerometer signal which is proportional to the vehicle's speed. Other concepts like [47], [48] determine the speed of the vehicle by the vibration induced on the road surface. For this method, the vibration analysis is done on the road surface, translating measure vibrations (as vehicle move across) to speed, providing an alternative to road speed monitoring systems like speed cameras.

For load detection, [49] aims to detect the presence of a person in a vehicle's rear seat, distinguishing between heavy objects and humans. The study is based on a piezoelectric

film designed to detect mechanical vibrations, focusing on extracting biological signatures from the acquire vibration signals. [50] investigates vehicle dynamics by post processing video information to extract vibration information, estimating the vehicle load in a more general basis.

2.4 Sensors and Filters

2.4.1 Accelerometer

An accelerometer measures the rate of change in velocity (acceleration) of the unit. Three popular techniques employed by industries in designing accelerometers are piezo electric, piezo resistive and capacitive methods. Although these concepts differ physically on their mode of operation, they all rely on the same principle of detecting small changes in mass movements, and translating the movements into an electrical signal [51], [52].

The piezo electric method works on the principle of piezo electric effect, which, in simple terms means electricity from pressure. These materials have properties that enable them generate an electrical charge from an applied mechanical stress. Piezo electric effect also works in reverse, that is, applying an electric charge on the material causes it to produce a pulse. As the accelerometer is moved, the weight exerted on the piezo electric material is varied, hence generating an electric charge that is a factor of the stress applied on the material, and then producing a voltage representing the acceleration. This design is appropriate when detecting shock and vibrations, as they are more suited for applications that sense high frequency movements, but they are subject to high noise.

The piezo resistive method uses piezo resistors, and its operation shares some similarities to that of piezo electric, in the sense that the electrical characteristics of the material changes regarding an applied mechanical stress. For piezo resistors, the effect of stress is on the electrical resistance of the material (piezo resistive effect). These accelerometers are relatively more expensive to manufacture (compared to piezo electric), and generally have low sensitivity, making them better suited for shock and pressure measurements.

The capacitive method is newer and more popular, it is commonly used in mainstream consumer accelerometers because they are cheaper to produce, and the manufacturing process is more reliable. They are made of silicon wafer, which is a thin slice of

semiconductor material used in fabricating integrated circuits. These accelerometers work by measuring capacitive changes in response to an applied mechanical stress. They have high sensitivities, very good temperature stability, and are well suited for applications measuring motion and steady state acceleration. Capacitive accelerometers are generally manufactured as Microelectromechanical Systems (MEMS), a technology that incorporates miniature mechanical (sensors, actuators, levers etc.) and electrical (resistors, capacitors etc.) components embedded in tiny semiconductor chips, to report changes in the physical properties of their environment.

The use of acceleration in this research is to derive displacement (to evaluate longitudinal road profiles) and velocity (to evaluate vibration). Following the equations of motion (2-1) and (2-2), displacement and velocity can be derived from acceleration by integration. Velocity is calculated by integrating acceleration, while displacement is obtained by integrating the velocity (i.e. a double integration on the acceleration). The major challenge in this process is overcoming the drift error that is imposed on the data over time, because small errors in the acceleration data will be magnified in the derived velocity and displacement calculations because of integration. Since velocity and displacement are a factor of their previous values in time, over successive periods, the calculated velocity and displacement begin to drift away from their actual values. Common practice implements high pass filters to overcome the drift error after integration, other advanced solutions involves the use of Kalman filters (discussed in section 2.4.5), which requires additional sensors.

$$v = u + at \tag{2-1}$$

$$s = ut + \frac{1}{2}at^2 \tag{2-2}$$

Where $\boldsymbol{v}$ is final velocity, $\boldsymbol{u}$ is initial velocity, $\boldsymbol{a}$ is acceleration, $\boldsymbol{t}$ is time and $\boldsymbol{s}$ is displacement.

2.4.2 Gyroscope

Gyroscopes are devices that measure angular velocity (°/s), which is the rate of change (speed) of rotation of an object about an axis. They are commonly used to determine orientation, and are typically present in applications that require autonomous navigation. The image below illustrates the three rotational motions (yaw, pitch and roll) measured by a gyroscope.

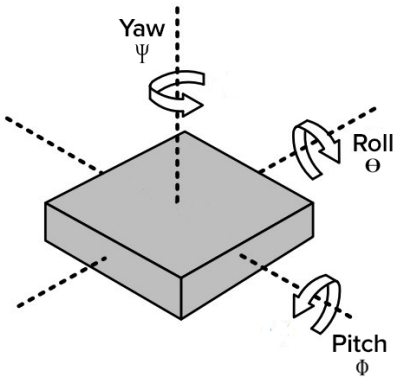

Figure 2-9 Yaw, Roll and Pitch axis

For applications that require orientation, accelerometers can determine the pitch (Φ) and roll (θ) angles from rest positions because of the effect of gravity on rotation an object about the x and y-axis. For the case of yaw (Ψ) or heading angle, since gravity has no effect on rotating an object around the z axis, the accelerometer is unable to determine heading on its own, hence, the need for a gyroscope or a magnetometer (compass) arises.

Angles calculated from gyroscope data also suffer from drift error like in accelerometers. This is a consequence of sensors that measure the rate of change of a physical attribute, rather than the attribute itself. Calculating the angular displacement about an axis using a gyroscope will require integration of the angular velocity measured, hence, errors in the measured data will result in a drift in the calculated angular displacement.

2.4.3 Laser Displacement Sensor

These are typically high precision sensors used to measure displacement between two points. The operation of the sensor is based from the principle of optical triangulation, as shown in Figure 2-10. A visible point of light is projected on a surface from a laser light source, and an image receiver positioned at a certain angle captures the reflection of the light spot, and then calculates the distance between the visible spot and its projection source.

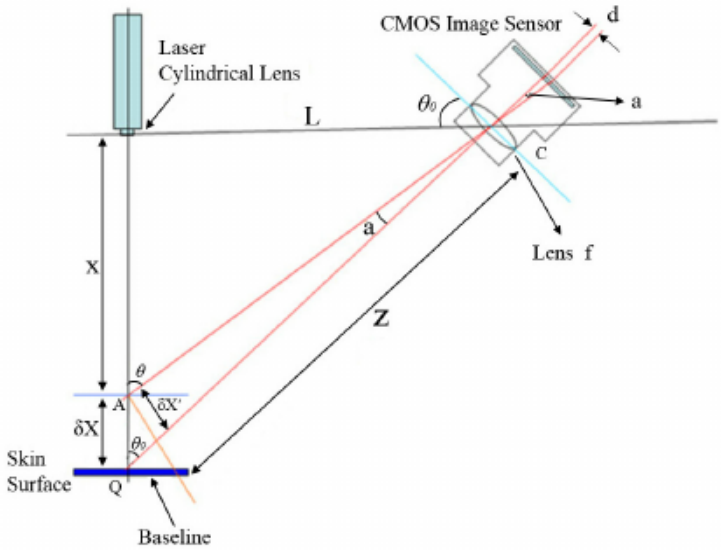

Figure 2-10 Optical triangulation geometry [53]

2.4.4 Doppler Speed Sensor

These are non-contact speed sensors used to estimate the speed and distance travelled of a moving object. It works with the Doppler effect principle, where the frequency of a wave propagated from a moving source increases or decreases depending on the direction of travel between the source and the observer. For example, the engine sound of a fast-moving vehicle heading towards a person increases in pitch (the frequency of a sound wave), this pitch begins to decrease when the vehicle moves passed (away from) the person, giving the sound a fading effect signifying a decrease in the sound's frequency.

A beam is emitted from the device towards the road surface, and then a receiver on the sensor captures the reflected signal, and compares the change in frequency to determine the speed at which the object is moving.

2.4.5 Kalman Filter

Error from sensors diminish the accuracy of the predicted trajectory, hence, there is a need to apply filtering to this acquired sensor data to minimise the effect of the error on the projected trajectory. The Kalman filter is a filtering concept commonly used in tracking and motion prediction applications [54]–[57]**.** It implements a predictor-corrector type estimator for real time data, and is ideal in the sense that it reduces estimated error covariance compared to other popular filtering concepts like moving average. Moving average estimates its filtered values by performing a moving average of current sensor data with a window of previous data. However, the disadvantage is

that, depending on the length of the averaging window (noting that the longer the window length, the smoother the estimated value), there is a respective phase difference between the actual and estimated value. In tracking/motion prediction applications, more emphasis is placed on the accuracy of the estimated result than there is on its smoothing, as smooth data could potentially (in high error cases) drift phase-wise away from actual values. In applications like missile control systems, delays are critical since real time tracking is essential to its operation. The Kalman filter is very effective in applications where accuracy and real time estimation is required, provided a mathematical model of the application can be established.

The basis for the Kalman filter is a series of mathematical expressions that define the prediction and correction aspects of its operation. It is a recursive filter (in the sense that the prediction of a future state depends on the present), and it relies on measurement samples taken at intervals at a constant rate. The basic theory to the filter involves making an estimation of the future, getting measurements from reality, comparing and moderating the difference between both values, then adjusting its estimate with the moderated value. Its accuracy is dependent on the derived mathematical model that defines its application.

2.5 Summary

Road roughness illustrates the longitudinal profile of a road due to deterioration of the surface over time, or simple bad construction. It is important to maintain good road conditions because of its effect on driver safety, ride comfort, fuel consumption, and maintenance cost (for both vehicle and road infrastructure). This requires regular monitoring of the pavement to detect or predict potential defects. The chapter discusses the theory of road roughness and monitoring, defining key terminologies. An assessment on current and popular methods of evaluating road surface conditions was also discussed, with focus on vehicle-based profilers (which the research is based on).

Vehicle-based profilers are high-speed profilers that allow faster road surface monitoring, with the convenience and flexibility of taking road measurements without the need for road closures. This is a major advantage compared to other profiling methods, but they do come with challenges as discussed in Chapter CHAPTER 4. Recent adaptations of these profilers rely on a combination of laser range finders, inertia sensors and cameras to evaluate road conditions, and these were discussed in Section 2.2.3.

The thesis also explores the idea of evaluating road surface conditions based on the vibration imposed on the vehicle. Section 2.3 reviews the theory of vibration analysis in both time and frequency domain, evaluating the various measurement techniques, representations, and terminologies. A typical vibration signal is formed of multiple signals with different frequencies and magnitudes. To precisely isolate and analyse each encompassing signal, a spectral analysis is required. Phenomenon like spectral leakages and discontinuity in FFT analysis require window functions to improve accuracy of the measured vibrations. Presently, there are several applications being explored to utilise vibration analysis in road transportation. Some of these are discussed in Section 2.3.5, and include monitoring, diagnostics, localization and speed detection.

CHAPTER 3

DESIGN OF THE LASER BASED ROUGHNESS MEASURING DEVICE

3.1 Overview

This chapter describes the hardware and software architecture of a Laser Based Roughness Measurement (LBRM) inertial profiler that was designed and developed for this research to verify the novel models and techniques described in Chapters CHAPTER 4 and CHAPTER 5. The LBRM device integrates several sensors constantly measuring different road properties to analyse road surface conditions. For this study, the device is mounted behind service vehicle as shown in Figure 3-1, a few inches above the ground, and in line with either tyre (i.e. left or right).

3.2 System Model

For road surface analysis, the LBRM device is expected to measure:

- Acceleration and angular velocity using the MPU-9150 Inertial Measurement Unit (IMU) that contains a MEMS (Microelectromechanical Systems) tri-axis accelerometer, tri-axis gyroscope, and a Digital Motion Processor (DMP) that allows it to process complex motion fusion algorithms.

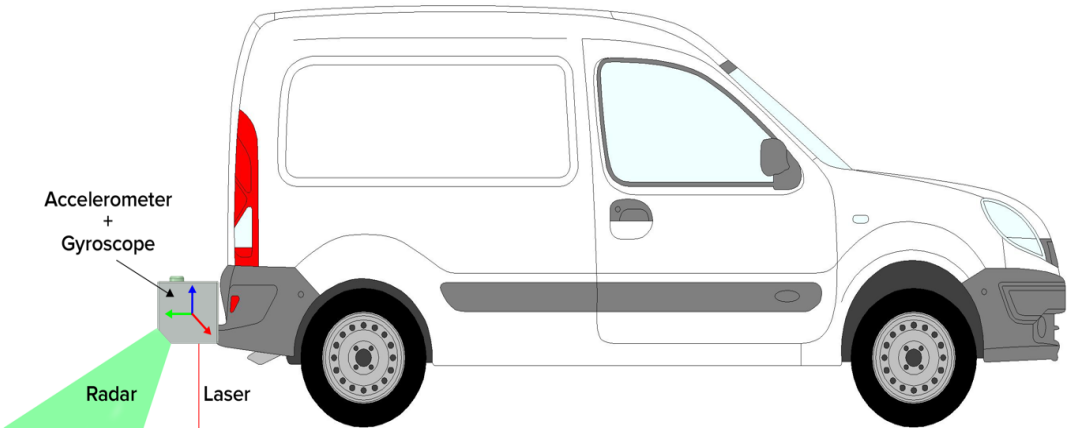

Figure 3-1 LBRM device mounted on a service vehicle

- Distance travelled with the PEGASEM GSS15 sensor that employs the Doppler effect principle to achieve a non-contact over ground speed sensing, with a measurement range between 0.1 – 400Km/h.

- Vertical displacement using the optoNCDT1700 laser sensor that applies the principle of optical triangulation to calculated displacement, and has a measurement range between 70mm – 320mm.

There are minimum power and communication requirements expected of the hardware to be suitable for use in the research. These requirements are primarily based off the four different sensors, which have all been chosen based on cost, size, ease of implementation, accuracy, and measurement range. Since the sensors operate separately, and each one of them have specific power and data transmission needs, an additional circuitry and MCU is needed to integrate with the sensors, compiling the different samples, and making them function as a single unit.

Table 3-1 Sensor power and communication information

Sensor	**Voltage**	**Current**	**Communication**	**Sampling rate**
MPU9150	2.375 – 3.465V	4mA	I2C	1000Hz
optoNCDT1700	11 – 30V	150mA	UART (RS422)	2500Hz
PEGASEM GSS15	8 – 32V	150mA	Digital IO pulse	100 pulses per meter

Using Table 3-1 above, the hardware's circuitry requires:

- a voltage input of at least 11V, that is able to supply more than 305mA of current (excluding current that will be used by other additional circuitry and MCU),

- an MCU that can support UART and I2C serial interfaces to communicate with the laser and IMU sensors,

- an interface that can communicate with a PC, transmitting samples at a frequency of at least 2500Hz. Hence, the chosen transmission protocol should support data transfer speeds between 37.5 – 75 kB/s (kilobyte per second) if a single sample is assumed to be between 15 – 30 bytes. This allows data to be transmitted in real time from the device, to be used in data analysis and processing.

- (optional) power saving features (enabling the ability to shut off power to the sensors when they are not in used), and a status feedback (which could be an LED, that allows device operators to easily identify the state of the unit during operation)

3.3 Circuit design and schematic

Following the requirements mentioned in section 3.2, the derived hardware is shown in Figure 3-2 below.

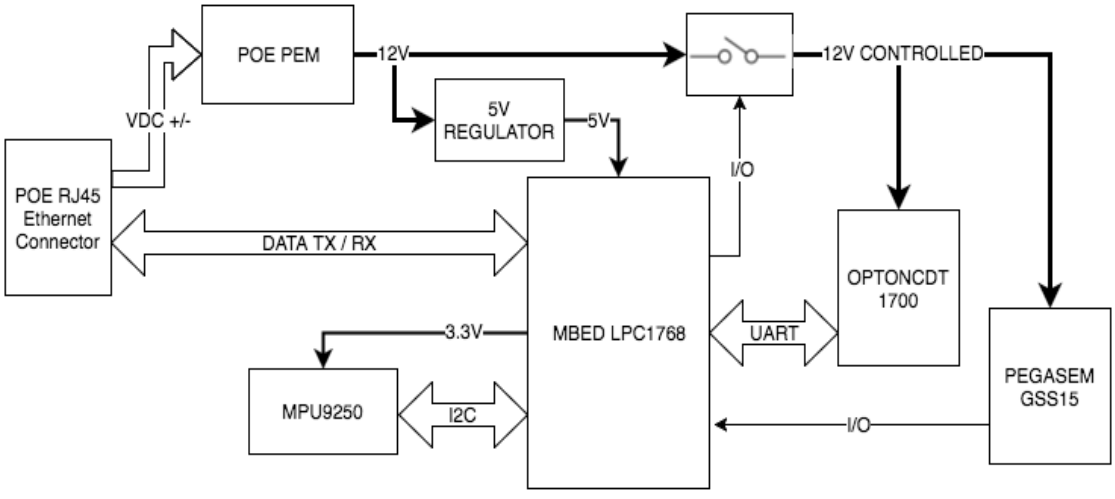

Figure 3-2 LBRM device circuit block diagram

Reviewing the circuit's block diagram, there are decisions that have been made to the design regarding the choice of circuitry, interface and communication protocol used, to conform to the proposed requirements, and these have been based on their respective

effectiveness, efficiency, and long-term reliability. Sections 3.3.2 and 3.3.3 concisely explains the operations of the different parts of the circuitry, and mentions any consequent trade-offs with commonly used alternatives.

3.3.1 Microcontroller Unit

The Microcontroller Unit (MCU) holds the firmware that defines the operations of the LBRM device, exposing a platform for communication and control, acting as a bridge between a user and the sensors. On request, the controller handles the power state, initialisation, and data collection from the sensors, compiling the different samples acquired into packets, and streaming them at set time intervals. There are fundamental features and specifications essential for an MCU to possess to properly carry out the tasks necessary for the LBRM device to function, some of these include, support for all communication interfaces required by the sensors, memory size, processor performance, and the amount of GPIO pins available.

The LBRM device is controlled by the MBED LPC1768 MCU, which was selected after a review of various MCUs based on the criteria mentioned earlier, and its relative ease of use.

MBED is one of many platforms and operating systems for creating Internet of things (IoT) devices. IoT devices are simply a series of embedded systems connected to the internet, that communicate with cloud/server based applications, working together to deliver a service.

The MBED LPC1768 is an ARM based microcontroller development board with a 40-pin Dual Inline Package (DIP) form factor. This microcontroller is designed for rapid prototyping of general embedded applications, and comes with built-in support for a host of commonly used I/O (Input / Output) interfaces including Ethernet, USB (Universal Serial Bus), SPI (Serial Peripheral Interface), I2C (Inter-Integrated Circuit), ADC (Analog to Digital Converter). With a 32-bit ARM Cortex-M3 core running at 96Hz, 512KB FLASH, and 32KB RAM (Random Access Memory), the LPC1768 is a high-performance system.

The LPC1768 is programmed in C/C++ language, and MBED provides a free online Integrated Development Environment (IDE), which is a browser-based editor where code is written, and then compiled on the cloud (i.e. on their web servers). The online IDE provides a C/C++ Software Development Kit (SDK) used for creating firmware/application that run on an MBED enabled device, which is programmable

via a simple drag and drop FLASH programmer. The SDK contains core libraries that interface with the microcontroller's Real Time Operating System (RTOS) and peripheral drivers, making application creation on the platform easier for developers.

3.3.2 Input Power Supply

The sole source of power to the LBRM device's circuitry is POE, which can source a voltage of 48VDC with maximum current of around 600mA. Current wise, POE supplies enough to power all sensors, MCU and any additional circuitry needed for the hardware to operate, but the supplied voltage should be extracted from the interface and stepped down to a value suitable for all internal components.

To extract the power from the POE interface, the hardware design includes a POE Power Extraction Module (PEM) PEM301C. The module, which is fully compliant with both the IEEE 802.3at and 802.3af standards, is designed to automatically detect the method of power transmission used by the POE injector/switch. The Injector infuses the power to the data line as described in the standards, stepping it down to 12VDC, which powers the LBRM device's circuitry.

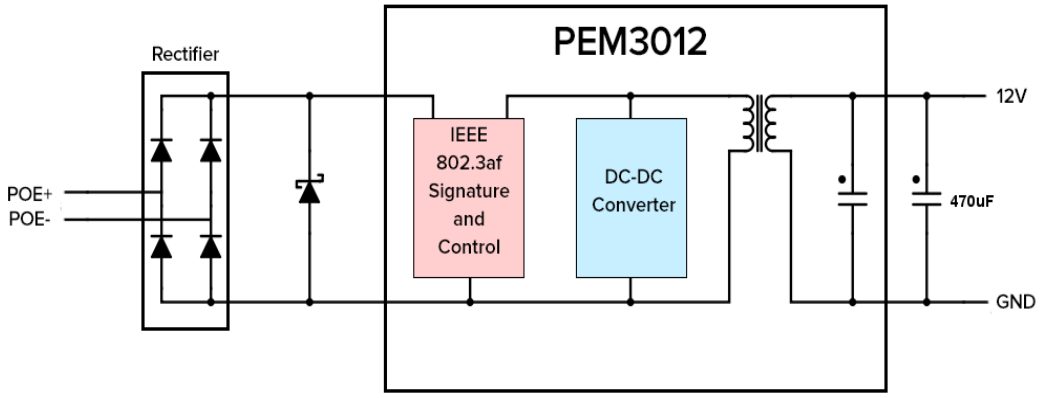

Figure 3-3 LBRM device's power supply circuitry

As shown in Figure 3-3, the POE voltage lines are passed through a bridge rectifier consisting of four diodes, which is used to maintain the polarity to the PEM3012 module, preventing any potential reverse polarity from the POE input. The PEM3012 contains circuitry that detects the standard used by the POE injector, steps down the 48VDC voltage to 12VDC, passes it through a transformer that isolated the output circuit from the input, and then provides a capacitor in parallel to the output terminals

for smoothing. The recommended additional smoothing capacitor of 470uF was added in the LBRM device design for extra smoothing.

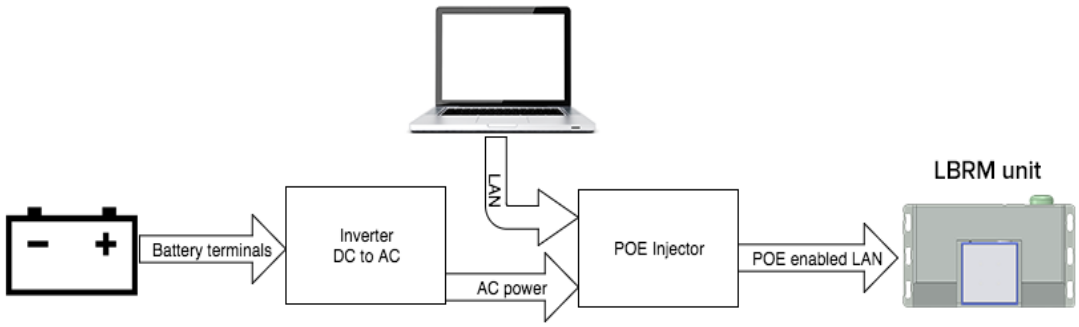

Figure 3-4 POE in a vehicular environment

Since the LBRM device is expected to be used in a vehicular environment, to supply POE in the vehicle during the research, an inverter was connected to the vehicle's 12V battery, converting the DC voltage to AC, as the POE injector being used is AC (mains) powered. Figure 3-4 shows the connection diagram.

3.3.2.1 Trade-offs with alternative designs

A common and typical concept usually employed in similar devices would be to directly power the unit via the car battery, eliminating the extra POE circuit from the resulting hardware, and potentially saving cost in design. The disadvantage to this model is that a separate interface needs to be made available to transmit information from the unit, in which case the popular choice (considering the desired transmission rate) would be USB.

Although using separate power and communication lines is a viable design model, POE offers the ability to supply power and transmit data to the unit from a single Ethernet RJ45 interface as described in section 3.3.4.1. This gives the convenience of having a single cable running out of the LBRM device, and allows the unit to benefit from all the features offered by the Ethernet interface, like LAN and internet communication, at a slight cost disadvantage.

3.3.3 Voltage distribution

This section describes how the 12V output voltage from the PEM3012 is distributed across the LBRM hardware circuit, powering the sensors and MCU during operation.

3.3.3.1 Conversion and regulation

The MCU in Figure 3-2 needs a voltage between 4.5V – 9V to be operational, in the hardware design, a step down switching voltage regulator converts the 12V PEM output to 5V, which is used to power the MCU. The switching voltage regulator is a more expensive and complex type of voltage regulation when compared to the more popular linear regulators, but the cost acquired in using switching regulators is gained in efficiency, which linear regulators suffer massively from.

Linear regulators operate with the same principle as voltage dividers, but they contain feedback loop, which is used to adjust/vary the resistance of the regulator depending on the load, resulting in a fixed output voltage. The main disadvantage to these regulators is their very low efficiency (as mentioned earlier), the power loss during operation is calculated using (3-1). The heat dissipated is directly proportional to the amount of current drawn by the load, hence, they are not ideal for high current applications, and would require heat sinks during operation.

$$W = (Vi - Vo) * I \tag{3-1}$$

Where, $\boldsymbol{W}$ is power loss in watts, $\boldsymbol{Vi}$ is input voltage, $\boldsymbol{Vo}$ is output voltage, and $\boldsymbol{I}$ is the load current.

Power to the MPU9150 (3.3V) is supplied via an on-board linear voltage regulator from the MCU. This is tolerable in this scenario since the current requirement for the IMU is only 4mA, and the voltage is stepped down from the MCU's operating voltage of 5V to the 3.3V required, hence, the power lost in heat is 6.8mW.

Switching regulators generally have much higher efficiency, and unlike their linear counterparts, the efficiency is not proportional to the amount of current supplied to the load, but is constant and usually stated in the component's datasheet. They operate by rapidly switching the input voltage on and off at a set duty cycle (which is determined via a feedback mechanism) to regulate the amount of charge that is received by the load. These regulators dissipate almost no power, but have the disadvantage of requiring a more complex circuitry compared to linear regulators.

$$W = VI\frac{(1-\eta)}{\eta} \tag{3-2}$$

Where $\boldsymbol{W}$ is power loss, $\boldsymbol{V}$ is output voltage, $\boldsymbol{I}$ is load current, and $\boldsymbol{\eta}$ is the regulator's efficiency

For a system that needs to step down a 12V input to 5V, with a load current of 300mA, the power loss in a linear voltage regulator is 2.4W, while the loss in a switching regulator rated at an efficiency of 80% is 0.375W using equation (3-2). This is a huge difference in the heat that will be dissipated by the regulator. For example, in room temperature of about 20°C, a regulator in a TO-220 package where the thermal resistance is typically rated around 70°C/W (temperature rise above room temperature per watt) will reach a temperature of 188°C for a linear regulator, compared to 46.25°C for a switching regulator. Therefore, a 5V switching regulator was preferred over linear to provide power to the LBRM unit's MCU.

3.3.3.2 Switching and control

The optoNCDT1700 and PEGASEM GSS15 sensors are powered from switchable source so they can be turned on or off when necessary, saving power during device operation. The switch consists of a transistor pair made up of a P-Channel MOSFET and NPN BJT, setup to act like an SPST (Single Pole, Single Throw) push button switch as shown in Figure 3-5 below.

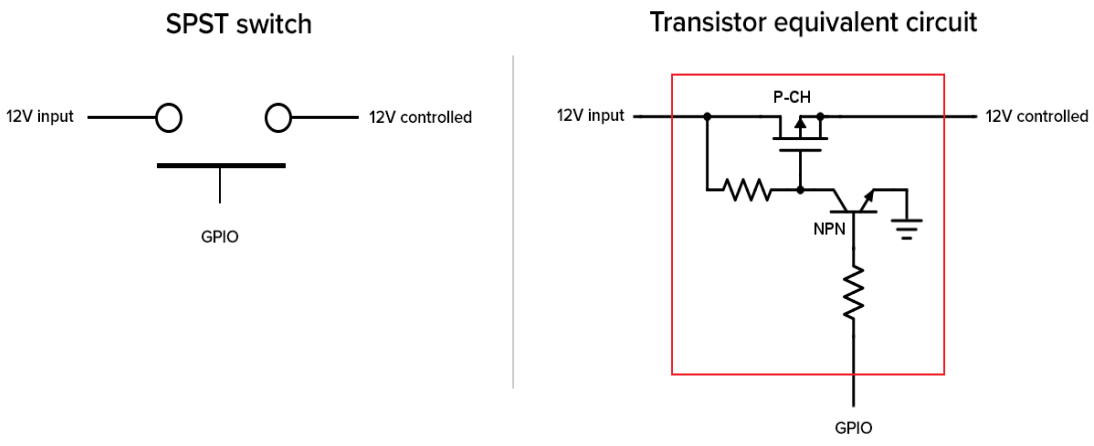

Figure 3-5 Transistor pair switch circuit

Controlling the on and off state of the switch is done by applying a voltage to the GPIO pin. When the voltage on the pin is 0V or left unconnected, no current flows through the base of the NPN transistor, leaving it in an unsaturated (off) state.

Therefore, the MOSFET's gate has the same voltage as the source via the pull-up resistor, while keeping the MOSFET in an off state, meaning that no power is supplied to the load. However, when voltage is applied to the GPIO pin, this allows current to flow through the resistor to the base of the NPN transistor. This subsequently turns on the NPN transistor permitting current flow from its collector to emitter, hence, pulling the MOSFET's gate to ground and creating a voltage difference between the gate and the source, turning on the MOSFET, and sending a 12V voltage to any load connected to the 12V controlled line.

3.3.3.3 Trade-offs in using transistors vs relays

Relays provide an alternate means of switching power to the sensors. Their equivalent circuitry is less complex, and they are generally used to switch both AC and DC loads, unlike transistors, which are only able to switch DC loads. Relays also provide complete isolation between the switching circuit and the load, which is a necessary when switching an AC load using a DC voltage. However, since they are electromechanical switches (i.e. they contain set of contacts with a physically moving part to open/close the switch), they suffer from wear and tear, degrading over time. This makes transistors more desirable for switching DC loads, when isolation is not a necessity.

3.3.4 Data transmission

3.3.4.1 PC to LBRM unit

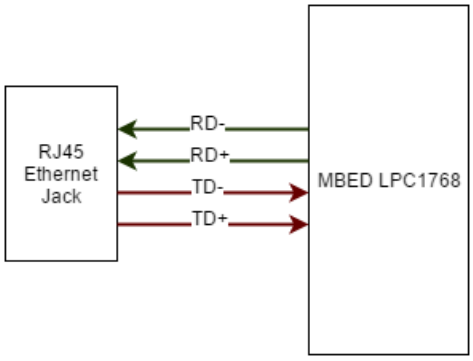

Figure 3-6 Ethernet data wiring

The LBRM unit communicates with a PC via Ethernet since the device is POE enabled (3.3.2), this allows data and power to be transmitted using a single RJ45 cable.

MBED LPC1768 has a built-in Ethernet network interface controller that enables network interfacing without the need for additional hardware. Figure 3-6 show the wiring diagram between the RJ45 jack and the LPC1768 that is required to connect

the unit to a network switch. Apart from the convenience of transmitting power and data over a single cable, communicating over the Ethernet interface lets the LBRM unit benefit from the features offered by TCP/IP, like being able to connect to a LAN or the internet, enabling remote data transmission.

Theoretically, the LPC1768 supports both 10 and 100Mbps data transfer speeds, but this is slower than the more popular USB standard (USB 2.0) which supports rates of up to 480Mbps. The newer USB 3.0 standard can transfer data upwards of 4Gbps, which is at least 40 times more than Ethernet. For applications that require very high data transfer rates, USB will have a clear advantage over Ethernet, but as the LBRM unit needs a transfer speed of around 500Kbps, which Ethernet is able to handle comfortably, it's the preferred option because of the other benefits and convenience it allows.

3.3.4.2 Laser to MCU

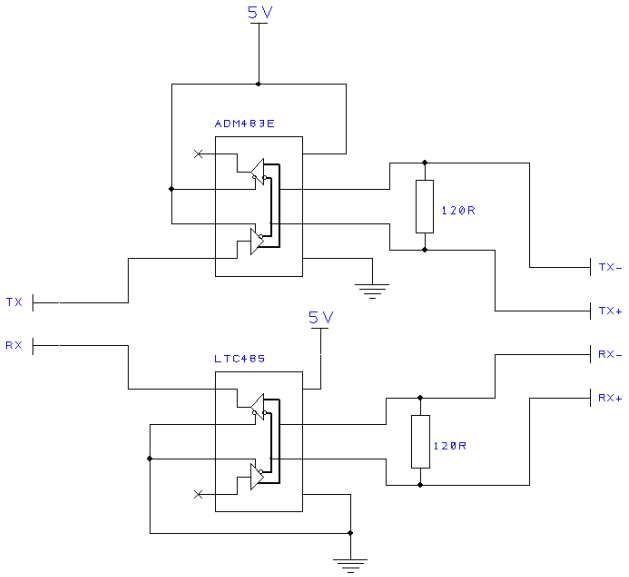

Figure 3-7 RS232 to RS422 transceiver circuit

The optoNCDT1700 laser sensor supports both analog (varying voltage) and digital (UART) means of retrieving measured data. In analog mode, the sensor varies the voltage on the signal line. The measured displacement is directly proportional to the voltage. Communication with the MCU is via the UART interface, but an additional hardware is required to transmit data from the laser sensor to the MCU as they support different standards (RS422 and RS232 respectively). The hardware shown in Figure 3-7 consists of low power data transceivers, and acts as the intermediary between the

interfaces, converting the signals received from one end to the appropriate signal on the other.

3.3.4.3 Radar to MCU

The PEGASEM GSS15 speed sensor has a UART (RS232), analog, and pulse output interface to obtain data from the sensor. Just like the laser sensor, the analog interface varies the voltage on a signal line, but in this case, the voltage on the line is dependent on the measured speed (1V per 100 km/h). The UART interface outputs both speed and distance travelled, where the sensor calculates the distance travelled using the measured speed and elapsed time. The LBRM's MCU communicates with the sensor over the pulse interface, mainly due to its ease of use and minimal wiring, since only one signal line is needed to obtain the measured information. The interface outputs the calculated distance travelled by generating a 5V pulse on the signal for every centimetre measured. Figure 3-8 shows a pull up resistor together with a smoothing capacitor, both used to stabilise the pulse line, preventing fluctuations on the signal when in Hi-Z (floating).

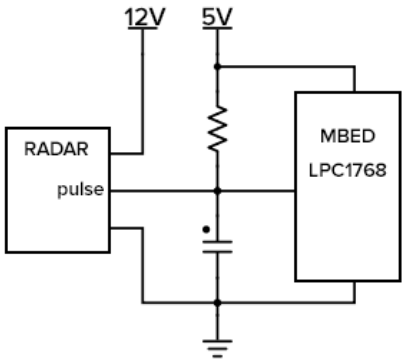

Figure 3-8 GSS15 wiring circuit

3.4 Firmware Architecture and Implementation

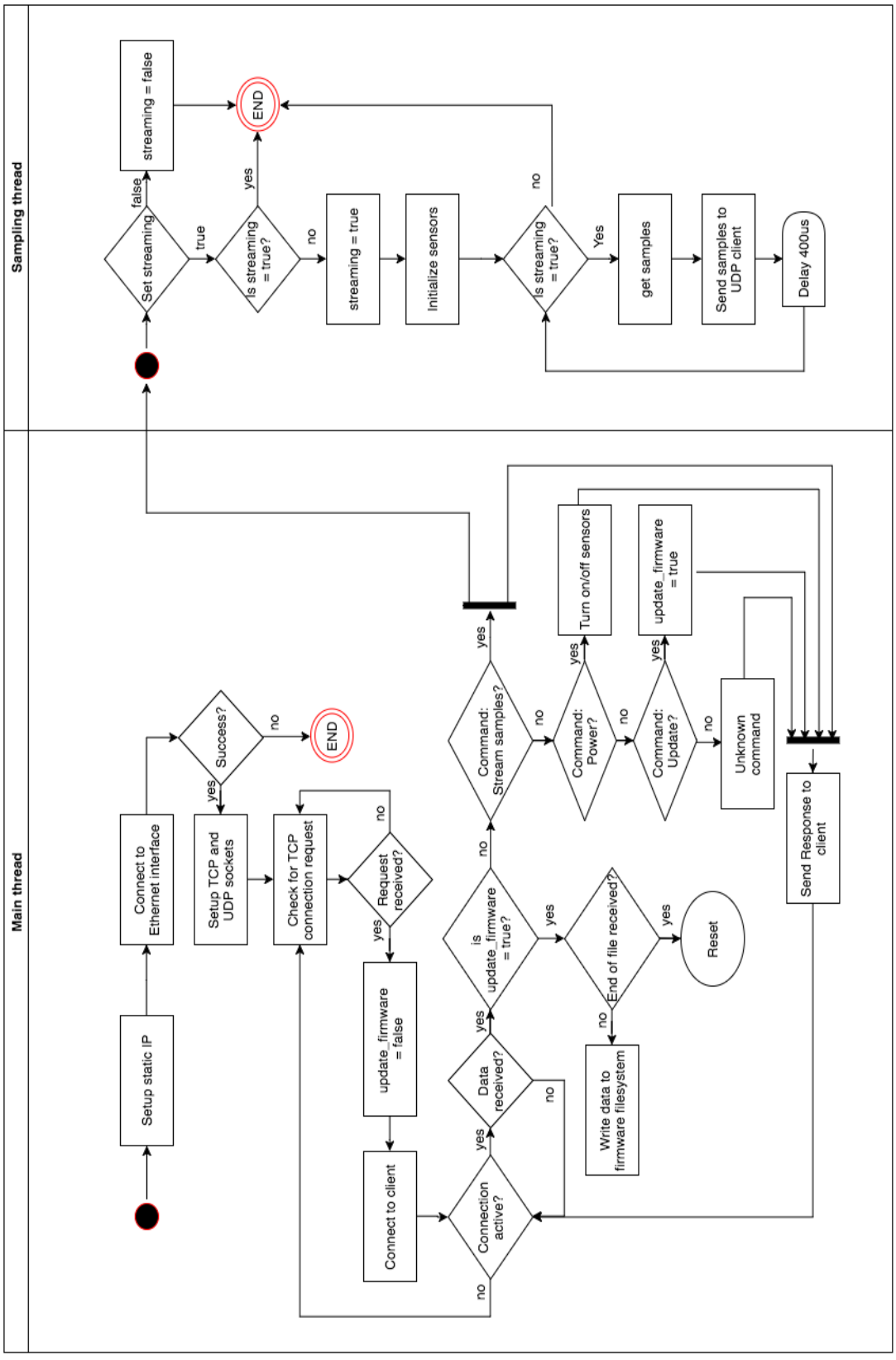

Figure 3-9 Firmware flowchart

The flowchart in Figure 3-9 above illustrates the operation of the LBRM device. The tasks has been split into different parts and analysed from Sections 3.4.1-**Error! Reference source not found.**, stating challenges experienced during development as a result of hardware and software limitations, and providing solutions to these issues.

3.4.1 Initialization and Ethernet setup

The initialization process occurs when the device is powered on, since there is no dedicated power on button/switch on the LBRM unit, connecting the RJ45 cable to a POE enabled switch (as described in 3.3.2) turns on the unit and begins this process.

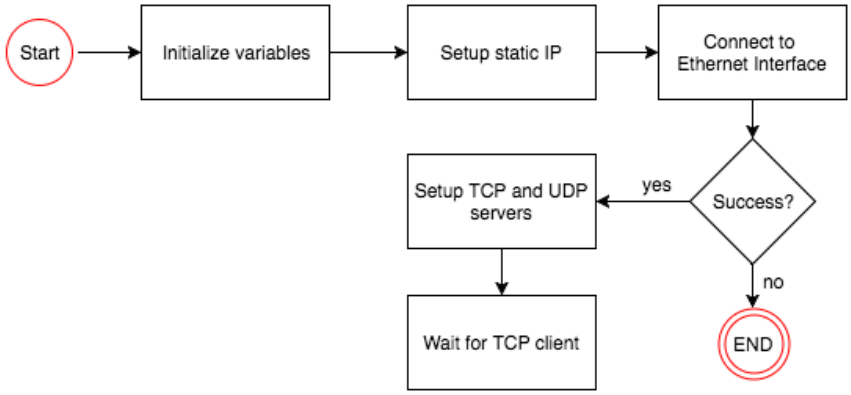

Figure 3-10 Initialization process

As shown in Figure 3-10, on device start-up, all variables are initialised, with GPIO pins configured as either inputs or outputs depending on their function. Most importantly the GPIO pin controlling the 12V power to the laser and radar sensors (as described in section 3.3.3) is configured as an output and set to 0, which guarantees that the switch is in an open/off state, cutting power to the sensors. The next step configures a static IP address for the unit's Ethernet interface. This means the chosen IP address will be used to connect to the unit when trying to establish a connection using a PC. Configuring the unit with a static IP eliminates the extra step of trying to figure it out, which would be the case in a DHCP setup, where the IP address is dynamically allocated by a DHCP enabled router. Although there are ways to eliminate this step in a DHCP environment, like setting up a hostname for the unit and making a TCP/IP connection using this hostname instead of an IP address, a DHCP setup would require a network router to allocate IP addresses dynamically. However, this is an extra hardware, which is unnecessary for current functionality.

To establish the connection between the LBRM unit and the PC using a static IP, the Subnet mask of both device needs to match, with the chosen IP address for the LBRM unit correlating with the IP address and subnet mask of the PC. A value of 254 on the subnet mask indicates that that section of the IP address needs to match for both communicating devices, while a value of 0 indicates that any value between 0 and 254 can be set for that section of the IP address. In a scenario where the IP address and subnet mask of the PC is set as 169.254.10.252 and 254.254.0.0 respectively, according to the subnet mask, the first two section of the IP address needs to match (i.e. 169.254.x.x), while the other two can be any value between 0 and 254. Therefore, an IP address of 169.254.0.100 will be a valid value for the LBRM unit to establish a successful connection between the devices.

If an Ethernet connection cannot be established between the LBRM unit and a PC after device start-up, the device goes into an inactive state, where a restart is required in-order to try to establish a new connection. Else, on successful connection the device begins a TCP and UDP socket server (described in 3.4.2 below) waiting for a client connection request.

3.4.2 TCP and UDP socket servers

The TCP and UDP servers provides the channel for communication between the device and PC. The TCP protocol handles device commands and response data, while the UDP protocols is used to stream samples from the device. The reasons for using the two different protocols for transmission is explained in section **Error! Reference source not found.**, which describes the problems experienced from using a single protocol i.e. either TCP or UDP, giving reasons why these problems occur, and how splitting the operations improves overall transmission performance.

The basic operation of the TCP server is shown in Figure 3-11 below, the LBRM device initialises the server, binds to port, and begins listening and waiting for a connection request from a client (which in this case is the PC wanting to communicate with the device). For a client to establish a successful connection with LBRM unit's server, it needs to know the static IP address assigned to the unit (as explained in section 3.4.1) and the port number the server listens on. Both TCP and UDP servers are bound to port 1000. When the server accepts the connection request from the client, data can be transmitted bi-directionally between server and client.

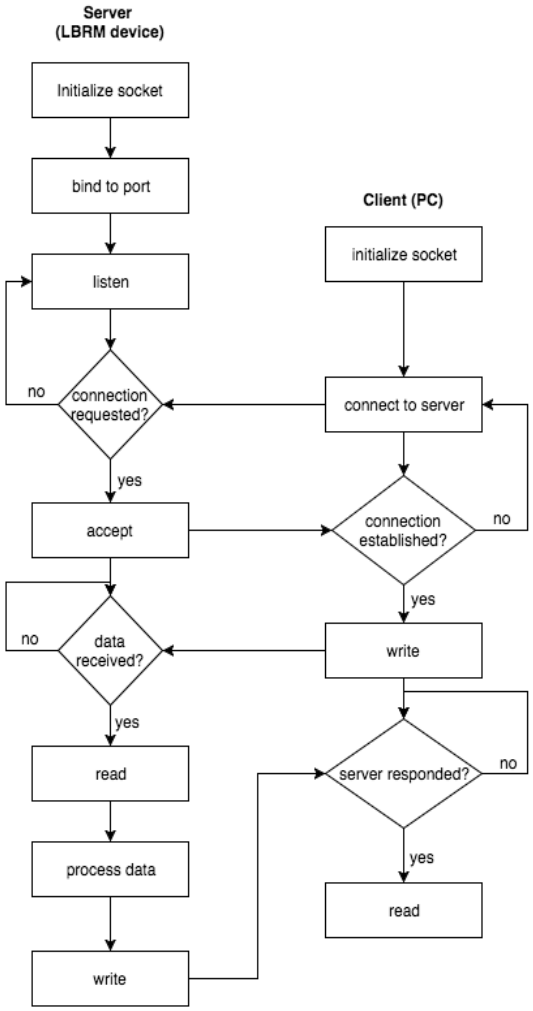

Figure 3-11 device TCP flowchart

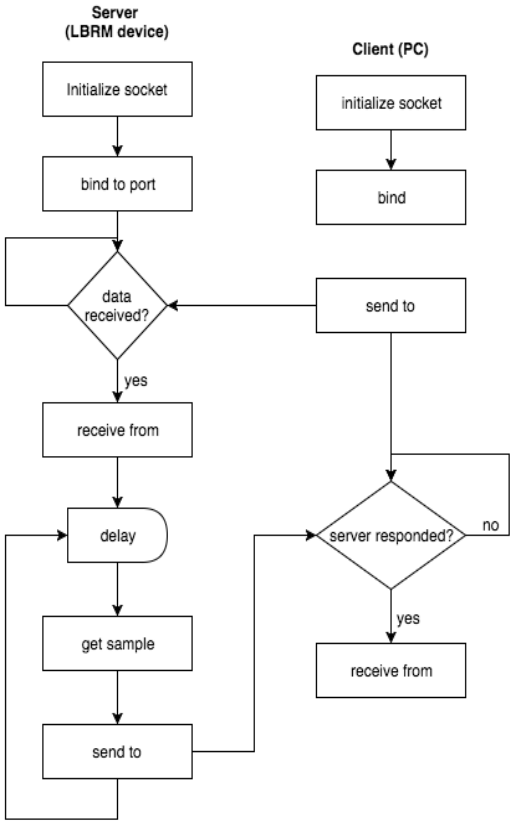

Figure 3-12 device UDP flowchart

Now the PC can send a command to the LBRM unit, which is then processed by the unit to determine and execute the necessary command, and then sending a success or fail response back to the PC. Details about command and response format is shown in section 3.4.3.

As shown in Figure 3-12, the UDP protocol requires no official connection request for data transmission to occur as it is a connectionless protocol. With the server's IP and port number, the client can send a message across, but for the client to be able to receive broadcasted message from the server, an initial message needs to be sent from client to the server for the server to establish a link to the client. The content of this message is irrelevant, as any sort of byte or text would do.

Initial designs for initializing the UDP server required it to run on a different thread to that of TCP. As UDP is only used to transmit sample packets, it is only needed

when a stream command is received from a client, hence, the UDP server is run on a new thread when a client requests samples to be streamed, and then the thread is terminated on end of stream. This implementation had a negative effect on the data transmission rate for both the TCP and UDP streams, which is described in section 3.6, explaining the effect of thread on performance.

An improved design runs the UDP server on the same thread as TCP, with both initialise at the same time. The flaw to this design is that the UDP server starts running at the same time as the TCP server, hence continues when not required and potentially using up resources, but since the amount of wasted resources is not very significant, the huge improvement in transfer rate compared with the previous model makes it preferred.

3.4.3 AT command set

The LBRM unit is driven via the Ethernet interface using a set of custom predefined messages referred to as AT commands. These messages have specific formats for different instructions, and the LBRM device has been programmed to understand and execute the instructions, sending a formatted response back to the client indicating success or failed, depending on the instruction being executed successfully or not.

The AT commands are case insensitive, begin with the prefix "***A T***", and should be terminated with a $<\boldsymbol{LF}>$ Line Feed character (hexadecimal code 0x0A) usually represented in ASCII as "$\backslash \boldsymbol{n}$".

$\boldsymbol{AT x0{=}0}<\boldsymbol{LF}>$ is a sample AT command showing the typical format of the messages.

- The first two characters "***A T***" is used by the unit to differentiate an AT command from ordinary data.
- The character "***x***" represents the nature of the command being sent, i.e. indicates a write or execution command, which is represented by "***w***" or "***x***" respectively.
- The character "***0***" indicates the type of command to be executed, which is a number between 0 and 2 inclusive.
- The fifth and sixth characters "=" and "***0***" in execution commands sends the specific instruction to be carried, for example enable, disable etc., but for write commands, the sixth character could be a text (i.e. a series of characters) used to send additional data to be used by the unit.

- The $<LF>$ line feed character is required to signify the end of the message.

Response from the LBRM unit has the format $ATR?<1>?<2>$ where:

- The characters "ATR" are identity characters that indicates the message is a response to a command.
- $<1>$ is the AT command being responded to.
- $<2>$ is the result of the command execution, which is either a "1" or "0" indicating whether the command executed successfully or not.

For example, if the command $ATx0=0<LF>$ was sent to the unit to be executed, to signify the command was successfully executed, the unit will reply with $ATR?ATx0?1$.

Sections 3.4.3.1 - 3.4.3.3 describe the three features that can currently be controlled using AT commands, and figure x below illustrates the decision making process taken when executing a command received from a client application.

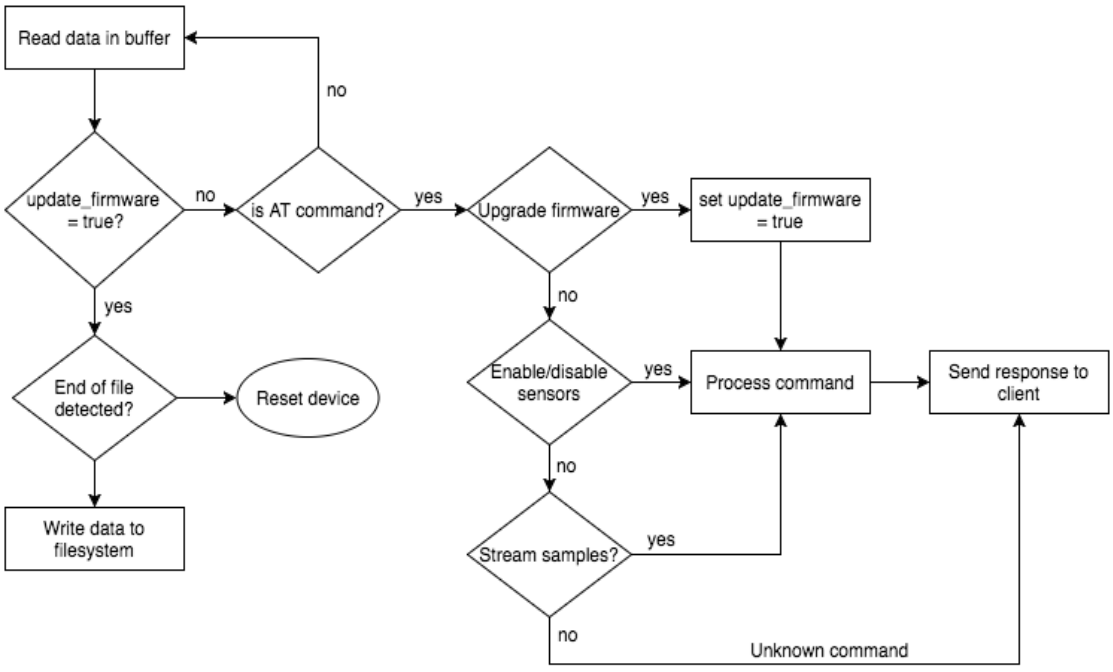

Figure 3-13 Firmware AT command handler

3.4.3.1 Device firmware upgrade

AT command: $ATw0=<1>$

Response: $ATR?ATw0?<2>$

This is a write command that is used to upgrade the LBRM device's firmware over a network. In the command, "*<1>*" is a text indicating the version of the firmware being transferred from the client. Reviewing Figure 3-13 above, sending this AT command enables the variable "***update_firmware***", which means that any subsequent data sent over the channel is written to a new file (which is named according to the firmware version) until the end of file message "***ATw0<CR><LF>***" is detected. Where *<**CR**>* is the carriage return character commonly denoted as "\r" in ASCII or 0x0D in hexadecimal. Before the new firmware is transmitted to the unit, the client must wait for the "***ATR?ATw0?1***" response from the unit, indicating that the device has successfully processed the command and is awaiting the new firmware, which needs to be sent as a series of bytes and terminated with the end of file characters.

After the new firmware has been written to the file system, the previous firmware is deleted and a reset command is executed. This process restarts the LBRM unit and forces it to run using the new firmware.

3.4.3.2 Sensors power on/off

AT command: ***ATx1=<1>***

Response: ***ATR?ATx1?<2>***

This is an execution command that enables/disables the 12V switch mentioned in section 3.3.3, thereby turning the sensors on/off. The commands:

- ***ATx1***=1 turns on the sensors, while
- ***ATx1=0*** turns the sensors off
- and both commands should respond with either ***ATR?ATx1?1*** or ***ATR?ATx1?0*** depending on whether the command executed successfully or not respectively.

3.4.3.3 Stream samples

AT command: ***ATx2=<1>***

Response: ***ATR?ATx2?<2>***

This is an execution command that begins the process of sampling data from all the sensors, compiling them into packets (as shown in Table 3-2) and streamed over the UDP channel to the client.

The data sampling process is executed on its own thread in order to avoid blocking the main thread. Since the main thread runs the base processes handling the TCP/UDP servers, blocking this thread to capture samples would significantly affect overall system performance. Also, since the samples from the sensors are expected to be captured at specific time intervals for data consistency (which greatly improves accuracy during post processing), having the process running uninterrupted is vital. The commands:

- ***ATx2***=1 starts streaming sample packets over UDP, but this command will fail if the sensors are powered off.
- ***ATx1=0*** stops streaming the samples,
- and both commands should respond with either ***ATR?ATx2?1*** or ***ATR?ATx2?0*** depending on whether the command executed successfully or not.

3.4.4 Sample structure

The LBRM unit streams a 23-byte sample through its UDP server. This 23-byte data contains measurement values from all the sensors, and is shown in Table 3-2.

Table 3-2 Device sample format

0	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***11***
'a'	't'	's'	'?'	S0	S1	L0	L1	AX0	AX1	AY0	AY1
12	***13***	***14***	***15***	***16***	***17***	***18***	***19***	***20***	***21***	***22***	
AZ0	AZ1	GX0	GX1	GY0	GY1	GZ0	GZ1	T0	T1	D	

From the sample structure shown in the table above, bytes:

- [0 - 3] are identity bytes used by client applications to detect the start of a sample.
- [4 – 5] is a 16-bit integer value that contain the number of bytes left to be read.
- [6 – 7] is a 16-bit integer value of the laser displacement measurement.
- [8 – 9] is a 16-bit integer value of the x-axis accelerometer measurement
- [10 – 11] is a 16-bit integer value of the y-axis accelerometer measurement
- [12 – 13] is a 16-bit integer value of the z-axis accelerometer measurement
- [14 – 15] is a 16-bit integer value of the x-axis gyroscope measurement
- [16 – 17] is a 16-bit integer value of the y-axis gyroscope measurement

- [18 – 19] is a 16-bit integer value of the z-axis gyroscope measurement
- [20 – 21] is a 16-bit integer value representing the temperature of the unit
- [22] is an 8-bit integer value representing the distance travelled, measured by the radar

Note: all 16-bit integers from the sample needs to be converted to their real world equivalent values as follows:

3.4.4.1 Laser displacement

To obtain the laser's displacement (in millimetre), the MSB (Most Significant Bit) of bytes 6 and 7 must be ignored before being used in equation (3-3) below. These bits are only used to signify low and high bytes in a 16-bit word, where 0 indicates that the byte is the lower byte, and 1 indicates the high byte.

$$displacement\ (mm) = \big((value * 0.0000623167) - 0.01\big) * 250) + 70 \tag{3-3}$$

A more detailed explanation of the displacement calculation can be found in the optoNCDT1700 datasheet.

3.4.4.2 Acceleration

Acceleration values have a range between -1 g and 1g.

$$Acceleration\ (g) = \frac{value * 2}{32768} \tag{3-4}$$

3.4.4.3 Angular velocity

Value have a range from – 125 to +125 degrees per second (°/s)

$$Angular\ velocity\ (°/s) = \frac{value * 250}{32768} \tag{3-5}$$

3.4.4.4 Temperature

$$Temperature\ (°C) = \frac{value}{340} + 36.53 \tag{3-6}$$

3.4.4.5 Distance travelled

This byte contains an incremental integer between 0 and 255 for ever centimetre travelled by the laser. This value is reset to 0 after 255 is obtained.

3.5 Validation Tests

Test items	**Requirements**	**Procedure**	**Result**
Sampling rate	2500Hz	Write PC application to read samples from LBRM unit. Set a timer on application to count number of samples read from LBRM unit per second Number of samples per second is expected to be approximately 2500	Pass
Data integrity	Laser proximity sensor	Mount the LBRM unit behind vehicle, elevated to about 120mm off the ground. Using pre-set blocks of known height, place blocks underneath laser point. New laser reading should be the difference between the mount height and block height	Pass
	Accelerometer	Place LBRM device on a flat surface. Repeat for all six sides of the device For each side, the axis on the vertical plane should reflect the acceleration due to gravity i.e. approximately $9.81 m/s^2$	Pass
	Doppler speed sensor	Mount LBRM unit behind vehicle. Drive vehicle between two points of known distance. Distance recorded by sensor should reflect the distance between the two points.	Pass

3.6 Design challenges

One of the major challenges was the effect of threading on the overall system performance. Due to the requirements for the profiler, the use of threads in the software is essential, because both sensors need to be sampled at specific intervals to maintain the profiler's output sampling rate. In addition, inconsistencies in the time between each sample would reflect post-processing inaccuracies when performing Fast Fourier Transform (FFT) analysis. The problem is, increasing the number of threads used causes a negative effect on system performance. For the MCU used, the maximum number of threads that could be active without affecting performance was three. This was not sufficient for all services that required concurrent execution. To overcome this, services were paired up and run on separate threads based on priority. For this case, thread 1 ran the main application, thread 2 handled the sensor sampling for the laser and accelerometer, and thread 3 handled the TCP and UDP servers.

Initial software design used the TCP protocol for sending sample data and receiving control commands. This resulted in samples not transmitted at the desired maximum rate of 2500Hz. This was due to two factors, first was the available resources on the MBED MCU, and the second was due to the nature of TCP protocols. TCP protocols guarantee the data transferred arrives in the same order, and this makes the protocol heavyweight because it performs flow control for congestion control and data reliability. On the other hand, UDP is lightweight, at the expense of not guaranteeing the order of arrival for packets sent, or in worst cases, complete loss of packets. Transmitting the sample over UDP allowed the desired sampling rate to be achieved, and since the connection to the user application is direct (switch), there is less chance of packet loss. Although this meant that the samples transmitted were not in the order they were sent from the device, each sample was tagged with an index, and rearranged in the client application to maintain the order.

3.7 Summary

This chapter describes the hardware and software architecture of a Laser Based Roughness Measurement (LBRM) device for profiling road surfaces. Essentially, the profiler contains a laser displacement sensor and an accelerometer, used to generate the pavement's profile. The profiler was developed to validate the profiling techniques proposed in this research. Section 3.2 outlines the properties of the profiler, which is capable of a sampling rate of 2500Hz, and supports Power Over Ethernet (POE). It

supports both TCP and UDP connections for communicating with a user application, and allows one client application per connection. The profiler acts as a server, waiting for a connection request over TCP. On successful connection, the client application can send control commands to the profiler using custom pre-defined AT commands listed in Section 3.4.3. UDP protocol was employed to stream measurement samples to the user application because of performance bottlenecks (described in Section 3.6) experienced from using TCP.

CHAPTER 4

HALF-WAVELENGTH PEAK MATCHING TECHNIQUE FOR EVALUATING LONGITUDINAL ROAD PROFILES

4.1 Overview

This chapter describes a novel integration-drift correction model based on a peak matching principle applied to the derived displacement from an accelerometer. The basis for the model compares the displacement of both accelerometer and laser, estimating a correction offset based on the laser's displacement.

Road profiling is an essential part of its maintenance, improvement and development in the transport industry. Longitudinal profiles evaluate the vertical/height irregularities of a pavement, identifying areas of concern, and from both vehicle and road user standpoint, the road condition plays and important role. For vehicles, the road surface texture can affect longevity of certain parts (typically unsprung mass), due to vibrations caused by consistently driving on rough or uneven road surfaces. As a road user, ride smoothness and comfort is the main factor of concern, and this is especially important as it is (to an extent) relative to the driver's control and balance during transit.

Realizing fast, efficient, scalable, and accurate measurement concepts for evaluating longitudinal road profile is necessary for the growth and development of the transport industry. Conventional methods like the 3m Rolling Straight Edge (RSE) is inefficient and requires manually intervention of a person to move and take measurements from the unit. Recent concepts use sensors like camera, laser, accelerometer, gyro, or a combination of these mounted on a vehicle, connected to an onboard computer to capture and process data from the sensors automatically. This allows faster profiling, at the expense of more complex processing methods. With localization devices like GPS used to tag road profiles to their respective locations.

A popular profiling technique uses a combination of an accelerometer and a laser sensor. The laser sensor measures the actual road irregularity (they offer precise displacement measurements at high sampling rates $>$ 2500 samples per second), while the accelerometer compensates for any vertical offset (during transit) from the vehicle's chassis. For example, for a design that has the laser and accelerometer unit mounted on the rear chassis of the vehicle, when driving over a road hump, at the instance where the rear tires are at the highest point of the hump, an angle of elevation is created between the base of the vehicle and the pavement. At this instance, the measurement recorded by the laser is at least greater than the height of the hump. For low frequency motions like this, an accelerometer can detect these displacements, reflecting this raised displacement. The actual road profile is evaluated by the difference between the displacement measured from the laser and accelerometer sensors. [58] describes a similar implementation of this system using two accelerometers (for acceleration accuracy) and a laser sensor mounted in front of the vehicle.

The major source of error in this approach is the process of estimating the displacement from the accelerometer samples. Acceleration is converted to displacement by double integration, but in practice, this process requires additional post processing techniques to obtain accurate results. The minimal requirement for the conversion process uses a high pass filter on the calculated displacement, to eliminate any offset imposed on the derived displacement due to integration. This is especially important in this application because the axis of concern (vertical axis) is affected by the gravitational pull. Hence a DC offset equal to the gravitational acceleration (approximately 9.81 m/s^2) is reflected on the accelerometer's z-axis. Integration amplifies errors present in a sample, and for applications like these where the integration is cumulative (i.e. an integral of a sample depends on the integral of its predecessor), a low frequency signal is injected into the derived displacement due to the cumulative error. Therefore, a high pass filter is required after integration to eliminate this low frequency error signals. Applications

requiring more accurate displacement results like [29], where an accelerometer is used to detect when an object's position changes, by computing the displacement from its previous location, a Kalman filter is used with the help of other supplementary sensors.

Table 4-1, extracted from the series 700 road pavement standard [59], highlights the significance of the proposed model for inertial profilers. The table defines the longitudinal thresholds (4mm and 7mm), stating their respective acceptable tolerance which determines the eligibility of a road. According to the standard, road surfaces exceeding 10mm have zero tolerance. Due to the fine margins between the thresholds, the precision (to the nearest mm) of road measurement equipment is vital.

Table 4-1 Road surface irregularity thresholds and limits

	Surfaces of each lane of carriageway, each hard strip and each hard shoulder for each irregularity limit				**Surfaces of each lane of bituminous binder courses for carriageway, hard strip and hard shoulder for each irregularity limit**				**Surfaces of lay-bys, service areas, and associated bituminous binder courses for each irregularity limit**			
Irregularity Limits	4 mm		7 mm		4 mm		7 mm		4 mm		7 mm	
Length (m)	300	75	300	75	300	75	300	75	300	75	300	75
Category A* Roads	20	9	2	1	40	18	4	2	40	18	4	2
Category B* Roads	40	18	4	2	60	27	6	3	60	27	6	3

The rest of this chapter is split into three sub sections; Section 4.2 describes the novel Half-Wavelength Peak Matching (HWPM) model. Section 4.3 show the results obtained from the tests carried out, analyzing all relevant findings, while Section 4.4 gives a summary and final thoughts on the chapter.

4.2 Half-Wavelength Peak Matching (HWPM) model

The HWPM model compares and matches the peak amplitudes of two waveforms, where one acts as a reference (which ideally is the waveform expected to be more accurate between the two), and the other is the active waveform that is adjusted based on the reference.

For this application, the aim is to correct any errors in the displacement derived from the accelerometer using the laser samples. Between both sensors, the laser sensors are a lot more accurate (±0.25% error margin as stated in the datasheet) and are better suited as the reference since, they are primarily displacement sensors compared to accelerometers, which are acceleration sensors, and must be converted to displacement using mathematical models.

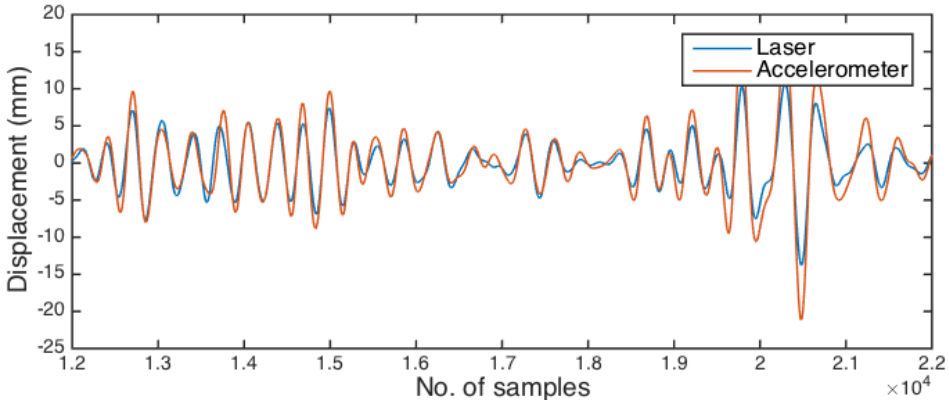

Figure 4-1 Laser and accelerometer samples with band-pass filter with f_c (1 – 4Hz)

Figure 4-1 show the low frequency (<4Hz) characteristics of both laser and accelerometer displacement samples. In the image, both samples have been passed through a band-pass filter with cutoff frequencies of 1 – 4Hz. The samples were taken using the LBRM device described in Chapter 4. The laser displacement (dL) indicates the vertical displacement from the unit's mount position (on the vehicle) to the road surface, while the accelerometer displacement (dA) indicates the longitudinal displacement of the chassis. In theory, for an even road surface, dL and dA are expected to be identical since the main source of change to the sensors is the movement from the chassis. In practical tests, the difference between dL and dA come from uneven road surfaces where the dL is expected to indicate lower or higher displacement compared to dA at points where the laser beam measures over a hump or pothole respectively.

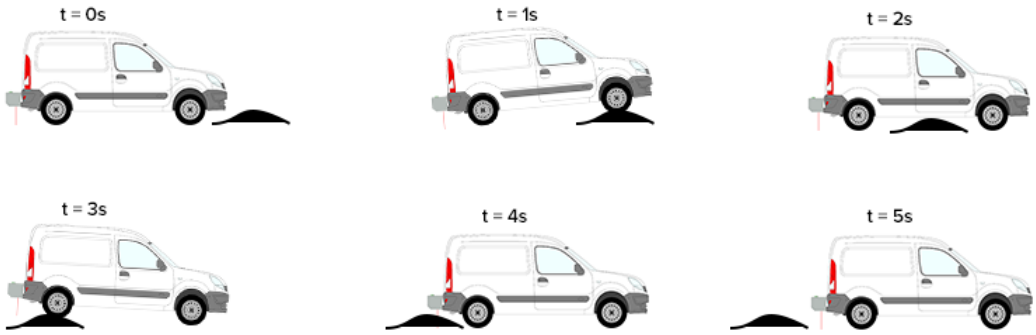

Figure 4-2 Vehicle dynamics over road bump

Consider the measurement vehicle driving over the road bump in Figure 4-2, ignoring the effect of the vehicle's shock absorbers for easier analysis. At 0 seconds, before the vehicle drives over the bump, the laser sensor measures a value equal to the vertical distance between the road surface and its mount position (h), while the accelerometer

measures minimal to no change on its vertical axis. At time 1 second, the front tires of the vehicle are at the highest point of the bump, this lowers the rear end of the vehicle (where the unit is mounted) slightly, due to the vehicle pivoting on the front wheel. This causes the laser sensor to record a value lower than (h), which is also reflected on the accelerometer. At the 2-second mark, the sensor measurements return to their neutral values indicated at the start. After 3 seconds, the rear tires are at the highest point of the bump, which raises the rear end of the vehicle, causing the laser and accelerometer sensors to measure a higher value. The crucial measurement required is at time 4 seconds, where the beam of the laser sensor is directly above the bump. At this point, there is a difference between dA and dL because there is no vertical displacement of the vehicle chassis, hence no change measured by the accelerometer, unlike the laser sensor, which measures a reduced displacement due to the bump. The goal of the HWPM model is to eliminate these false displacements (observed between 1 – 4 seconds, caused by front/rear tires moving over road anomalies) sensed by the laser.

The HWPM model is implemented in four phases as described below:

i. Peak detection
ii. Peak classification
iii. Peak matching
iv. Edge smoothing

4.2.1 Peak Detection

This is the first part of the HWPM process, which identifies low frequency peaks (i.e. P1 – P5 in Figure 4-3) in both reference and active waveform that needs to be corrected/matched. There were two methods considered in the peak identification:

4.2.1.1 Mean Threshold (MT)

The MT method identifies peaks based on the absolute maximum value (i.e. ignoring polarity) between two points that intersect the mean value of the waveform, where a positive peak is the highest value greater than the mean, and a negative peak is the lowest [60], [61]. Analyzing the waveform in Figure 4-3, P1, P3, P5 are examples of positive peaks, while P2, P4, P6 are negative peaks. Where 'x' indicates the points on the waveform that intersect with the mean value, while a peak is the highest or lowest value between two consecutive 'x' points.

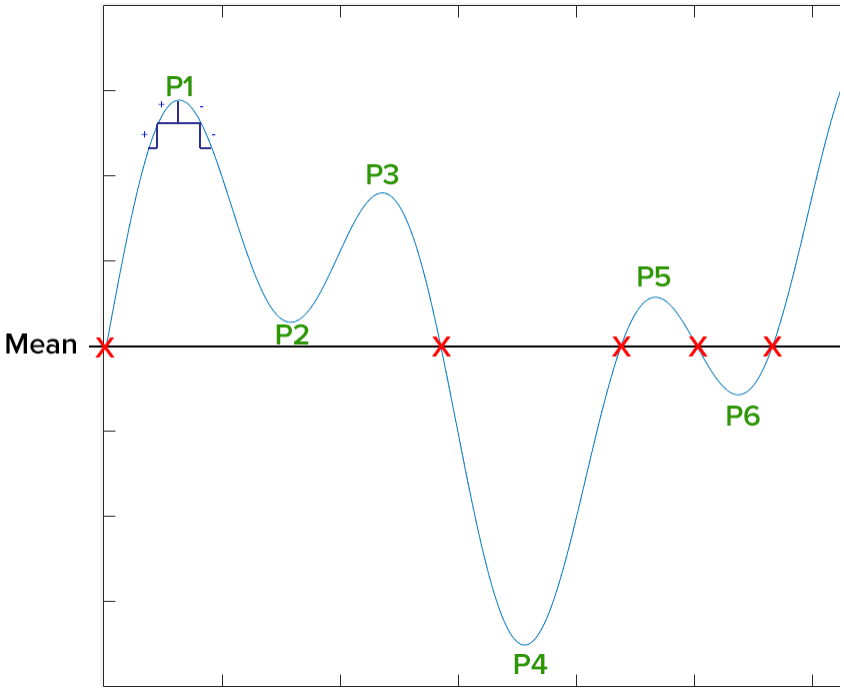

Figure 4-3 Waveform showing peak detection terminologies

Pseudocode: MT peak detection

```
1.   calculate mean for all samples in 'input_buffer'
2.   subtract mean value from all samples in 'input_buffer'
3.   initialise 'output_buffer' with values equal to 0
4.   FOR each sample 'S' in 'input_buffer'
5.       IF polarity of 'S' is the same as 'previous_S'
6.           IF polarity is –, & 'S' is less than 'peak_S'
7.               set 'peak_S' to 'S'
8.           ELSEIF polarity is +, & 'S' is greater than 'peak_S'
9.               set 'peak_S' to 'S'
10.      ELSE
11.          set value in index of 'output_buffer' to 'peak_S'
```

The pseudocode above shows the program flow for detecting peaks in a signal using the MT technique. Firstly, the mean is calculated for all values in the sample set, and then this calculated mean is subtracted from each sample to eliminate the DC offset in the signal moving it to the zero-threshold mark. The purpose of this is to be able to compare polarities of the samples when identifying positive and negative peaks. Using a loop to iterate through the samples, the polarity of each sample is compared with the previous sample in the buffer, and a threshold is determined as the point where the

polarity between successive samples differ. At which point the highest detectable peak is stored in the *output_ buffer*, and the process is repeated (lines 4 – 11).

This method of peak detection is unsuitable for this application because it assumes all positive and negative peaks occur above and below (respectively) two consecutive mean intersections, therefore, peaks P2 and P3 will ignored since P1 is the highest value between the intersections.

4.2.1.2 Difference in Gradient Polarity (DGP)

Unlike MT, this method uses the difference in gradient polarity between the values of successive samples to detect peaks in the waveform [62], [63]. Essentially, the process runs through all samples in the waveform sequentially, calculating the gradient using equation (4-1) between each sample. By comparing susccessive gradients, a peak is identified as the point at which the polarity of the gradient changes as represented in equation (4-2), i.e. from positive to negative, or vice versa as illustrated in Figure 4-3 (P1).

$$m_n = \frac{\Delta y}{\Delta x} \tag{4-1}$$

$$m_{n-1} + m_n \neq |m_{n-1}| + |m_n| \tag{4-2}$$

Where $\boldsymbol{m_n}$ is the gradient, $\boldsymbol{m_{n-1}}$ is the previous gradient, $\boldsymbol{\Delta y}$ is the difference between the samples, and $\boldsymbol{\Delta x}$ is the change on the x-axis, which will always have a value of 1 in this case, since the comparison is done between two subsequent samples, hence:

$$m_n = \Delta y \tag{4-3}$$

Pseudocode: DGP peak detection

```
1.   initialise 'output_buffer' with values equal to 0
2.   FOR each sample 'S' in 'input_buffer'
3.       calculate current gradient (with previous sample)
4.       IF current gradient is not equal to previous gradient
5.           IF gradient is greater than or equal to 0
6.               set value at index of 'output_buffer' to -1
7.           ELSE
8.               set value at index of 'output_buffer' to 1
9.       update previous gradient equal to current gradient
```

The pseudocode above shows the program flow for detecting peaks in a signal using the DGP technique.

The purpose of the above implementation is to generate a buffer (from either dL or dA sample set) with a value of either 0, 1 or -1, where 1 represents a positive peak and -1 represents a negative peak. The generated *output buffer* is created with a size equal to that of the *input buffer* to be analysed, with all values in the *output buffer* initialised to 0 to begin with (line 1). For every sample in the *input buffer* a gradient is calculated between itself and the previous sample in the buffer. This gradient is then compared with the previously calculated gradient, and a difference between the current and previously calculated gradient suggests a change in direction, indicating a peak (line 4). For instances where the compared gradients are different, a change in gradient from negative to positive and vice versa reflects a negative peak and positive peak respectively (lines 5 – 8). Finally, line 9 updates the value of the previously calculated gradient to that of the current gradient to be used in the next iteration of the loop.

For both MT and DGP methods, the presence of high frequency noise components on the waveform will affect the detection process and cause false peaks to be detected, especially if there is a considerable difference in frequency between the main and noise signal. Running a low pass filter (with a suitable cutoff frequency) on the waveform pre-analysis should minimise this effect.

4.2.2 Peak Classification

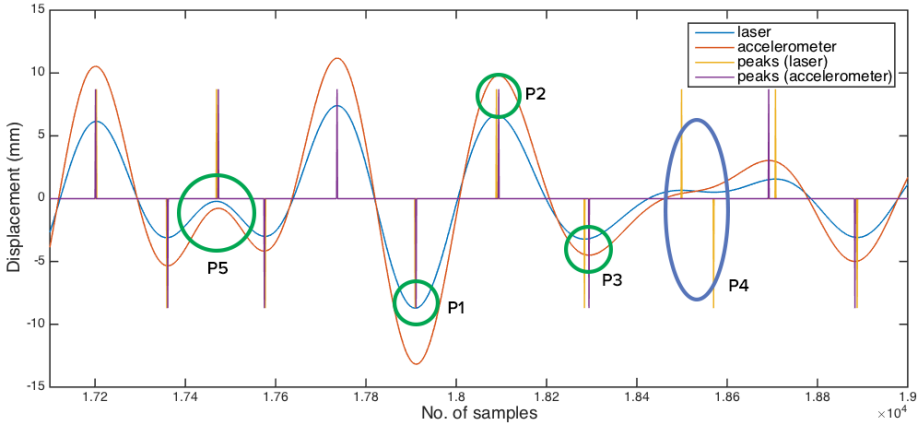

Figure 4-4 Waveform describing peak classification terminologies

Detected peaks are classified into two categories:

4.2.2.1 Common peak (CP)

These are peaks present in both dA and dL waveform for any given wavelength, for example P1 in Figure 4-4. These peaks imply a correlation between both signals, and indicates the point of interest in the HWPM algorithm described in this chapter. CPs suggests periods where the vehicle chassis is displaced vertically due to its dampers, road bumps etc. In which case the algorithm assumes an equal displacement is expected in both dA and dL samples.

4.2.2.2 Uncommon peak (UP)

These are peaks on any given wavelength present in only one of the samples i.e. either dA or dL as opposed to CPs. When present in dA samples, this typically suggests noise due to low frequency component of the acceleration samples. Whereas, if present in dL samples, this indicates longitudinal surface irregularities, which is one of two key areas of interest in the HWPM algorithm (mentioned in section 4.3).
The pseudocode below is an illustration of the method used to identify and isolate the CPs after the peak detection procedure in section 4.2.1.

Pseudocode: CP identification

```
1.   FOR each sample ‘A’ and ‘B’ in ‘buffer_A’ and ‘buffer_B’
2.       IF ‘A’ is not equal to 0
3.           set ‘peakA_temp’ equal to ‘A’
4.           IF ‘peakA_temp value’ is equal to ‘peakB_temp_value’
5.               add ‘peakA_temp’ to ‘CP_buffer_A’
6.               add ‘peakB_temp’ to ‘CP_buffer_B’
7.       IF ‘B’ is not equal to 0
8.           set ‘peakB_temp’ equal to ‘B’
9.           IF ‘peakB_temp_value’ is equal to ‘peakA_temp_value’
10.              add ‘peakB_temp’ to ‘CP_buffer_B’
11.              add ‘peakA_temp’ to ‘CP_buffer_A’
```

Essentially, the code scans through the buffers (*buffer_A* and *buffer_B* in line 1) generated from the peak detection process, which holds the sample index for all discovered CPs and UPs in both the reference and active signals. Since peaks are represented as either a 1 or -1 to indicate buffer index locations of positive and negative peaks, the loop checks for values in the buffer not equal to zero (lines 2 and 7). When a non-zero number is identified in the buffer, this value is stored in a temporary location (*peakA_temp* and *peakB_temp* for *buffer_A* and *buffer_B* respectively), which is then compared against each other (lines 4 and 9). If the values of both temp values are the

same, this indicates a common peak, and the index of the temporary values are stored in their respective buffers (CP_buffer_A and CP_buffer_B). Cases where the temporary values do not match indicate a UP, which is ignored by the program.

4.2.3 Peak Matching

After peak classification, the next phase of the HWPM is matching/scaling detected CPs on the active signal using the reference signal. The matching is done in half wavelengths on the perimeter of both sides of the peak point as shown in Figure 4-5.

For regular periodic signals, this process is done by calculating the wavelength of one cycle of the signal, and performing matching process at set intervals on exactly half of the calculated wavelength. However, for waveforms as shown in Figure 4-5, using this approach will result in false calculations, as it is irregular, and contains multiple signals with different wavelengths.

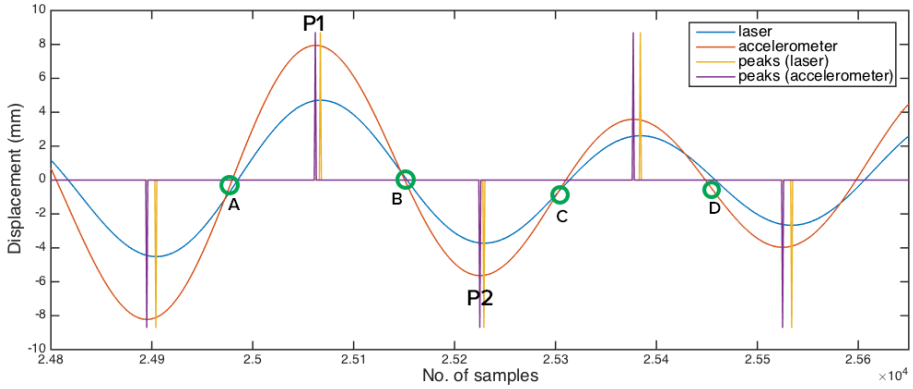

Figure 4-5 Waveform showing half-wavelength intersections

The method used in the HWPM calculates the midpoint between each opposite peak (in half wavelengths) on the active signal using equations (4-4) and (4-5). The calculated midpoint is based on the index on the x-axis and not the actual value of the sample (illustrated in Figure 4-5 as A, B, C and D). This establishes a quarter wavelength with a length between the peak point and calculated midpoint, and both quarter wavelengths on either side of the peak position forms the half wavelength to be corrected.

$$LM = \left\lceil \frac{x(t)_p + x(t)_{p-1}}{2} \right\rceil \tag{4-4}$$

$$RM = \left\lfloor \frac{x(t)_{p+1} + x(t)_p}{2} \right\rfloor \tag{4-5}$$

Where $\boldsymbol{LM}$ is the midpoint to the left of peak (point A in Figure 5), $\boldsymbol{RM}$ is the midpoint to the right of peak (point B in Figure 5), $\boldsymbol{x(t)_p}$ is the current peak's x-axis index, $\boldsymbol{x(t)_{p-1}}$ is the previous peak's value to the left of $\boldsymbol{x(t)_p}$, and $\boldsymbol{x(t)_{p+1}}$ is the next peak's value to the right of $\boldsymbol{x(t)_p}$.

When the half wavelength to be corrected is defined, the perimeter around the CP on the active signal is adjusted based on equation (4-6). The equation linearly scales the target samples within an acceptable minimum and maximum range determined by the CP of the reference signal.

$$f(t) = ya_{min} + \frac{(yb_{max} - yb_{min}) * (ya(t) - ya_{min})}{(ya_{max} - ya_{min})} \tag{4-6}$$

Where $\boldsymbol{ya(t)}$ is the amplitude of the active signal with respect to time, $\boldsymbol{ya_{min}}$ is the minimum value of the active signal's quarter wavelength, $\boldsymbol{ya_{max}}$ is the maximum value of the active signal's quarter wavelength, $\boldsymbol{yb_{max}}$ is the maximum value the active signal can be scaled to, and $\boldsymbol{yb_{min}}$ is the minimum value the active signal can be scaled to.

The values used in the parameters in equation (4-6) differ depending on the part of the wavelength that is being corrected. This is done in four sections:

- Positive peak, left quarter wavelength

This is the perimeter between points A and P1, in this instance $\boldsymbol{ya_{min}}$ and $\boldsymbol{yb_{min}}$ is the value at midpoint A, $\boldsymbol{ya_{max}}$ is the peak value P1 of the active signal, and $\boldsymbol{yb_{max}}$ is the peak of the reference signal.

- Positive peak, right quarter wavelength

This is the perimeter between points P1 and B, in this instance $\boldsymbol{ya_{min}}$ and $\boldsymbol{yb_{min}}$ is the value at midpoint B, $\boldsymbol{ya_{max}}$ is the peak value P1 of the active signal, and $\boldsymbol{yb_{max}}$ is the peak of the reference signal.

- Negative peak, left quarter wavelength

This is the perimeter between points B and P2, in this instance $\boldsymbol{ya_{max}}$ and $\boldsymbol{yb_{max}}$ is the value at midpoint B, $\boldsymbol{ya_{min}}$ is the peak value P2 of the active signal, and $\boldsymbol{yb_{min}}$ is the peak of the reference signal.

- Negative peak, right quarter wavelength

This is the perimeter between points P2 and C, in this instance $\boldsymbol{ya_{max}}$ and $\boldsymbol{yb_{max}}$ is the value at midpoint C, $\boldsymbol{ya_{min}}$ is the peak value P2 of the active signal, and $\boldsymbol{yb_{min}}$ is the peak of the reference signal.

4.2.4 Edge smoothing

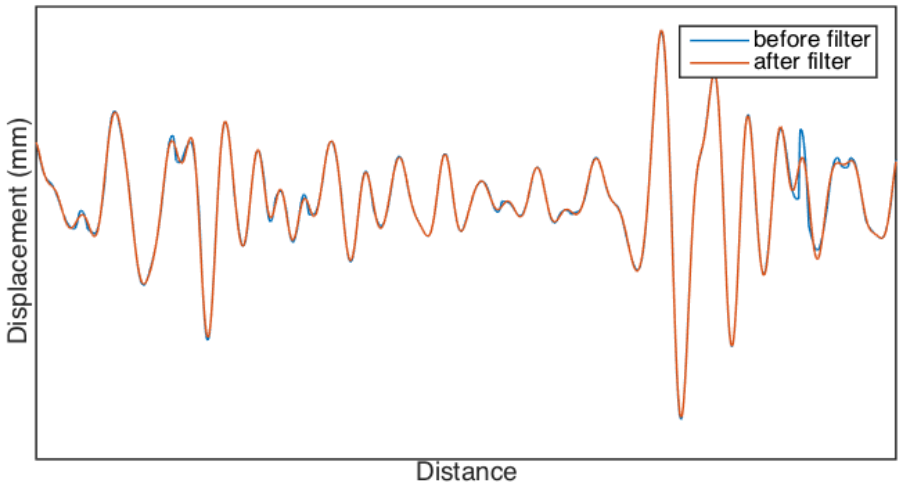

Figure 4-6 Effect of low pass filter on compensated waveform

The peak matching process causes slight discontinuity on the waveform as shown in Figure 4-6, which primarily occurs at the mid-points between the peaks. This is expected since the scale factor for each half wavelength is only dependent on its CP difference, hence, the midpoint at which they join show features of discontinuity. The purpose of this part of the algorithm is to eliminate this property, giving the waveform a smooth transition between the corrected half wavelengths. The edge smoothing is accomplished using a low pass filter with an acceptable cut-off frequency (4Hz in this application).

4.3 Experimental Results and Analysis

The results shown in this section reflect post analysis done on the samples acquired from the LBRM unit. All samples studied where obtained with the LBRM unit mounted on the rear bumper of the vehicle, and aligned to the right tire as shown in Figure 4-7.

Figure 4-7 Test setup of LBRM device mounted on vehicle

The decision on sampling rates and vehicle speed was chosen based on parameters that will yield the most precise and accurate analysis of the road texture. For these results, a sampling rate of 625Hz was chosen purely due to convenience. Since the analysis is targeted towards evaluating the surface mega texture and unevenness, sampling at the stated rate is adequate without compromising the profile accuracy.

Another factor that determines the texture measurement range is the speed of the vehicle, increasing the speed increases the texture wavelength that can be measured. According to the ISO 13473-5 standard [6] that characterises pavement textures by use of surface profiles, the wavelength of mega textures and unevenness is between 50mm – 0.5m and 0.5m – 50m respectively. To enable measurements within this range, the maximum speed required by the vehicle is approximately 15m/s (30mph) for mega

textures, and 150m/s (300mph) for unevenness (following Nyquist equation). Measurements were taken at various speeds from 20mph to 50mph, which allows for road texture investigation within 14mm – 22m wavelength range (using equation (4-7)).

$$Road\ texture\ measurement\ range = V * \frac{n}{fs} \tag{4-7}$$

Where $\boldsymbol{v}$ is the vehicle speed, $\boldsymbol{fs}$ is the sampling rate, and $\{\boldsymbol{n} \in \mathbb{R} \mid 1 \leq \boldsymbol{n} \leq fs\}$ which is used to calculate the possible measurement range where a value of $\boldsymbol{n} = \mathbf{1}$ and $\boldsymbol{n} = \boldsymbol{fs}$ are the minimum and maximum (respectively) road textures that can be evaluated.

The purpose of the accelerometer in laser based roughness measurement devices is to eliminate the low frequency displacements (caused by the vertical motion of the vehicle chassis) from the measured laser samples. First, the acceleration samples are converted to displacement via double integration, and then this calculated displacement is subtracted from the laser samples. If the LBRM unit is continuously displaced vertically creating a sinusoidal waveform as shown in the image, both laser and accelerometer samples are expected to show the same displacement waveform. In an ideal environment, subtracting dA from dL should produce a DC signal at 0.

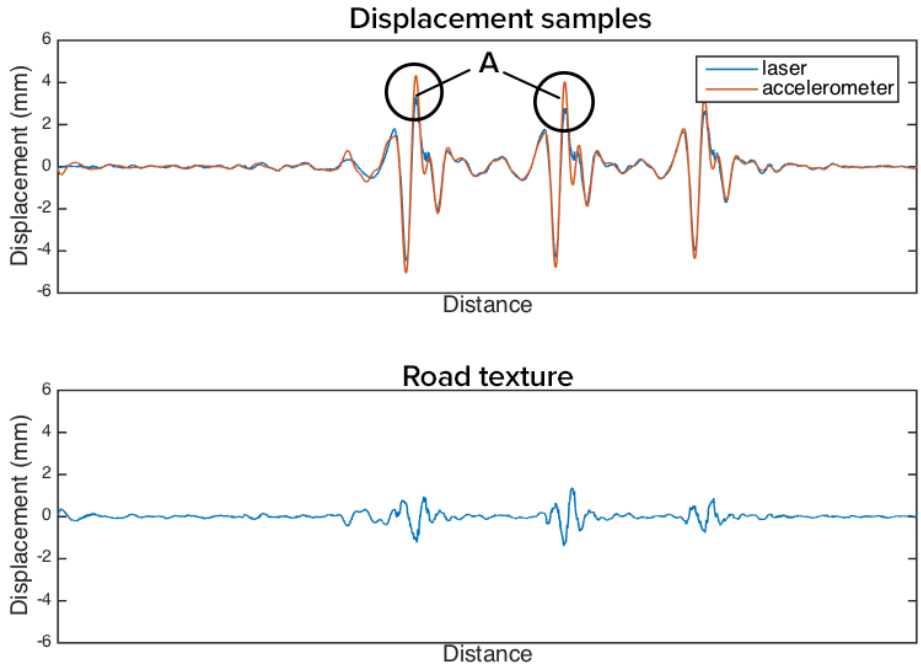

Figure 4-8 Effect of drift on evaluated road texture

In practice, this cannot be obtained as there are several factors affecting the accuracy of the measured samples. These include device tilt angle caused by vehicle dynamics, road surface reflection affecting laser sensor measurement, integration error caused by converting from acceleration to displacement etc. There are several methods to minimise the effect of these error sources, which have been discussed in Chapters 2. Figure 4-8 shows the actual waveform (with the corresponding road texture) derived from displacing the LBRM unit vertically to generate a sinusoidal waveform. Although there are similarities in the low frequency components of both dL and dA samples, there is a difference between their respective peak amplitudes, and in some cases, phase. The goal of the HWPM algorithm is to detect and correct this peak amplitude difference to derive a more accurate road texture representation.

The two key scenarios that determine and longitudinal irregularity with this form of road analysis are differences in amplitude between dL and dA samples ('A' in Figure 4-8), and phase shifts between wavelengths. Studies on the waveforms reveal that irregularities which show up as a difference in amplitude are more likely to be expressed as UPs when comparing the samples, while ones that appear as phase shifts in the wavelength are expressed as CPs.

4.3.1 Frequency analysis of laser and accelerometer samples

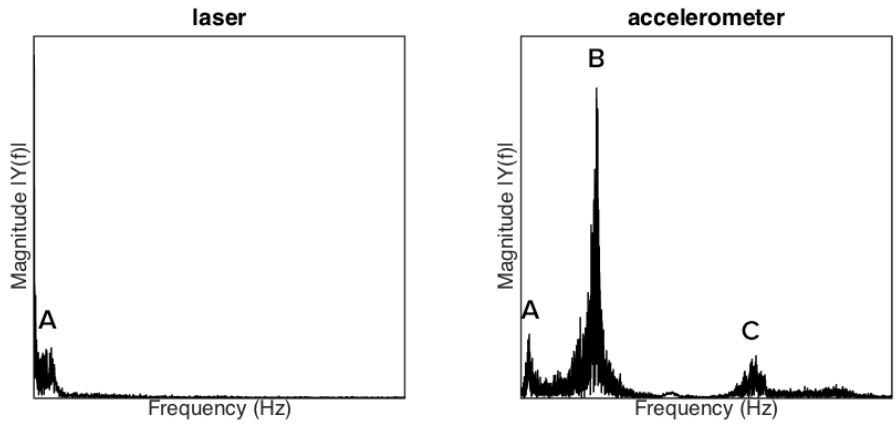

Figure 4-9 FFT representations of laser and accelerometer samples

Figure 4-9 shows the FFT of a typical sample set (for both laser and accelerometer sensors) acquired from the LBRM device after measuring the road surface texture. The laser sensor shows the presence of a single frequency band between $0 < f < 4$Hz, which reflects the frequency of the vehicle's vertical displacement caused by the longitudinal irregularities of the road, and this corresponds to the acceptable frequency range of sprung mass in commercial vehicles. Unlike the laser displacement samples, the

accelerometer (vertical axis) reveals three different frequency bands (A, B, and C as shown in Figure 4-9). Bands B and C are a result of higher frequency vibrations acting on the chassis, likely caused by the engine or other vibration sources. The focus for this application is on band A, as it reflects the same oscillation properties as the laser sensor. Hence, a low pass filter is required on the acceleration samples to remove frequencies greater than band A.

4.3.2 Effect of a High pass filter

The high pass filter is particularly important for the acceleration samples. It removes the drift that is imposed on the evaluated dA signal, which occurs due to low frequency components in the acceleration samples.

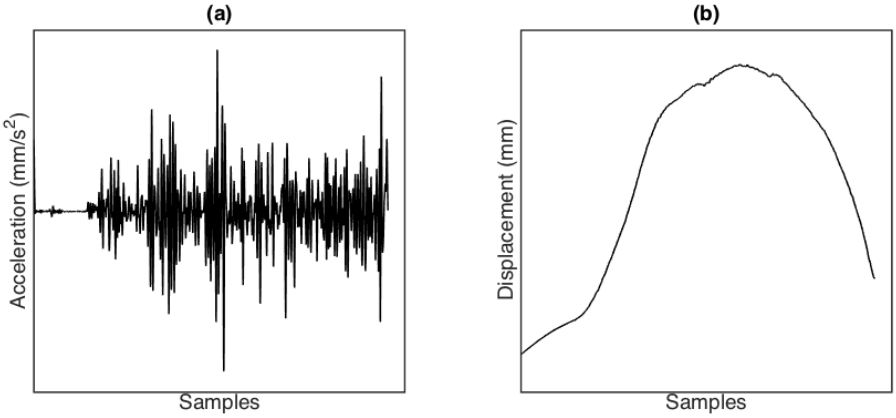

Figure 4-10 (a) acceleration samples low pass filtered (fc = 4Hz). (b) acceleration samples converted to displacement (without high pass filter)

Figure 4-10 shows the acceleration signal before (a) and after conversion to displacement (b). This acceleration signal in Figure 4-10a has been low pass filtered with a cutoff frequency of 4Hz to eliminate the unwanted frequency bands (B and C in Figure 4-9 FFT).

Figure 4-11 shows the effect of a high pass filter at different cutoff frequencies between 0.1Hz – 1Hz. The decision on selecting the right cutoff frequency is dependent on the application. In this instance, since 0 < signal frequency range < 4Hz, choosing a cutoff frequency outside these parameters or too far away from the lower limit will affect the profiling accuracy. In ‘a’ and ‘b’, when the cutoff frequency is set too low, the drift on the signal is still present (albeit reduced), masking the actual chassis vertical movement from the vehicle.

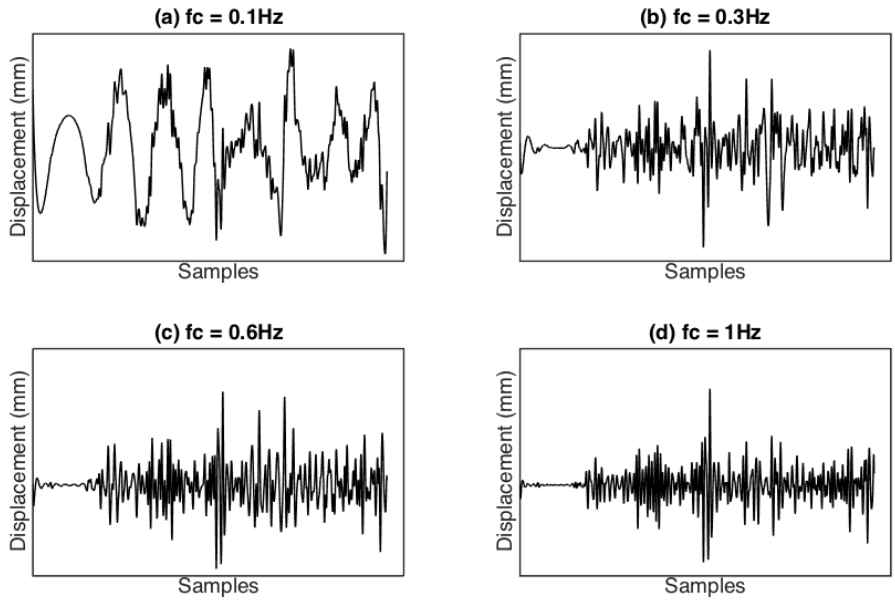

Figure 4-11 dA sample high pass filtered at 0.1Hz, 0.3Hz, 0.6Hz, and 1Hz cut-off frequencies

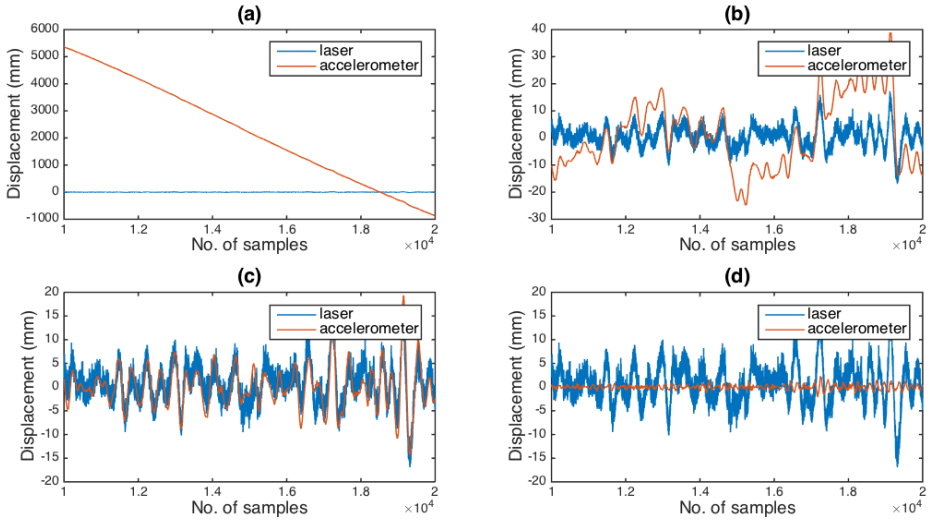

Figure 4-12 Comparing dA and dL at different cut-off frequencies. (a): fc = 0Hz, (b): 0.3Hz, (c): 1Hz, (d): 2Hz

As the cutoff frequency is increased, dA becomes identical to dL, because the filter eliminates more of the drift in dA, as shown in Figure 4-12. There is a limit to this increment, because the amplitude of the signals is attenuated at higher cutoff frequencies, hence, losing useful information as indicated in (d).

The cutoff frequency is based on the required road texture wavelength. Assuming a road texture wavelength of 10m, for a vehicle speed of 20m/s (45mph) the frequency of the texture wavelength is 2Hz. Therefore, the high-pass cutoff frequency should not exceed this value.

4.3.3 Effect of speed on the high pass cutoff frequencies

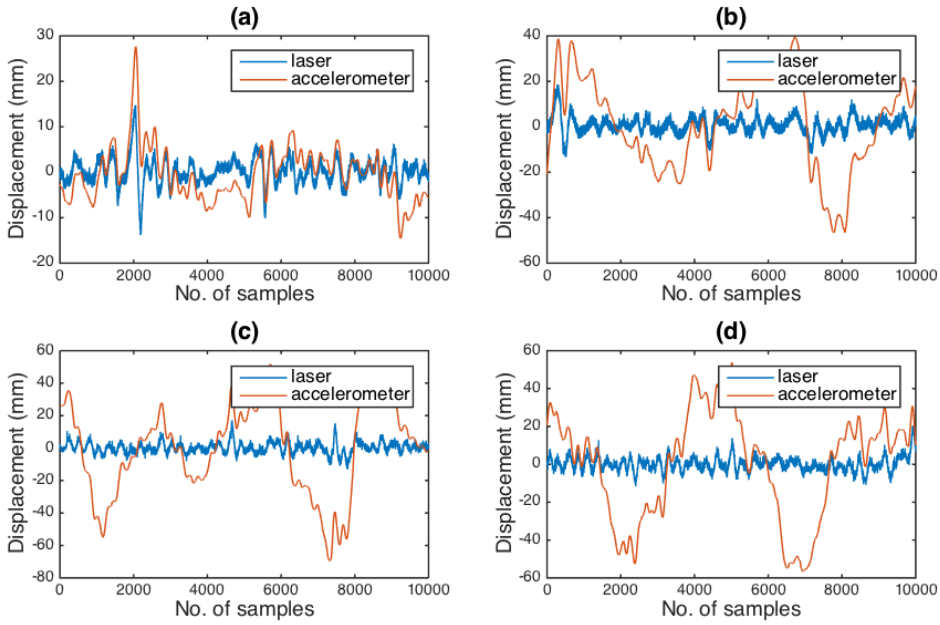

Figure 4-13 dL and dA samples high pass filtered with a cut-off frequency of 0.1Hz, showing the effect of speed on the dA drift. (a) v = 20mph, (b) v = 30mph, (c) v = 40mph, (d) v = 50mph.

Results gathered for the experiment suggests that the speed of the vehicle influences the amount of drift imposed on the dA signal after double integration. This is a direct relationship where the drift increases with speed, and requires a higher cutoff frequency from the high pass filter to eliminate it.

Figure 4-13 illustrates this theory. Comparing the waveforms obtained at the different speeds, the drift on the dA signal increases progressively as the speed rises from 20mph to 50mph, and this is more evident when the cutoff frequency is at 0.1Hz. This is mainly due to the difference in wavelengths occurring at these speeds. Ultimately, the frequency sensed by the vehicle chassis is limited because of the dampers, and although this frequency range is consistent, the equivalent wavelength changes because it is dependent on the speed of the vehicle. On the waveform, a wavelength with a frequency of 1Hz will have an equivalent travel distance of 8.9m at a vehicle speed of 20mph

(8.9m/s), and 22.4m when the vehicle is travelling at 50mph (22.3m/s). The low frequency component of the waveform equates to a larger distance at higher speeds, hence, the increase in drift after integration.

4.3.4 Effect of HWPM on Drift

The plots (Figure 4-14 – Figure 4-17) demonstrates the purpose of the HWPM algorithm in correcting drift in the dA signal. The samples were obtained for four different speeds at 20mph, 30mph, 40mph and 50mph respectively. Each image has two sets of data (separated by a column) showing the result of the evaluated dA waveform before and after applying the HWPM algorithm. Note that the profile samples were taken at different locations.

As the cutoff frequency increases, the dA waveform becomes progressively equivalent to its dL counterpart irrespective of the HWPM. However, comparing both sets of results for the various speeds, better drift correction is achieved with the HWPM.

Without the HWPM, a cutoff frequency of at least 1Hz is typically required to sufficiently eliminate any drifts in the waveform (as shown in the plots), especially in higher vehicle speeds. Whereas, with the HWPM the same level of drift elimination is achieved at half (50%) the cutoff frequency. The HWPM reduces the amount of low frequency signal that is ignored by the high pass filter that could potentially contain useful profiling information.

This is better observed in low speed conditions. Figure 4-18 shows this data loss, where the vehicle is used to sample a road hump at 10mph. Comparing the waveforms in the image, there is a significant change at 'x' which depicts the hump. As the cutoff frequency increases, the filter progressively alters the representation of the hump. This is because the hump produces a very low frequency signal, which is increasingly ignored by the filter, as the cutoff frequency gets higher.

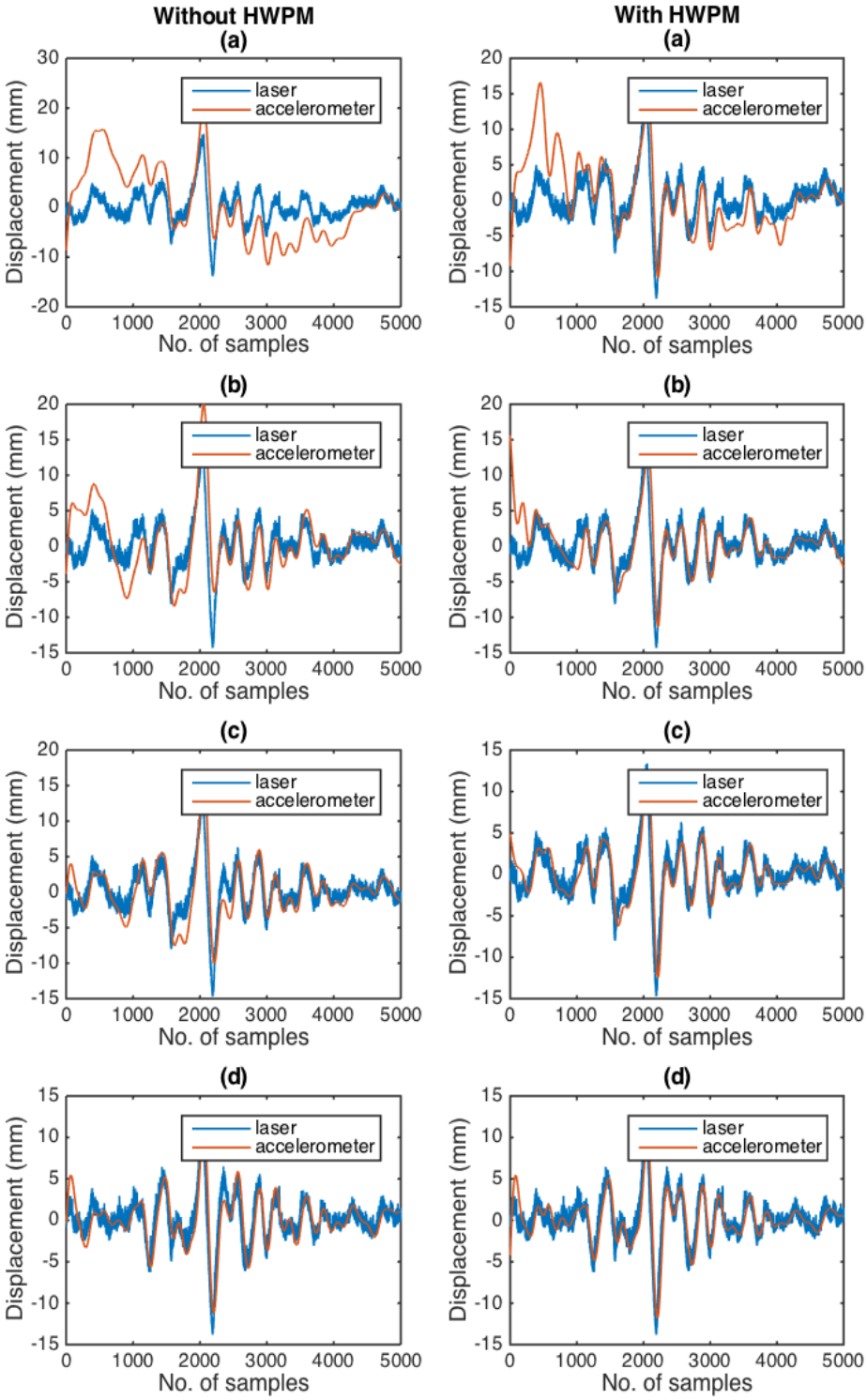

Figure 4-14 dL and dA samples measured at 20mph to evaluate the drift correction done by HWPM at different cut-off frequencies. (a) fc = 0.1Hz, (b) fc = 0.3Hz, (c) fc = 0.5Hz, (d) fc = 1Hz.

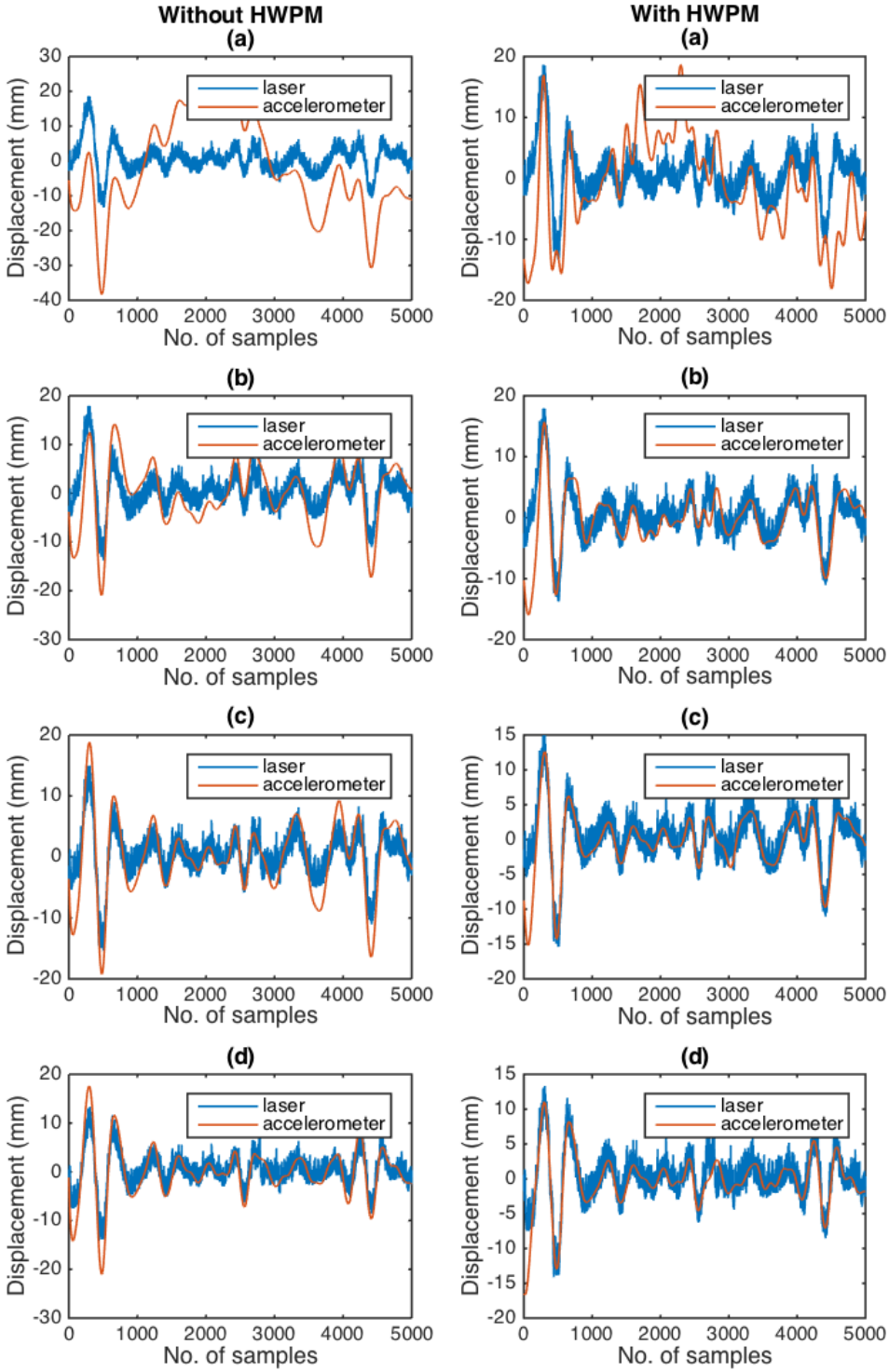

Figure 4-15 dL and dA samples measured at 30mph to evaluate the drift correction done by HWPM at different cut-off frequencies. (a) fc = 0.1Hz, (b) fc = 0.3Hz, (c) fc = 0.5Hz, (d) fc = 1Hz.

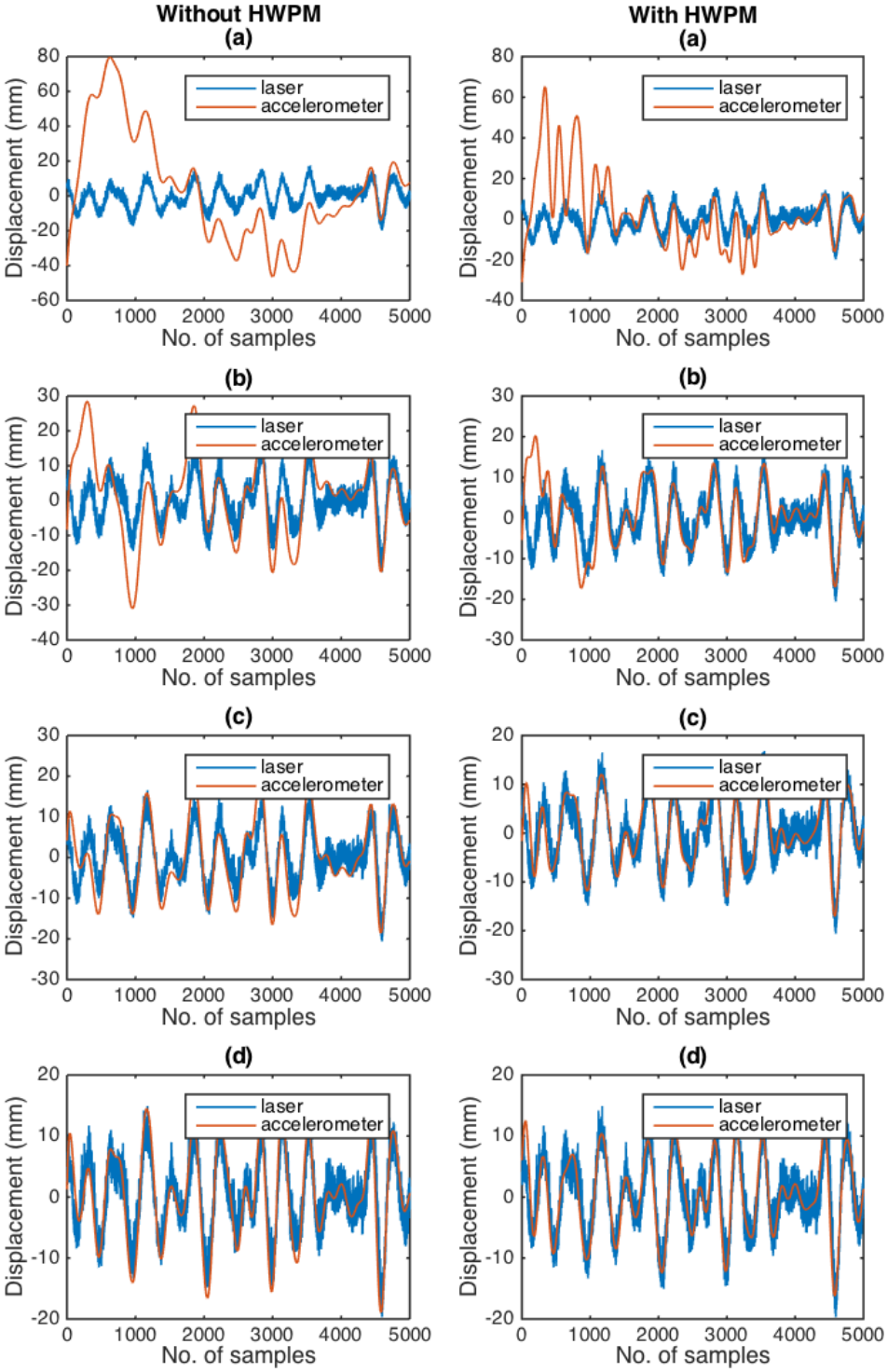

Figure 4-16 dL and dA samples measured at 40mph to evaluate the drift correction done by HWPM at different cut-off frequencies. (a) fc = 0.1Hz, (b) fc = 0.3Hz, (c) fc = 0.5Hz, (d) fc = 1Hz.

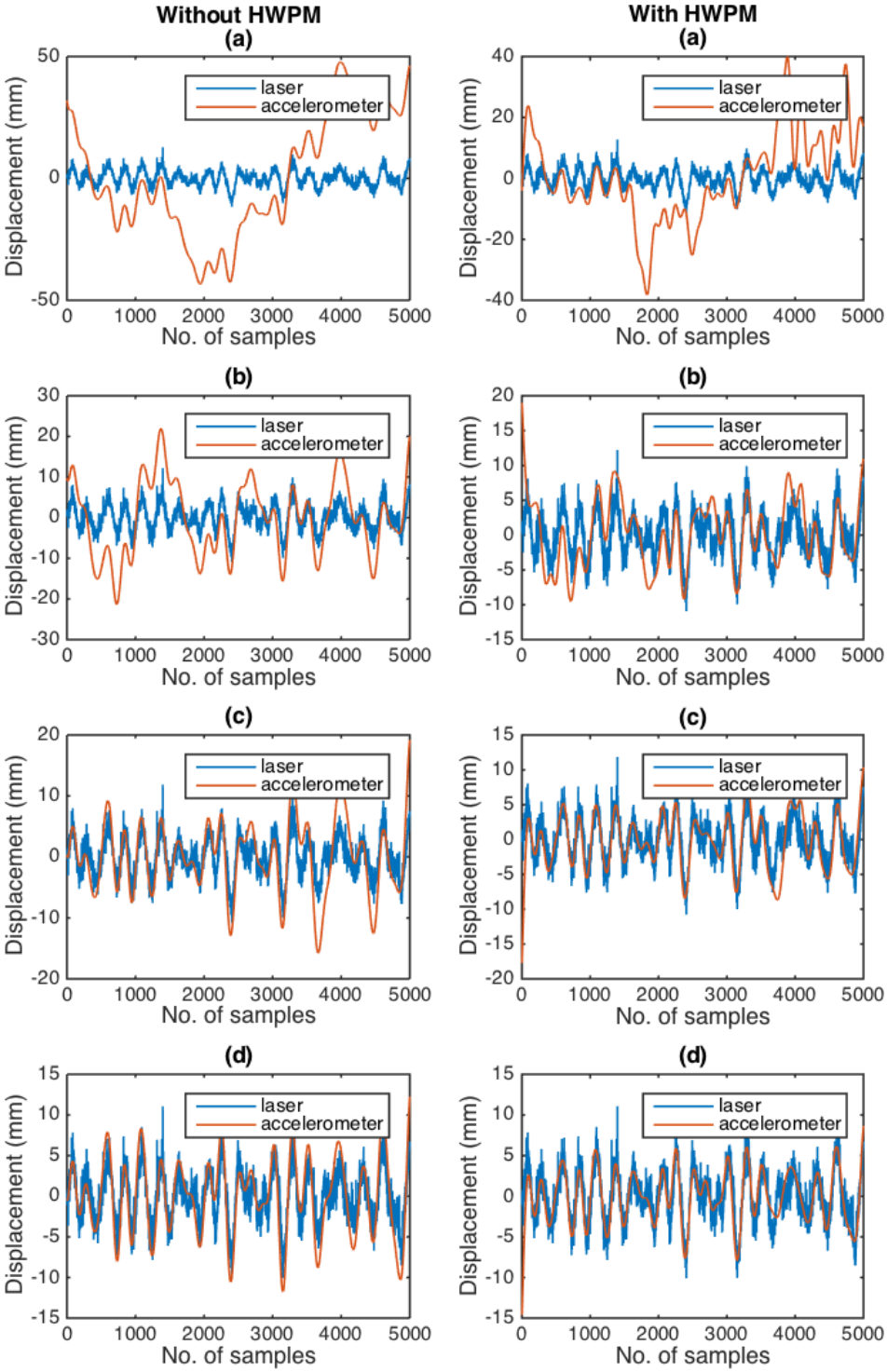

Figure 4-17 dL and dA samples measured at 50mph to evaluate the drift correction done by HWPM at different cut-off frequencies. (a) fc = 0.1Hz, (b) fc = 0.3Hz, (c) fc = 0.5Hz, (d) fc = 1Hz.

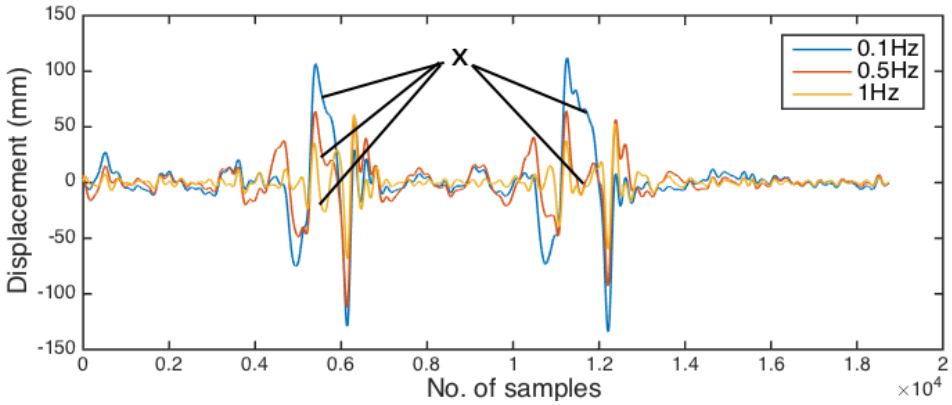

Figure 4-18 dL sample of road humps.

Since the laser signal is used as the reference for the HWPM algorithm in this application, it is vital to maintain the originality of this reference signal as best as possible, because inaccuracies are translated to the active signal in HWPM.

For laser based road profilers, it is important to select an appropriate high pass cutoff frequency for the dA signal. Because of the longitudinal irregularity thresholds defined in the series 700 road pavement standard (Table 1), an incorrect cutoff frequency could cause a pavement to be evaluated above or below the outlined thresholds.

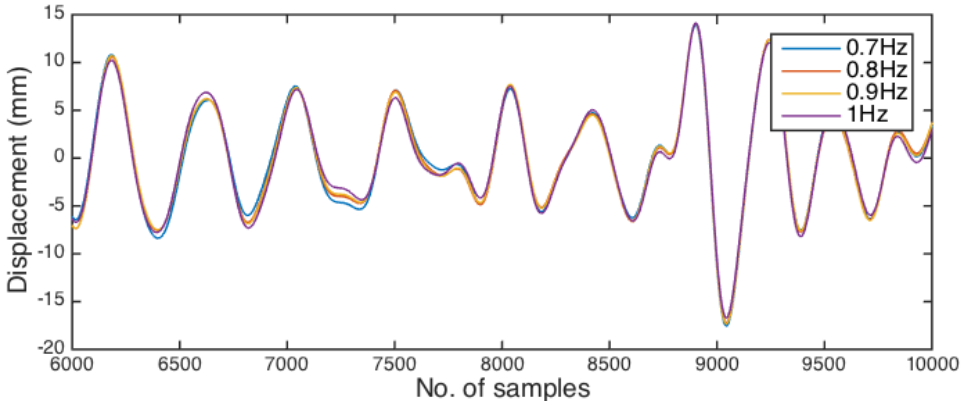

Figure 4-19 dL sample after high pass cut-off frequency between 0.7Hz – 1Hz.

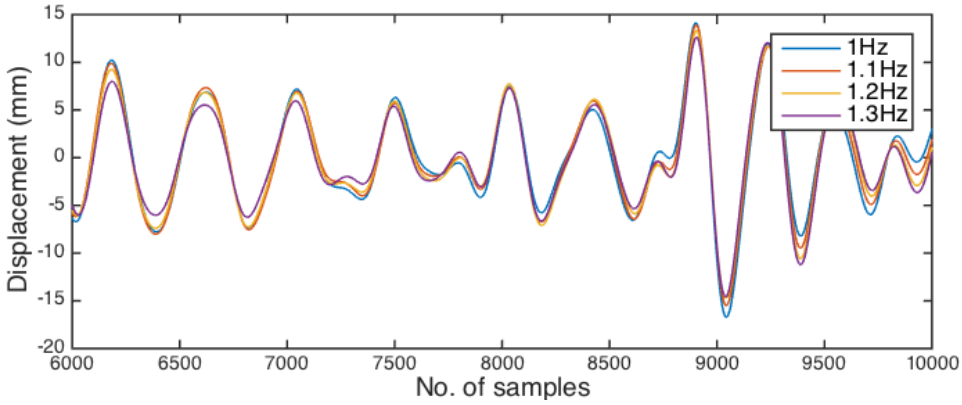

Figure 4-20 dL sample after high pass cut-off frequency between 1Hz – 1.3Hz.

Figure 4-19 and Figure 4-20 illustrates the low frequency component of the laser displacement sensor at high pass cutoff frequencies between 0.7Hz – 1.3Hz. The purpose of the comparison is to justify the appropriate cutoff frequency to be used in profiling, which should validate the expected frequency range of the vehicle's sprung mass.

Since the dL is the reference signal for the HWPM in this application, the goal is to find the highest cutoff frequency at which the dL signal begins to show significant change in its waveform. Comparing both images, significant transformation on the signal is more evident as the cutoff frequency is greater than 1Hz, which corresponds with the frequency range of a vehicle's sprung mass.

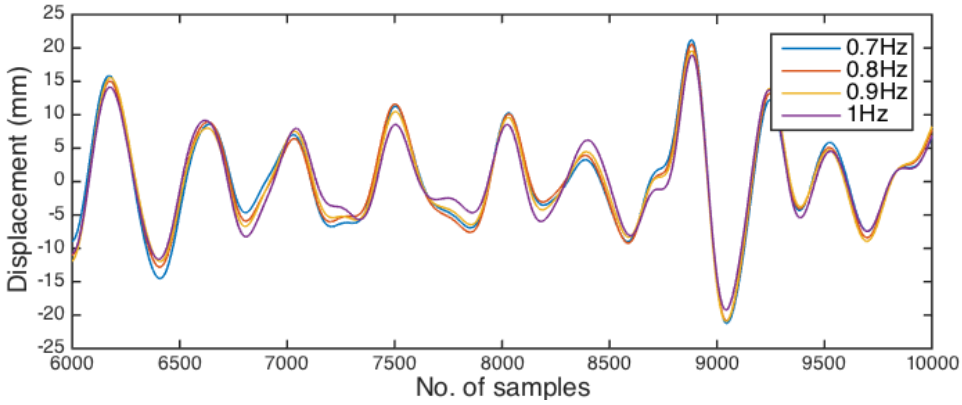

Figure 4-21 dA sample after high pass cut-off frequency between 0.7Hz – 1Hz.

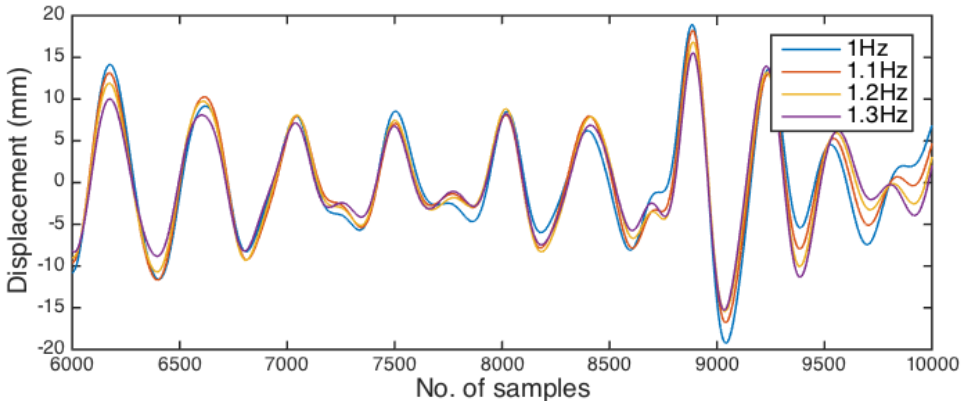

Figure 4-22 dA sample after high pass cut-off frequency between 1Hz – 1.3Hz.

For the dA samples as shown in Figure 4-21 and Figure 4-22, the same interpretation is observed where significant transformation to the waveform occurs, as the cutoff frequency gets higher than 1Hz.

HWPM minimises the distortion effect of the high pass filter on the dA samples. This allows consistency in profile results around a range of frequencies as shown in Figure 4-23 and Figure 4-24. From the HWPM plots, the waveform for the dA signal is maintained even for cutoff frequencies slightly greater than 1Hz.

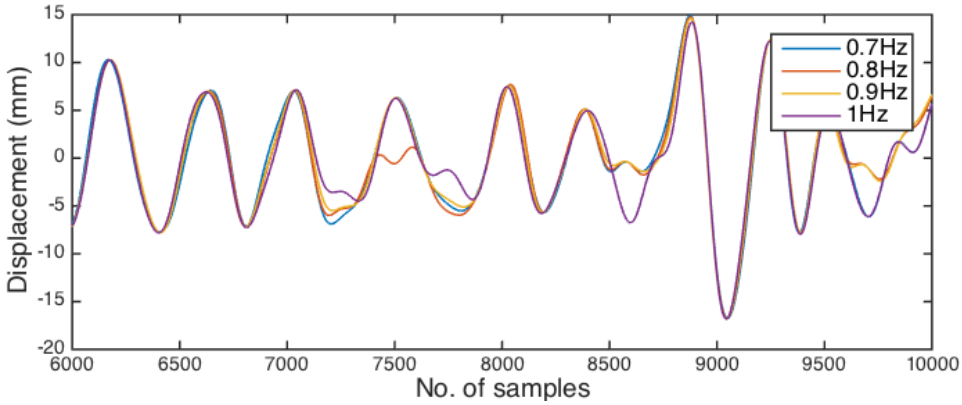

Figure 4-23 HWPM sample after high pass cut-off frequency between 0.7Hz – 1Hz.

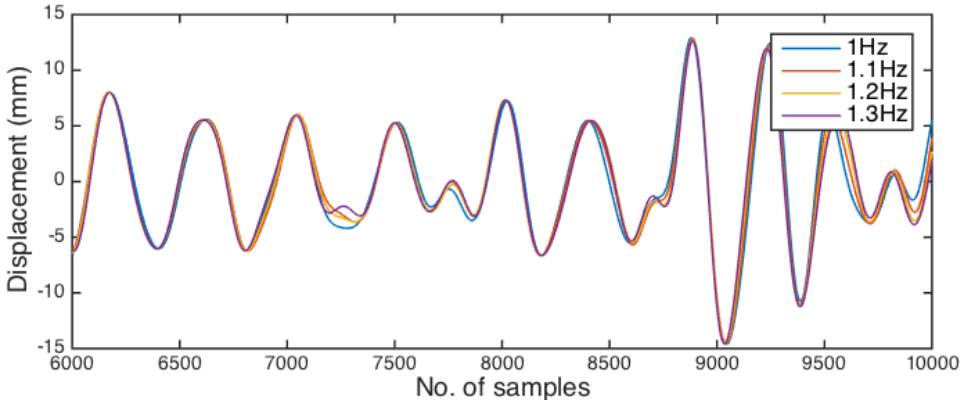

Figure 4-24 HWPM sample after high pass cut-off frequency between 1Hz – 1.3Hz

4.3.5 Comparing the evaluated road profile against a 3m Rolling Straight Edge

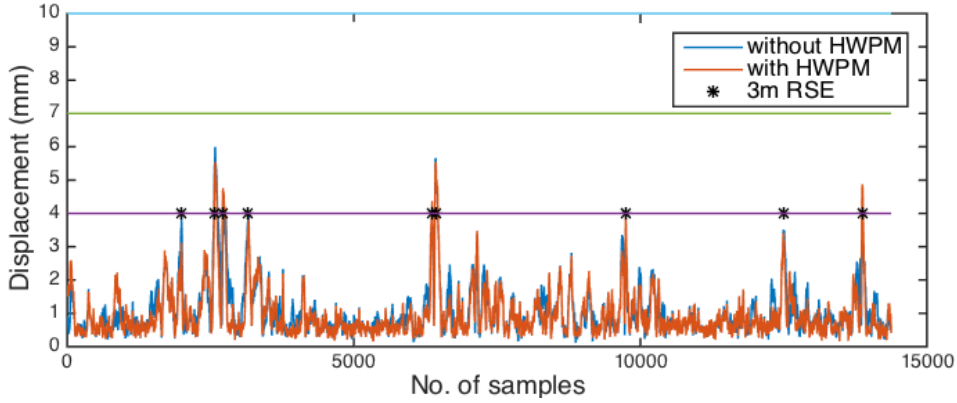

Figure 4-25 Longitudinal profile at 20mph (run 1)

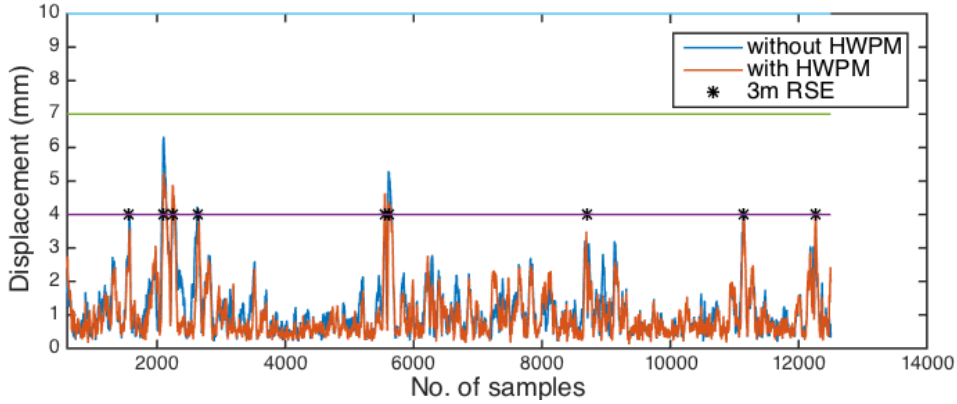

Figure 4-26 Longitudinal profile at 20mph (run 2)

The longitudinal profile obtained using the LBRM unit is compared with the results of a 3m rolling straight edge (RSE). The profile shown is the absolute value of the difference between dL and dA. The results illustrate the improvements made by the HWPM algorithm.

Before comparing the results between the LBRM unit and 3m RSE, both sets of profiles had to be viewed under similar characteristics. This is particularly important since the LBRM unit uses a laser, which has a precision diameter of approximately 1mm to measure the pavement irregularities, compared to the RSE, which uses a wheel with a diameter of 12cm. The evaluated profile data from the LBRM device was filtered to simulate the characteristics of a 12cm wheel running through the pavement. This was done by applying a moving average to the samples with a window equivalent to the wheel diameter. Also, the profiles from the LBRM device uses a cutoff frequency of 1Hz for both laser and acceleration samples, and the justification for this value is explained in the previous section.

Figure 4-25 and Figure 4-26 shows the profile of a pavement taken at separate times to validate the consistency and accuracy of the LBRM unit and HWPM algorithm. The measurements taken with the 3m RSE where rounded to the lowest threshold i.e. 4mm, 7mm and 10mm, according to the standard. For example, an irregularity of 6mm, 8mm and 12mm measured with the RSE will be indicated on the plot at the 4mm, 7mm and 10mm mark respectively. All measurements lower than 4mm are not plotted.

The RSE is used to validate the profile generated by the LBRM device, and the evaluated profiles verify the consistency and accuracy in the profiling technique, especially for irregularities that cross the threshold marks.

The effect of the HWPM is more apparent in Figure 4-27, where the measurement is taken at a higher speed of 50mph. As discussed in section 5.3.4, the drift imposed on the accelerometer samples after conversion from acceleration to displacement increases as the vehicle speed increases. This requires a higher cutoff frequency from the high pass filter to correct the signal, but this has a consequence on data loss. The pre HWPM profile show in Figure 4-27 indicate higher sensitivity at this speed, which is corrected post HWPM. The corrected profile is validated against the RSE with consistent results.

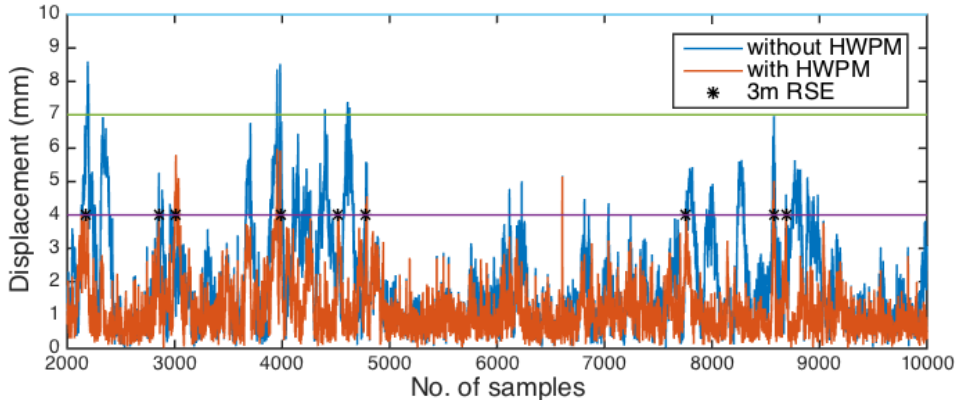

Figure 4-27 Longitudinal profile at 50mph

4.4 Summary

The chapter proposes a novel technique that improves the accuracy of vehicle-based road surface profilers, by minimizing the accumulated drift in the accelerometer's displacement signal. Drift from accelerometers are typically removed with high-pass filters with an appropriate cut-off frequency. Choosing an inappropriate cut-off frequency will result in loss of data (if too high), or inaccuracy due to drift (if too low). Being able to determine the ideal cut-off frequency in this application is a challenge, because as demonstrated in Section 4.3.3, the vehicle speed at which the measurements are taken influences the amount of drift imposed in the displacement signal derived from the accelerometer. Where a higher speed (around 50mph) means an increase in drift imposed on the signal, and vice versa. Which in turn means a higher cut-off frequency is required to sufficiently remove drifts on the signal at higher vehicle speeds, compared to lower speeds. This requires a model to estimate the cut-off frequency required, depending on the speed of the vehicle. This could be based on the mathematical relationship, or estimated preset values taken from multiple tests. The propose Half-Wavelength Peak Matching (HWPM) model minimises the drift on the accelerometer's samples by comparing it with measurements taken from the laser displacement sensor. The model estimates a correction offset on the accelerometer samples based on common wavelengths between both sensors as described in Section 4.2.2. A high-pass filter is still required, but the advantage is, the drift compensation will be maintained for a band of frequencies around a chosen cut-off.

Tilt compensation, which is beyond the scope of this study, adjusts the displacement measured from the laser for times where the sensor is tilted. This could occur in scenarios where the front wheel of the vehicle is raised, creating an angle between the

laser beam and the road surface. The profiles discussed in Section 4.3.1 do not account for this tilt, because the aim of the chapter is to validate the accuracy of the HWPM model. A common approach to correcting the tilt error in the laser sensor, is attaching a gyroscope used to measure the tilt angle, then the compensation is calculated based on the measured angle.

The evaluated profile was benchmarked against a 3m Rolling Straight Edge (RSE) before and after applying the HWPM model to the profile. The results discussed in Section 4.3.5 was predominantly positive to HWPM, with significant differences observed at higher vehicle speeds. Since the laser sensor is used as a reference signal to compensate the drift in the accelerometer sample, it is vital to maintain its accuracy.

CHAPTER 5

VIBRATION MEASUREMENT BASED ON A MEMS ACCELEROMETER

5.1 Overview

This chapter describes an algorithm that achieves accurate and precise vibration measurements using MEMS accelerometers. The purpose for this in road analysis is to be able to design portable low cost vibration meters, which can be adapted and integrated with the LBRM device to evaluate road surface conditions based on the vibrations caused on the vehicle's unsprung mass (i.e. brakes, wheel, tyre etc.).

Pavement irregularities induces a force on the unsprung mass, caused by tyre compressions, and this determines the amount of vibration experienced. Shock absorbers dampen the effect of this vibration on the vehicle's chassis, improving overall ride comfort during transit. Since the LBRM device is mounted on the chassis, it is unable to sense these vibrations, hence the need for a separate vibration measurement unit, of a smaller form factor, mounted on the vehicle's unsprung mass, and working alongside the LBRM unit. This makes it easy to evaluate the effect of road surfaces on wheel vibrations, by comparing it with data obtained from the analysis described in Chapter CHAPTER 4.

The method described calculates the vibration velocity, which is the displacement per second, measured in mm/s. This estimate how much a body moves every second in the three axes (x, y and z), from a reference point. Severity of vibration is a factor of moved distance and frequency of the movement, hence why vibration velocity is best suited for detecting machine unbalance, looseness, harmonics etc. This is ideal in this application as opposed to vibration acceleration and vibration displacement, which are more suited to rotational and low frequency (< 5Hz) vibrations respectively.

Fundamentally, the basis for this theory relies on the equation of motion in equation (2–1), which describes the relationship between acceleration, velocity and displacement. According to this relationship, integrating acceleration samples generates velocity, while integrating velocity generates displacement (which essentially is the double integration of acceleration). Accelerometers form the basis of most vibration meters, because they are readily available, and theoretically, their samples can be converted to either velocity or displacement as mentioned earlier. Unlike piezoelectric accelerometers, commonly used in commercial meters (because they generally offer a higher measurement and frequency range), the MEMS capacitive accelerometers are relatively cheaper and offer good measurement accuracy for frequency ranges within the requirements of the application ($10 - 1000$Hz).

The chapter focuses on calculating the vibration velocity from an accelerometer, by processing the acceleration samples in the frequency domain and estimating the velocity in the same domain. The process eliminates the need to integrate the acceleration samples in the time domain, which typically requires a high pass filter to eliminate any DC offset induced on the calculated velocity. Common causes of this DC offset are due to noise on the measurement samples, or gravitational force acting on the sensor (especially in MEMS accelerometers). Conventional practice involves pre-processing the acceleration samples before integration, the purpose for this is to minimise the amount of noise, and eliminate any DC signals, because these are amplified after integration generating increased noise on the derived velocity.

Usual pre-processing methods include mean elimination, high pass filtering, and Kalman filtering. Mean elimination (which the model in the chapter employs) subtracts the calculated mean of the signal from each sample, and this is the quickest and most efficient method of removing any DC offset caused by gravitational force, mainly because it has no attenuation effect on the signal, but simply shifts its midpoint towards zero. High pass filtering also eliminates any DC signals present, but this attenuates signals lower than a specified cut-off frequency, only permitting frequencies higher than

this cut-off. The side effect to this method is that signals with frequencies very close to the cut-off are attenuated as well, but the amount of attenuation is relatively insignificant. Kalman filters work by estimating and adding a correction value to the measured samples, using a mathematical model derived from its relation to other sensors. This is comparatively complicated in terms of implementation, since the filter relies on data from other sensors (hence the need for additional sensors), because its correction accuracy depends on the number of related sensors that make up the mathematical model.

Common practice of obtaining vibration velocity using accelerometers integrates the acceleration samples, deriving a velocity waveform, then this waveform is then passed through several filters to minimise the effect of the error obtained from the integration process, before any vibration analysis is carried out on the post filtered signal. The method described in this chapter attempts to improve the accuracy of MEMS based vibration systems, by bypasses the effect of integration error on the results, calculating the velocity of the signal directly from the acceleration's frequency domain. This technique obtains the acceleration samples from the sensor, performs an FFT analysis on the acceleration, and evaluates the velocity based on this acceleration's frequency domain.

The rest of this chapter is split into three sub sections; Section 5.2 explains the novel approach to vibration calculation, which the chapter is primarily based on. Section 5.3 discusses the design and test process used to validate this novel technique, and then Section 5.4 show results obtained from the experimental tests, analyzing all relevant findings and comparing the data with other common practices, processes and methods.

5.2 DFT based velocity calculation with RSS approximation

This process is formed of four stages, from data collection to velocity approximation as shown below:

- Acceleration measurement (Data collection)
- Discrete Fourier Transform (DFT) analysis
- Acceleration to velocity conversion in the frequency domain
- Root Sum Squared (RSS) moving average

5.2.1 Acceleration measurement (Data collection)

Data collection is a critical aspect of any signal-processing algorithm, as this has a ripple effect to the rest of the analysis carried out on the signal. Inaccuracy at this stage will affect the results obtained in subsequent stages that form the algorithm. Depending on the nature of the signal, inaccuracies could be from the measured sample obtained from the sensor, or, in the case of continuous signals, the time interval between each measured sample (this is particularly important in applications where consistency between the measured samples is expected, especially in frequency analysis). Several factors affect the measurements from the sensor, producing incorrect samples, the most critical of these factors are:

5.2.1.1 Mounting

Incorrect sensor mounting is a usual source of measurement error, and this could be because of sensors not fitted properly in terms of position and location, or, for measurements involving the movement of a body, the sensor's tightness to that body. The tightness of the sensor's mount is particularly important in this application, since we intend to measure vibration. This tightness is directly proportional to the frequency at which the body is vibrating, because looseness has an increasing negative effect on the measurement accuracy as the vibrating frequency rises. The objective, especially in this application, is to eliminate the possibility of moving parts during measurement. Regarding mount position and location, since accelerometers measure acceleration at different axis (i.e. x, y and z-axis in a tri-axial sensor), the mount position is a major factor. Vibration occurring along one axis will not be reflected on the other two axes, and in more advanced situations, depending on the orientation of the sensor, the vibration may occur between two or all the axes. To evaluate the measured acceleration in a three-dimensional space, all three axes need to be considered. Section 5.3 describing the experimental setup, show the methods employed to minimise errors from mounting.

5.2.1.2 Sampling

For continuous data measurement, sampling is another source of error mainly caused by choosing inappropriate sampling rate, or inconsistency between the samples. The sampling rate used cannot exceed the maximum sampling rate stated for the selected sensor, because in best-case scenarios, this will cause duplicate values to be reflected for subsequent time intervals, potentially affecting any succeeding calculations. Several other factors need to be considered while selecting an appropriate sampling rate, these include, the measurement range desired, the FFT requirements for the application (for

applications that require analysis in the frequency domain as described in section 5.2.2), and the memory limits of the controller used for sampling. Consistency between each sample is also vital for applications requiring frequency analysis. It is important to maintain this sampling interval (in practice, this entails maintaining a minimal deviation across each sample interval for the signal to be processed). Three techniques can be used in sampling a sensor as shown in the pseudo code below.

Pseudocode: Technique A

```
1.   main code
2.       while (obtaining samples)
3.           read sample;
4.           delay;
```

Pseudocode: Technique B

```
1.   timer interrupt routine
2.       read sample;
3.
4.   main code
5.       setup timer interrupt;
```

Pseudocode: Technique C

```
1.   time interrupt routine
2.       sample now = true;
3.
4.   main code
5.       setup timer interrupt;
6.       while (obtaining samples)
7.            if (sample now is true)
8.                read sample;
9.                sample now = false;
```

Technique A (TA) is the simplest form of data sampling, which continuously reads samples from the sensor with a predefined delay (determined by the required sampling rate) between each read. The main advantage to this technique is its simplicity, which is sufficient for some applications, but for applications requiring consistent time interval between each sample, this method of sampling performs poorly. The reason for this is, any delay caused by the process of reading a sample (at line 3 in TA) will result in an additional delay for that interval, hence, the reason for the inconsistency between each sample read. The potential for delay in reading a sample from a sensor is a factor of the communication protocol between the controller and sensor, the controller's processing capability, and the sensors sampling speed. Individually, this delay is somewhat insignificant, but in most cases, can be pronounced when analysed cumulatively over an entire signal, as shown in Table 5-1.

Technique B (TB) tries to solve the problem of sampling inconsistency occurring in TA by using a timer feature, which is usually present in MCU's. The idea is to setup a timer function that constantly triggers an interrupt routine after a predefined time has elapsed, to read a sample from the sensor when the interrupt routine is executed. In theory, this method of sampling seems to solve the problem of inconsistency, because timers in MCUs are maintained in parallel with executed code, which means that the timer ticks are not affected by delays in the main code. This way, the interrupt is expected to be triggered on a consistent basis, at the predefined period selected. Although this method performs better than TA in terms of consistency of sampling, it is generally not good programming practice to execute long running commands in an interrupt routine, as most MCUs require the interrupt routine to be idle before triggering successive interrupts, hence, any potential delay in reading a sample could result in delayed triggering of the timer interrupt.

Technique C (TC), which combines TA and TB, utilises the advantages offered by both methods. It aims to solve the issue of running long running commands in the interrupt routine experienced in TB, by reading sensor samples on the main code based on a flag set from the interrupt routine. TC maintains the timer feature from TB, but rather than read samples on the interrupt routine, it sets a flag "sample now" (at line 2 in TC) which determines when a sample should be read, and this flag is monitored from the main code (at line 8 in TC). Setting a flag is one of the shortest command sets in an MCU, so it will have an insignificant delay to the interrupt routine. The main code monitors this flag, and when the flag is set from the interrupt routine, a new sample is read, and the flag is cleared (lines 9-10 in TC). The clearing of the flag is an important aspect of the process, because without that, the "sample now" flag is always

set which will cause the main code to read a sample from the sensor for every loop, ignoring the required timing. With the flag cleared after every read, the next read will only occur when it is set from the timer interrupt routine, this way, the sensor is read with regards to the timer.

For optimal performance of TB and TC, the sampling period needs to be greater than the maximum time required to obtain a sample from the sensor as shown in equation (5–1), else this will also result in inconsistent sampling.

$$T_s > DS_{max} \tag{5–1}$$

Where $\boldsymbol{T_s}$is the sampling period, and $\boldsymbol{DS_{max}}$is the maximum time taken to obtain a sample.

Table 5-1 Sampling performance (512 samples at 250Hz)

	TA	**TB**	**TC**
Total time taken (s)	2.092	2.048	2.048
Average time per sample (ms)	4.09	4	4

Table 5-1 shows the time taken to read 512 samples at a frequency of 250Hz for all three data capturing techniques. Theoretically, for a sampling rate of 250Hz, the period per sample is 4ms, which equates to 2048ms for 512 samples. Analyzing the results shown in the table, TB and TC show better performance to TA (as expected) in this application. Although TB and TC perform identically, TC is preferred based on programming practice as explained previously. TB is more likely to cause unexpected problems in applications where the time taken to read a sample is greater than timer's timeout interval, because it is essential to process interrupt routines as fast as possible in MCU's, as it will be unable to trigger another interrupt until idle.

The test was carried out with an additional MCU dedicated to timing. The main MCU (MCU_M) is setup to switch the state of a digital IO pin (Pin1) when it begins and ends the sampling process i.e. Pin1 is set high when sampling begins, then low when finished. This allows the timer MCU (MCU_T) to monitor state change on Pin1, starting a timer when a rising edge is detected (signifying sampling is starting), and stopping the timer when a falling edge is detected. The purpose of using a separate MCU dedicated to

timing is because it is less likely to be affected by concurrent processes. If another timer is run on MCU_M to record the time taken to perform tasks, depending on the processing power of MCU_M, having an additional timer running concurrently with the main task can potentially cause a lag on the timing, hence why it is good practice to have a dedicated unit, especially for timing operations.

5.2.2 Fast Fourier Transform (FFT) analysis

So far, the captured acceleration samples discussed is with respect to the time domain, but a significant part of vibration analysis requires investigating the signal in its frequency domain. The aim is to determine the frequencies and magnitude of the different signals that make up the, making it easier to identify the nature and source of vibration based on their respective frequencies, which will be more difficult to extract from the signal's time domain only. Fast Fourier Transform (FFT) algorithm [35], converts a signal from time to frequency domain, where the maximum frequency indicated by the algorithm is based on the signal's sampling rate, and follows the Nyquist theorem in equation (5–2).

$$f_s \geq 2f_{max} \tag{5–2}$$

Where $\boldsymbol{f_s}$ is the sampling frequency, and $\boldsymbol{f_{max}}$ is the maximum frequency that can be detected on the signal.

For applications that require FFT analysis, the number of samples, and sampling rate for the original (time domain) signal should be decided based the required FFT result. Each sample in the frequency domain is known as a bin, and total number of bins is half the FFT size as shown in equation (5–3), where each bin has a resolution (frequency difference per bin) based on the sampling rate and FFT size as shown in equation (5–4).

$$n_{bins} \geq \frac{N}{2} \tag{5–3}$$

$$FFT\ resolution\ (\Delta f) = \frac{f_s}{N} \tag{5–4}$$

Where $\boldsymbol{n_{bins}}$ is the number of bins, $\boldsymbol{N}$ is the FFT size, $\boldsymbol{\Delta f}$ is the FFT resolution, and $\boldsymbol{f_s}$ is the sampling rate.

Note: in vibration monitoring, the time interval of a sample set is crucial, and this interval is defined by the chosen sampling rate and size. Analysing shorter time intervals (< 3 seconds) at higher sampling frequency is favourable in terms of accuracy, especially in a vehicular environment regarding road analysis. Considering, the speed at which the vehicle travels determines the length of road to be analysed, in a high-speed environment, a greater distance will be travelled per time. If the time span of a sample set is large, any road surface irregularities occurring in a fraction of that sample set will be overshadowed (having reduced magnitude on the frequency domain). In this environment, a smaller time span for the sample set is preferred, to obtain a more accurate profile of the surface irregularities.

5.2.3 Acceleration to velocity conversion in frequency domain

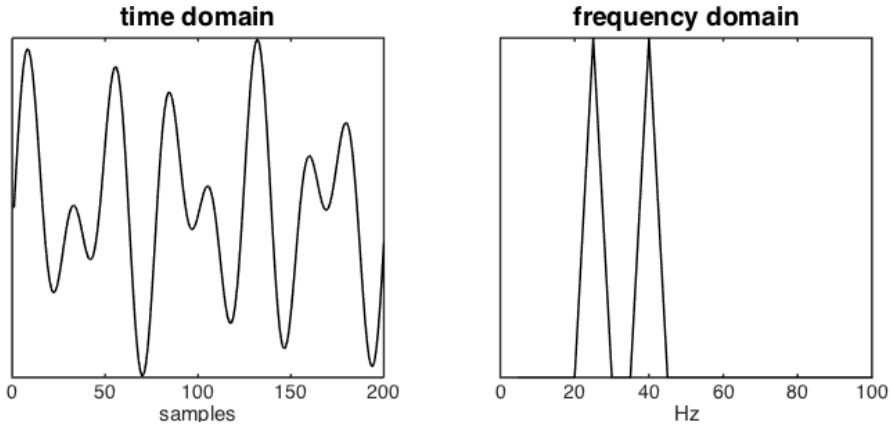

Figure 5-1 Sample vibration signal in time and frequency domain

After performing an FFT analysis on the acceleration samples, one of the core aspects of the technique is evaluating the vibration velocity based on the acceleration spectrum. Vibration signals are sinusoidal in nature, and Figure 5-1 shows a typical representation of a vibration signal caused by oscillations from two sources, i.e. 25Hz and 40Hz in this case. A vibration signal is typically made up of several sinusoidal waveforms, where each waveform indicates a different source of vibration. Therefore, it is essential to observe the spectrum of a signal sampled in the time domain, to identify all unique vibration sources and their respective magnitudes. The intensity of vibration is reflected as the magnitude of the vibration source, and this is unaffected by the offset (which indicates the presence of a DC signal, and is reflected by a vertical shift in the overall signal on the amplitude axis) and phase shift (a shift in the time domain between two signals) of the signal.

$$\omega = 2\pi f \tag{5–5}$$

$$a(t) = \cos(\omega t) \tag{5–6}$$

$$v(t) = \int a(t)dt \; = \frac{\sin(\omega t)}{\omega} + C = \frac{\cos\left(\frac{\pi}{2} - \omega t\right)}{\omega} + \; C \tag{5–7}$$

$$v(t) \simeq \frac{a(t)}{\omega} \simeq \frac{a(t)}{2\pi f} \tag{5–8}$$

Where $\boldsymbol{\omega}$ is the angular velocity, $\boldsymbol{f}$ is the frequency, $\boldsymbol{a(t)}$ is the acceleration signal, $\boldsymbol{v(t)}$ is the derived velocity, and $\boldsymbol{C}$ is a constant.

For a sinusoidal signal, each revolution is turned by an angle of $\boldsymbol{2\pi}$ radians (i.e. 360°), which equates to $\boldsymbol{2\pi * f}$ when evaluating the angle turned per second, where $\boldsymbol{f}$ is the number of revolutions per second (i.e. frequency of the signal in Hz). This is the angular speed (ω) as shown in equation (5–5).

Following the relationship between acceleration and velocity in equation (2–1), velocity is the first integral of acceleration. Therefore, for an acceleration signal $\boldsymbol{a(t)}$ in equation (5–6), its velocity is evaluated as shown in equation (5–7). The integrated signal includes an unknown constant (C) and a phase shift of $\boldsymbol{\pi/2}$ radians (90°). The constant imposes an offset on the signal, and as mentioned previously, is a common source of error when performing integration in the time domain. Since vibration is based on the magnitude of the vibrating signal, the phase difference and offset on the signal is irrelevant. Ignoring both characteristics of the signal, the equivalent equation for velocity is expressed as in equation

(5–8), which is the division of the acceleration signal by the angular velocity. This simplification is the basis for deriving the velocity from the acceleration signal's spectrum, and is expresses as shown in equation (5–9). For each bin on the acceleration's frequency spectrum, the magnitude of the velocity can be derived using the equation.

$$|v(n)| = \frac{|a(n)|}{2 * \pi * n * \Delta f} \tag{5–9}$$

Where $\boldsymbol{|v(n)|}$ is the magnitude of the velocity, $\boldsymbol{|a(n)|}$ is the magnitude of the acceleration per bin, $\boldsymbol{\Delta f}$ is the FFT resolution, and $\boldsymbol{n}$ is the bin index.

5.2.4 Root Sum Squared (RSS) moving average

As the name suggests, this step runs an RSS moving average across the bins in the frequency domain. RSS takes the square root of the summed squared of all the values as shown in equation (5–10).

$$RSS = \sqrt{b1^2 + b2^2 + \cdots + bn^2} \tag{5–10}$$

Where $\boldsymbol{b1 - bn}$ are the values to be averaged.

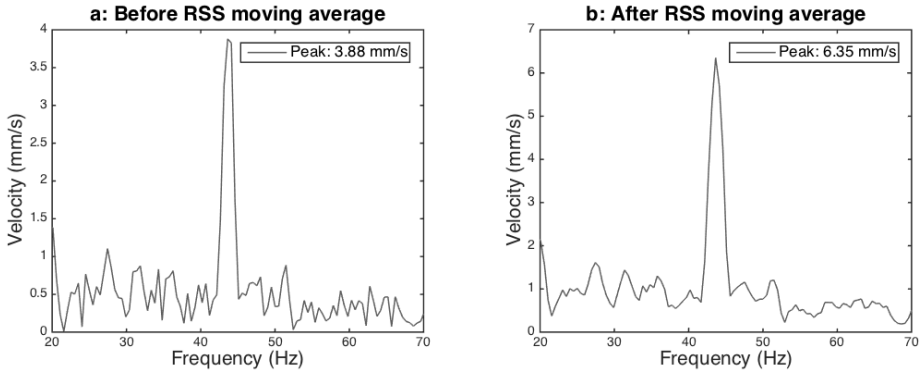

Figure 5-2 Effect of RSS moving average

The purpose for this is due to inconsistencies in the frequency spectrum, where (in some cases) the peaks for the same vibration levels can appear narrow (at a single frequency bin) or wide (spanning across multiple bins). A narrow peak reflects a more accurate magnitude when compared to wide peaks, because the magnitude of a wide peak is reduced and spread across neighbouring bins as shown in Figure 5-2 (reflecting the magnitude changed before and after RSS moving average).

$$a(n) = \sqrt{\begin{matrix} v(n-l)^2 + \cdots + v(n-1)^2 \\ +v(n)^2 + \\ v(n+1)^2 + \cdots + v(n+l)^2 \end{matrix}} \tag{5–11}$$

Where $\boldsymbol{a}$ is the RSS magnitude, $\boldsymbol{v}$ is the velocity, $\boldsymbol{l}$ is the number bins on either side of active peak, and $\{\boldsymbol{n} \in \mathbb{R} \mid \boldsymbol{1 \leq n \leq N}\}$.

The RSS moving average aims to correct this inconsistency in the frequency domain, by evaluating the magnitude of a bin based on the RSS average with its neighboring bins using equation (5–11), where the averaging window length defined in equation (5–

12) is the number of bins to be averaged. The RSS moving average used is symmetrical i.e. equal number of adjacent bins are used to calculate the RSS average.

$$L = 2l + 1 \tag{5–12}$$

Where $\boldsymbol{L}$ is the window length, and $\boldsymbol{l}$ is the number of adjacent bins.

5.3 Experimental design and Testing

This section describes the hardware, software, and test procedure used to validate the vibration calculation technique mentioned in section 5.2.

5.3.1 Hardware setup

Figure 5-3 Hardware Setup

Figure 5-3 illustrates a block level diagram of the measurement hardware, which consists of a:

5.3.1.1 DsPIC33EP256MC202 (Microcontroller)

An MCU developed by Microchip, which has dedicated hardware to facilitate Digital Signal Processing (DSP). It is the brains of the operation, programmed to handle sensor sampling, processing and vibration calculation, and section 5.3.2 describes the MCUs software operation in detail. The factors considered in choosing an appropriate MCU for this application, arranged in order of importance, include DSP capabilities, Random Access Memory (RAM) limit, availability and price. Since the MCU is expected to perform FFT analysis on the sampled signal, it is vital for the selected controller to be able to perform such operations quickly and efficiently, and the availability of an integrated DSP means the DsPIC33EP256MC202 is suitable. The RAM indicates how much memory is available to the MCU during run time, and this is a temporary storage system that is cleared on reset. RAM size is equally as important since sensor samples need to be read in real time and temporarily stored, to be able to perform any signal

processing on them. The RAM needed for this application is directly proportional to the FFT size, and since the memory on these types of MCUs are relatively limited, a decision must be made based on the sampling rate and frequency resolution required for the application.

5.3.1.2 ADXL345 (Accelerometer)

A MEMS 3-axis accelerometer manufactured by ANALOG DEVICES, with a rated maximum sampling rate of 3200sps. With the market for MEMS capacitive accelerometers saturated, the decision for choosing a sensor becomes less about accuracy and sampling rate, as most major manufacturers offer similar performance. The major factor influencing the decision to use the ADXL345 in this research is its popularity, giving the convenience of having an abundance of recourses available regarding its setup, integration with MCUs, programming and performance.

5.3.2 Software design

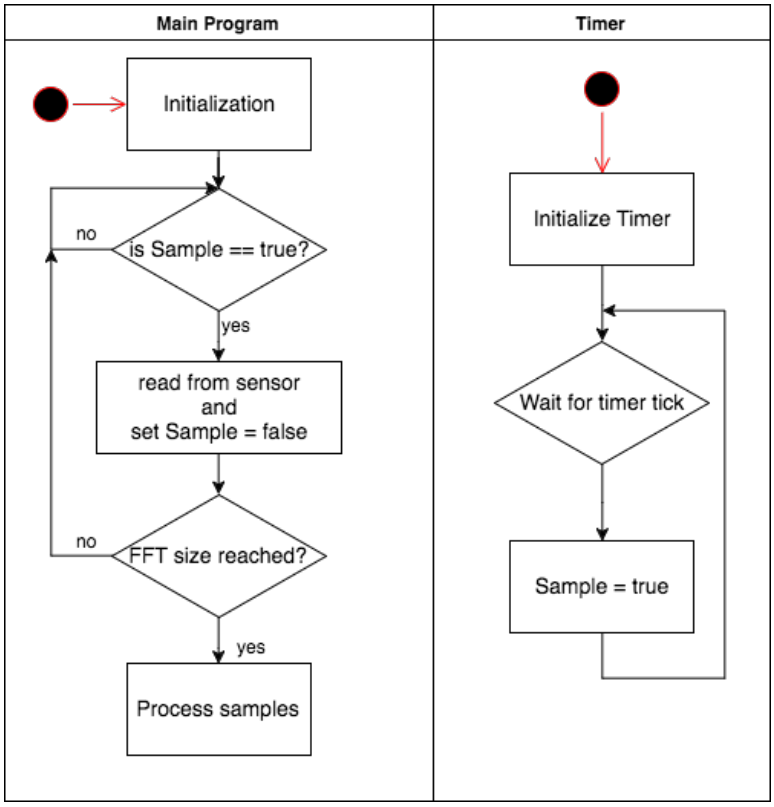

Figure 5-4 Software flow chart

The flowchart above describes the software operation carried out by the designed test unit. The tasks have been split into different parts, which are discussed below.

5.3.2.1 Initialization

This part of the task defines and initialises all the variables used in the program. This includes the system clock (defining the MCU speed), sample buffer (holds the samples to be sampled, where the size of the buffer is dependent on the FFT size), sample flag (indicating when the accelerometer is to be sampled), and timer (with a timeout set according to the required sampling rate).

5.3.2.2 Timer

The system uses TC (in section 5.2.1) for data sampling, hence, the timer, setup in the initialization process sets the sample flag via an interrupt that is triggered after the defined timeout period. This timer is run in parallel (concurrently) with the main process that handles the data sampling and processing.

5.3.2.3 Measurement and processing

At this point, if the sample flag is cleared, the main process waits in a loop for this flag to be set from the timer interrupt. A set flag indicates that the selected sampling period has elapsed, therefore the main process breaks out of the wait loop and reads a sample from the accelerometer, keeping a record of the number of samples read by incrementing the sample size for every sample. If the sample size is less than the selected FFT size, the measured sample is stored in the buffer, clears the sample flag, and goes back to waiting for the flag to be set by the time interrupt. This process continues until the number of samples in the buffer is equal to the FFT size, in which case the buffer is processed, calculating the vibration velocity based on section 5.2. After the calculation, the whole process is repeated, with the program going back to obtaining new samples.

5.3.3 Testing

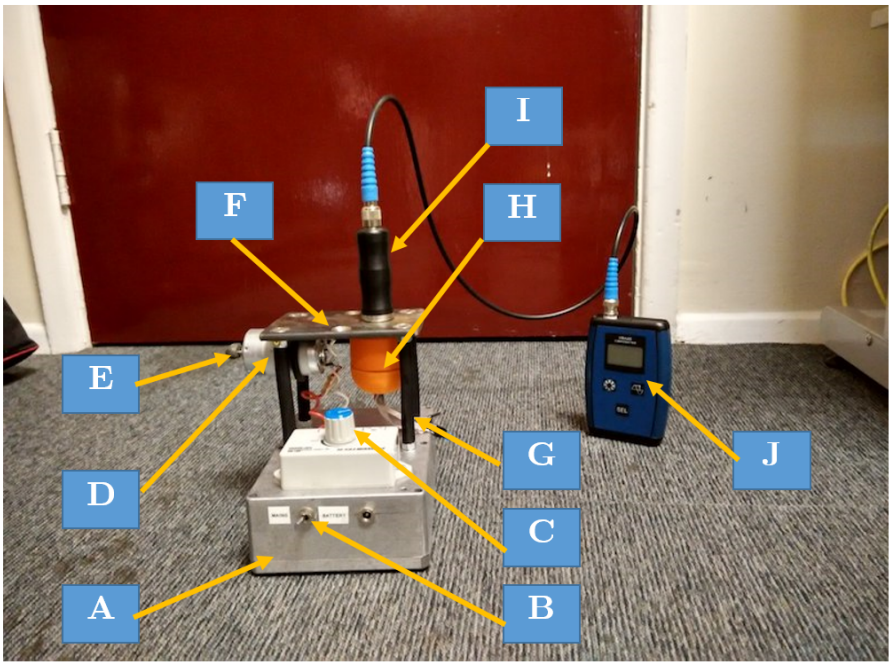

Figure 5-5 Image of test setup. A: base, B: power switch, C: potentiometer, D: motor, E: motor shaft, F: vibration plate, G: rubber support pillars, H: custom vibration meter, I: HS-620 accelerometer, J: HS-620 control unit.

To verify results obtained using the proposed method, and industrial grade vibration meter (HS-620) manufactured by Hansford Sensors was used. Figure 5-5 shows the test setup, where the HS-620 and the developed vibration meter are mounted at the same spot on the designed test rig, to simulate different vibration levels.

Essentially, a 12v DC motor (D) is used to generate the desired vibration on the rig. The base (A) of the vibration rig contains a rechargeable battery, charging circuit, 12v DC input socket, and a switch to control power to the motor. The power circuit is designed to source power from the battery when no external DC supply is connected. If an external source is detected, the battery supply to the motor is cut off automatically using a P-Channel MOSFET (PMOS), and power is sourced directly from the external DC input, while simultaneously charging the battery via the charging circuit, monitoring the state of the battery, and stopping the charging process when full.

The motor is fitted to a rectangular metal plate (F) (where the vibration meters will be mounted to take measurements), which is attached to the base of the rig with rubber tubes (G), to allow more pronounced movements of the plates while the motor is

running, generating increased magnitudes. With this system, the frequency of vibration can be set, and is controlled using a potentiometer (C), by varying the speed of the motor, but the magnitude on the other hand isn't linearly adjustable, but is determined by the combined effect of the motor's speed, weight on its shaft (E), and rubber stands. The weight attached to the shaft of the motor determines the intensity of the stress applied to the rubber stands while in operation, and this relationship is directly proportional i.e. increasing the weight increases the intensity, hence, higher vibration magnitudes can be generated for the set frequency range of the motor. As mentioned earlier, this is not a linearly adjustable process, and required trial and error to produce required magnitudes.

An important part of the test was discovering areas on the metal plate where the vibration is expected to be the same, but the non-uniform nature of the vibration rig makes this difficult and near impossible, based on the measurements carried out by the HS-620 meter. The purpose of this was to be able to mount both meters on the rig, concurrently monitoring the vibration levels measured by the meters, while changing the motor's frequency and shaft weight.

In the tests carried out with the HS-620 meter, different points on the plate produced different vibration measurements for the same motor speed and shaft weight. Although some points did reflect measurements closer than others, the measurements are not conclusive because the test was done using only the HS-620, and further investigations revealed any slight pressure applied to another section of the plate while taking measurements resulted in a change in vibration measured by the meter. Therefore, having the developed test meter mounted on a different location with the HS-620 is expected to change the rig's vibration, since it is equivalent to adding pressure to the plate equal to the weight of the unit.

To ensure the same vibration is concurrently sensed between both vibration meters, they had to be mounted at the same spot on the plate. Using this theory, the test was carried out by fixing the meters on either side of the plate (using a double-sided threaded bolt), i.e. top and bottom as shown in Figure 5-5.

It is also vital for the MEMS accelerometer and PCB to be firmly fitted to the enclosure to prevent movements, since this would cause the magnitude of oscillations to be multiplied by a factor dependent on amount of looseness. The unit developed for the research is potted with epoxy resin (Oxirane-containing oligomers, which cure through the reaction of epoxide groups with a suitable curing agent [64]) to guarantee this, and

to increase durability, ensuring all components are sealed and compact, forming a solid structure. With all electronic components potted inside the enclosure, the power and data (UART) lines were exposed from the PCB using cables, to be able to power and communicate with the MCU and sensor externally.

5.4 Results and Analysis

This section compares the vibration calculation process described in this chapter (*Method I*) with the conventional method. Measuring vibration velocity (mm/s) using accelerometers requires conversion from acceleration to velocity. Normally, this is achieved by integrating the acceleration samples from the sensor (in the time domain) to calculate the velocity, then the vibration frequency analysis is done on this derived velocity (this will be referred to as *Method II*). The problem with this method is the accumulated drift that is added to the calculated velocity signal. This is because any DC component present in the acceleration signal before integration causes each integrated sample in the resulting velocity to increasingly stray away from the actual value, causing the drift.

The goal of the described method of vibration analysis is to eliminate this drift in the derived velocity, by evaluating the signal based on the acceleration's frequency domain, and bypassing sample integration in the time domain.

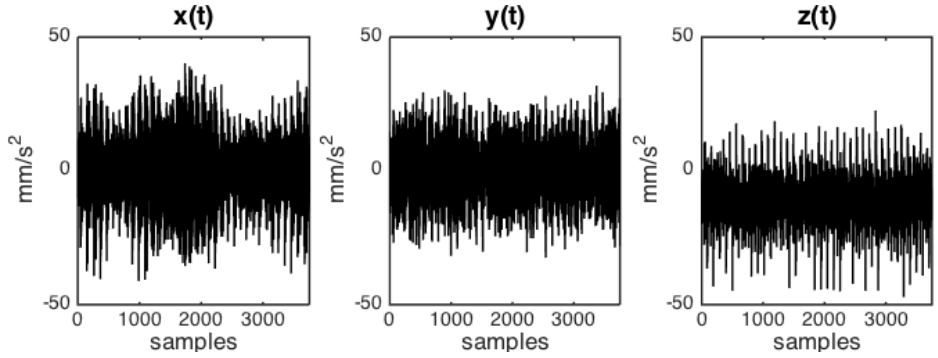

Figure 5-6 Graph of acceleration samples for x, y and z axis

Figure 5-6 shows the acceleration samples acquired for a vibration velocity of 9 Vrms generated with the test bed described in section 5.3.1. The measurement range for the in this application is between 10 – 1000Hz, which is best expressed as vibration velocity as opposed to displacement or acceleration as explained in Chapter 3. Following Nyquist theorem, a sampling rate of at least 2 x 1000Hz is required to suitably record

vibrations within the defined range. For the results shown in this section, a sampling rate of 2500Hz was used.

Observing the acceleration samples for all three axes, the DC offset causes a drift in the calculated velocity as shown in Figure 5-7. This drift appears more pronounced on the z-axis, because the measured acceleration on the axis has the most DC offset (approximately -10mm/s^2) compared to x and y. This offset on the z-axis is due to the mount position of the measurement device on the test bed, where the z-axis of the accelerometer is parallel to the earth's gravitational force. This drift imposed on the vibration signal has a significant effect on the velocity's frequency domain as shown in the plot, making it impossible to visualise the properties of the vibration signal accurately.

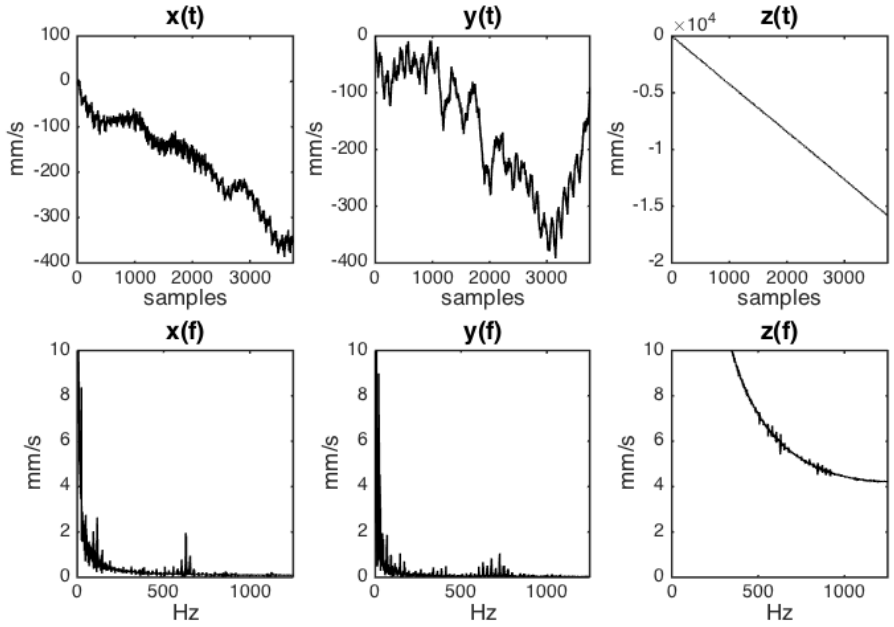

Figure 5-7 Time and frequency domain representation of derived velocity signal for x, y and z axis (without high-pass filter)

Reducing the DC offset on the acceleration samples using a high pass filter generally reduces the drift in the calculated velocity as shown in Figure 5-8, producing a much clearer representation of the vibration signal on the frequency domain. This is the widely adopted practice for vibration analysis using accelerometers.

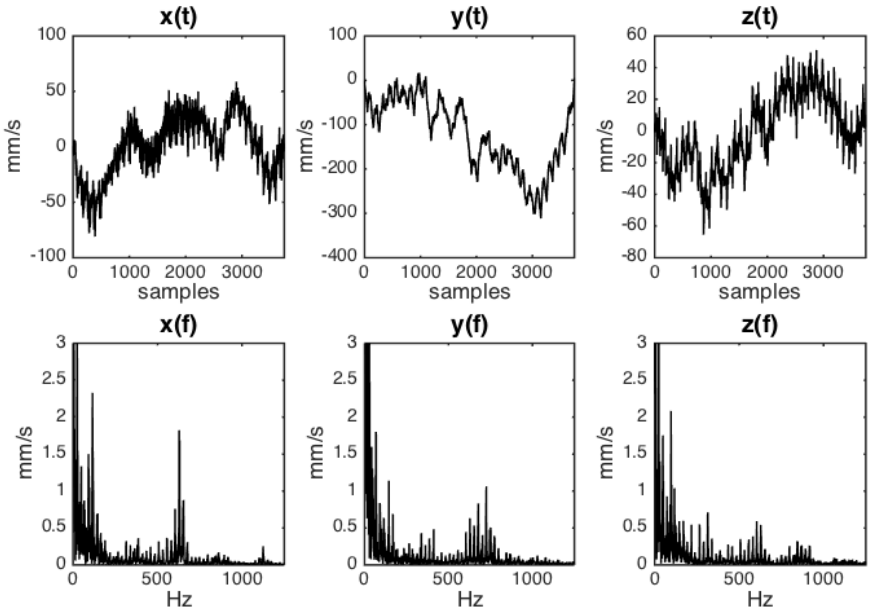

Figure 5-8 Time and frequency domain representation of derived velocity signal for x, y and z axis (with high-pass filter f_c=10Hz)

Figure 5-9 illustrates the difference between the result derived using *Method I* (A), and the result obtained from *Method II* (B and C, which shows the effect of a high-pass filter). The plot 'A' was evaluated from the raw acceleration samples without a high-pass filter. The absence of drift in 'A' is in direct contrast to 'C', which was also evaluated without a high-pass filter, using *Method II*. With a high pass filter applied to 'C', a more appropriate representation of the vibration signal 'B' is observed, which reflects similar characteristics to 'A'. The main difference between 'A' and 'B' is more evident on the lower frequency component of the graph (<5Hz) as shown in the image. This is a consequence of the high-pass filter applied to 'B', causing the magnitude of all frequencies lower than the applied cut-off frequency to be attenuated.

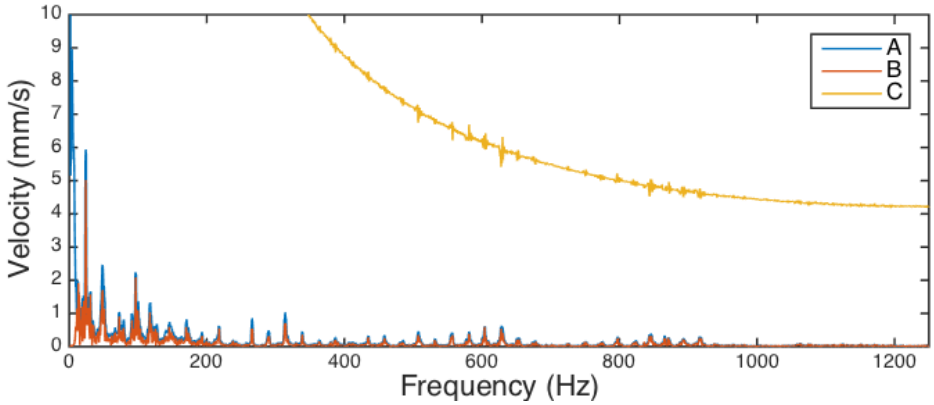

Figure 5-9 A: vibration velocity derived from *Method I*, B: Vibration velocity derived from *Method II* (with high-pass filter f_c=10Hz), C: vibration velocity derived from *Method II* (without high-pass filter)

When using *Method II*, a high pass filter with an appropriate cut-off is acceptable as longs as it is within the defined measurement limits. A benefit to *Method I*, is the irrelevance of a high pass filter, hence there is no attenuation or loss of data to any part of the signal.

Figure 5-10 show more comparison results from *Method I* and *Method II* at different vibration levels. The image shows the vibration analysis made by both methods ('a' and 'b') for two successive measurements (1 and 2) taken at the same vibration level. In this scenario, the theoretical expectation is to observe very similar vibration properties from successive measurements, since the vibration signal is unchanged. It can be observed that the peaks labelled 'x' in 1a and 2a have a significant difference in their magnitude. This is typically a result of spectral leakage as discussed in Section 3.3.4, which is rectified by *Method I* as shown in 1b and 2b. The same observation can be realised in Figure 5-10.

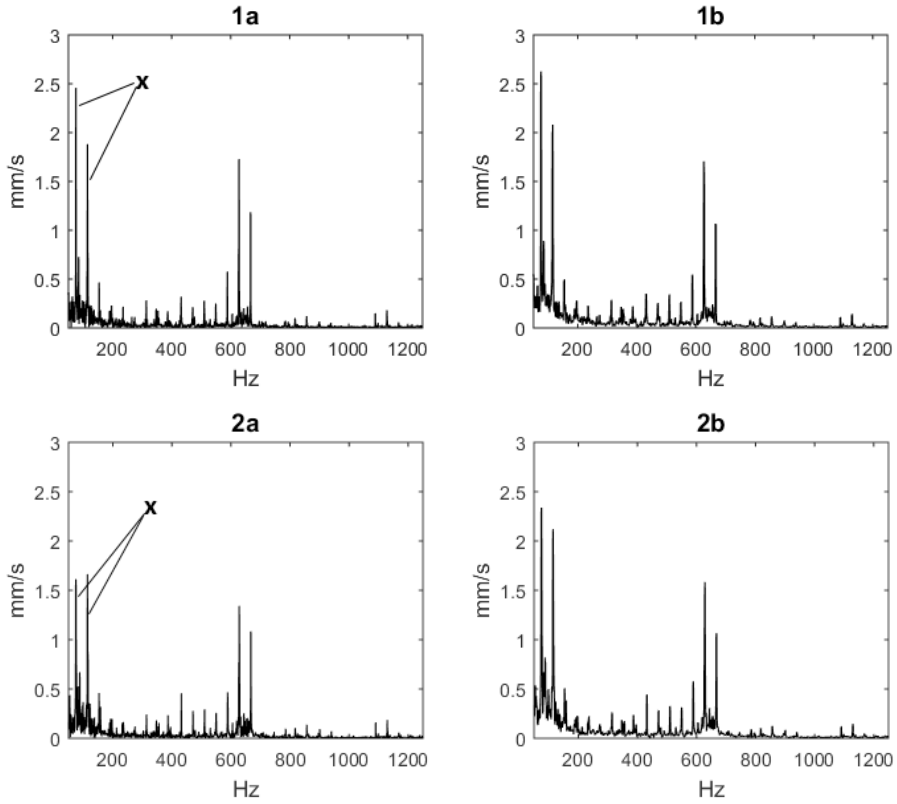

Figure 5-10 1a and 2a: vibration velocity derived using *Method II* (with high-pass filter f_c=10Hz), 1b and 2b: vibration velocity derived from *Method I.*

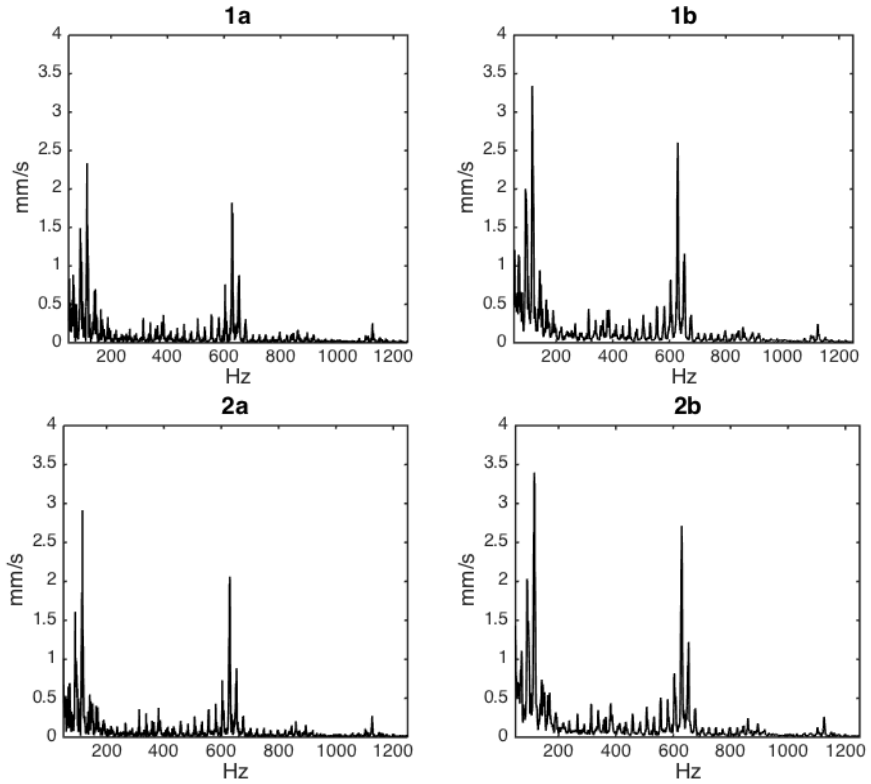

Figure 5-11 1a and 2a: vibration velocity derived using *Method II* (with high-pass filter f_c=10Hz), 1b and 2b: vibration velocity derived from *Method I*.

Another advantage of *Method I* is in relation to its computational requirements, which is especially important for microcontrollers with limited available memory. With *Method II*, the memory that is required to process the vibration velocity is at least three times the amount of memory needed to store all the acceleration samples necessary for computing the analysis. Essentially, for any number of samples measured from the sensor (to evaluate the vibration analysis), deriving the velocity and FFT bins will require twice the memory needed by the acceleration samples. So, for a sampling rate of 2500Hz, if the vibration analysis is done for two seconds, 5000 samples will be acquired from the sensor. Assuming 1 byte per sample, the minimum memory required to compute the vibration velocity in the frequency domain is 15000 bytes i.e. 5000 bytes for the acceleration samples, velocity samples, and FFT bins respectively.

For *Method I*, the amount of memory required to compute a vibration analysis follows the relationship defined in equation (5–13).

$$memory = M(1 + A) \tag{5–13}$$

Where $\boldsymbol{M}$ is the memory required to store the acceleration samples, and $\boldsymbol{A}$ is the ratio of the frequency range evaluated to the maximum FFT frequency.

With *Method I*, the minimum memory requirement depends on the expected frequency range to be evaluated. For worst-case scenario (where A = 1, i.e. the full frequency range is evaluated), the total memory required to compute the analysis is 2M. The convenience of the technique over the former is that the memory requirement is linearly proportional to the range of frequencies evaluated i.e. if only half the vibration spectrum needs to be computed (i.e. A = ½), the memory required will be 1.5M. This is ideal in a memory-constrained environment.

5.5 Summary

The objective of this chapter was to investigate and propose alternatives methods of calculating vibration velocity using readily available technologies. In this chapter, the focus is solely on accurately depicting the measured vibration on a spectrum, and not to analyse or evaluate the actual source or meaning of the vibration. A MEMS capacitive accelerometer was preferred over piezoelectric, even though piezoelectric sensors are generally preferred in vibration analysis for their wider frequency range. Capacitive accelerometers offer flexibility in terms of size as these come in very small surface mount packages. This allows measurements to be taken at space-limited areas.

In relation to the research, the purpose of this study is to cooperatively analyse the effect of road surface texture on vehicle vibration, exploring a different dimension to roughness evaluation. The use of accelerometers creates challenges in this application, especially when a conversion is required from acceleration to velocity. Difficulties in estimating the initial velocity and the presence of drift on the derived velocity, affect the accuracy of the corresponding result. These are well known issues in this area, and there are proven techniques to overcome these problems as discussed in section 2.7.

The method proposed (*Method I*) improves on some vital factors when compared against the usual method used in most vibration meters (*Method II*). Firstly, it eliminates the need for a high-pass filter, which is an essential part of *Method II* to minimise the accumulated drift that is caused by the presence of a DC signal on the acceleration samples, as illustrated in Figure 5-9. This is a combined side effect of not being able to determine the initial velocity during integration from acceleration. This feature conveniently saves on processing time, since this extra step is irrelevant. The computational process for *Method I* also requires less memory, as defined by equation (5–13), providing the flexibility of only processing desired frequency band, conserving memory and processing time. For modern day computers, the memory and processing difference is insignificant due to the abundance of resources present. However, the

application is targeted to resource-limited microcontrollers, which would realise an increase in overall performance.

For vibration analysis in a three-dimensional plane, the resultant magnitude of the vibration using a tri axial accelerometer is derived using vector addition shown in equation (5–14), which is true for vector 90^{o} apart. A comparison was done determine the effect of evaluating the resultant magnitude before or after performing the vibration analysis. Results showed that better precision and consistency was obtained when the magnitude was calculated after the analysis.

$$x^2 + y^2 + z^2 = r^2 \tag{5–14}$$

Where $\boldsymbol{r}$ is the resultant magnitude, and $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$ are acceleration values for the x, y and z-axis of the sensor.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

Preceding chapters have described novel techniques for evaluating irregularities in road surfaces, highlighting improvements to usual techniques, with details on experimental tests and corresponding results. This chapter presents the conclusions and feasible future research directions. Section 6.1 outline the deductions made from the results obtained, justifying the experiments carried out. Section 6.2 expresses some thoughts on further work and future direction of the research.

Faster, portable, and efficient technologies that advances the transport industry is a continuous necessity. Some of the fundamental and key motivations are safety, maintenance, and environmental awareness. The research in this thesis aspires to target these motivations by improving current methodologies in road monitoring, providing a convenient and scalable approach to road surface measurement. Chapter 1 concisely defines the problem domain, which supports the research drive to improve the current state of pavement monitoring, facilitating faster growth in applications relating to road maintenance.

The research introduces a novel Half-Wavelength Peak Matching (HWPM) model, to improve the accuracy of longitudinal road profiles generated by high-speed inertial profilers. This model operates on a peak matching principle that estimates an error offset on a sensor, based on measurements from another, and is best suited for applications where movements are bound to a single axis. Details of the proposed

correction technique is comprehensively described in Chapter CHAPTER 4, including related mathematical models and program implementations in pseudo code. To test the HWPM model in real world conditions, an inertial profiler was developed for this research, made up of a laser displacement sensor, accelerometer, and Doppler speed sensor. Details of the hardware and software implementation of the developed profiler is shown in Chapter CHAPTER 3, including validation tests done to ensure and verify the accuracy of the data obtained from the profiler.

6.1 Outcomes

Vehicle based rapid road profiling has become a popular concept in pavement analysis, largely due to the demand for fast and portable means of road monitoring. Its ability to scan road surfaces without the need for road closures is a huge advantage to other more conventional profilers, which are typically manned or static. Traditional profilers like the 3m Rolling Straight Edge (RSE) are still very much in use, since they generally offer more consistent accuracy. Although, these are relatively cheaper than the rapid profilers, their efficiency progressively reduces as the length of road increases. Due to fine differences in longitudinal road profiles (usually in millimetres), vehicle based profilers are better suited for evaluating road textures with wavelength greater than 50mm (which account for mega textures and unevenness). Multiple factors play a part in the consistency of these high-speed profilers, which range from sensitivity of the sensors, to the profiler's mount location on the vehicle. Hence, why their profilers are usually evaluated as a moving average of a defined distance. Studies are consistently being done to improve the accuracies of these profilers because they offer more room for growth, since they can easily be automated.

The HWPM model proposed in this thesis show significant improvements to the roughness profiles generated with a vehicle based profiler. A major concern for high-speed profilers that use a combination of laser displacement sensors, and accelerometers to evaluate roughness of a road surface, is the accumulated drift from integration. Integration is essential in this application to convert acceleration samples (from the accelerometer) to displacement. The theory of these profilers is to eliminate the vehicle's vertical dynamics (caused by shock absorbers due to the road's longitudinal irregularities) from the samples recorded by laser displacement sensors. This is the purpose of the accelerometer. The relationship between acceleration and displacement make them a convenient sensor in this application. Unlike displacement sensors, which requires a physical reference to determine the distance from a point, accelerometers, do

not, because they measure the rate of change of velocity, hence, are well suited to determine the vehicle dynamics. However, as previously mentioned, the issue is the accumulated drift on the displacement signal, after double integrating the acceleration samples. The cause and effect of this drift on the resulting profile is discussed in Chapter CHAPTER 4.

The vertical motion of the inertial profiler during measurement (while mounted on the vehicle) is irregular, since its movement is determined by the road surface irregularity and its effect on the vehicle's dampers. Hence, the frequency of the displacement waveform will be in the range of motion tolerated by the dampers, and will alternate between this range for different sections of the measured road, due to varying surface irregularities.

The usual practice used to eliminate this drift is by applying a high-pass filter on the signal with a high enough cut-off frequency to remove the drift bias, but low enough to maintain the accuracy of the signal. Investigations revealed that increasing the vehicle speed at which the measurements are taken, progressively increased the drift, hence, a higher cut-off frequency is required. The studies in Chapter 5 proved there is a limit on the high pass cut-off frequency (where the signal become attenuated or altered), but in some cases, this cut-off frequency is insufficient to properly remove the drift (mostly at higher speeds). Choosing an appropriate cut off frequency for the high pass filter is essential to the accuracy of this calculated displacement. From the samples measured during the research, a slight offset of ±0.3 Hz from the ideal cut-off frequency causes an offset (between ±5mm) to the calculated displacement. In road surface monitoring this has a huge effect to the evaluated road profile, since the thresholds indicated by the standard [59] for acceptable surface irregularity heights are between 4mm, 7mm and 10mm. The results from the proposed HWPM model (shown in section 4.3) highlights the benefits of the technique in this application, because the error offset on the calculated displacement is approximately ±1mm for a high pass cut-off range between ±0.5Hz from the ideal cut-off frequency.

Part of the success of the HWPM model is its ability to further reduce the drift on a signal at lower cut-off frequencies, without compromising the signal's originality. To verify the necessity for HWPM, the laser sensor and accelerometer where displaced from a static position, with the laser pointing to a fixed reference to ensure both sensors recorded the same displacement. Samples obtained before applying the HWPM model showed a difference in peak amplitudes between the displacement measured from the laser and accelerometer. Although the waveforms were consistent in terms of phase and

wavelength, theoretically, the waveform from both sensors were expected to have the same properties. Results achieved after HWPM for the same test case show the benefits and accuracy of the model in correcting the differences in peak amplitudes. A Laser Based Roughness Measurement (LBRM) device (described in Chapter 4) was designed and developed to validate the work done in this research. The LBRM device, which consists of a laser displacement sensor and an accelerometer, is fixed to the rear bumper of the vehicle to measure actual road surface samples. Section 4.3 reviews the samples obtained from the device, analysing the effect of the HWPM model on the corresponding road profiles, which was largely positive. Although, it was also evident that little difference was observed for road samples taken at speeds less than 20mph, mainly because there is less accumulated drift at low speeds.

The vibration measurement technique discussed in ChapterCHAPTER 5, which uses a MEMS capacitive accelerometer provides an alternative to a market saturated with vibration meters based on piezoelectric sensors, which are generally preferred for vibration analysis because they allow for high frequency measurements (greater than 1000Hz). The aim of the study is to investigate the effect of road roughness on vehicle vibration. At the stage of the thesis, a hardware was designed and developed with a novel evaluation process to measure the vibration velocity of a body, and to be integrated with the LBRM profiler. The measurement range for the target frequency is between $10 - 1000$Hz, which is based on the requirements from the IS0 10816 standard. Test results were compared against certified industrial meters to verify the accuracy, with favourable outcomes. There is also an improvement in terms of memory requirements for computing the vibration velocity's spectrum. This is valuable when low-end microcontrollers with limited resources are used for the vibration computation. Capacitive accelerometers typically have an advantage of size over their piezoelectric counterparts, and this study encourages the flexibility of its use in other applications requiring vibration analysis, especially in a space-constrained environment.

6.2 Future Directions

The previous section (6.1) hinted at a possible area for further explorations on the HWPM model. The model relies on the theory that low frequency peaks on the laser samples which are also present on the accelerometer's displacement samples reflects the vehicle's vertical dynamics, while peaks out of phase (180^{o}) or inconsistent between both signals reflects roughness. This is based on numerous investigations carried out from analysing multiple road surfaces. The studies done were focused on these three

scenarios to identify areas of the signal that require correction, but the significance of the phase difference between the peaks needs to be explored, to identify relationships with the HWPM's peak correction offset.

An area for future work the ability to generate a multi-point longitudinal profile. Currently, the LBRM profiler generates a two-dimensional longitudinal representation of the road surface along a line the size of the laser's diameter. This is a single-point profile that takes no consideration for the road's lateral roughness. With the profiler designed for portability and scalability, attaching multiple profilers along the bumper of the vehicle allows the lateral profile of the road (equal to the width of the vehicle) to be visualised along with its longitudinal profile. Increasing the number of profilers would increase the precision of the lateral profiler generated. This will require a revision of the current sampling process, because each sample from the profilers would require synchronization to maintain accuracy of the profile results. This can be done by either performing the data synchronization internally on the profiler, externally on the user application, or a combination of both. The preferred approach is the combination of the two methods, to prevent modifications to the hardware architecture, or complex algorithms on the user application. Since the profiler communicates with the user application via a TCP connection, multiple devices connected to the application would be able to communication peer-to-peer, where one device operates as a master while the others operate as slaves. This means the master device can handle the time synchronization between all connected devices, while the user application samples, and sorts the data according to their timestamps.

Another possible advancement to the research is achieving real time driver awareness of road conditions. The overall architecture for this service will require a regularly updated database of road profiles, tagged with their respective localisation information. Capable vehicles can request this road profile, giving them an awareness of the road surface condition and allowing vehicles to notify drivers as they approach potentially hazardous areas. There are challenges that need to be overcome to enable services like this, and one of the major hurdles is the ability to store the amount of data generated by profilers. The LBRM profiler developed for this research is capable of 2500Hz sampling rate. Considering the amount of data that would be acquired over several kilometres, this will require complex compression and data formatting algorithms to shrink the data that is stored, while maintaining sufficient information about the profile. There is also the challenge of the communication channel being able to handle this amount of data to be transmitted to several nodes (vehicles in this instance). This

is a big data challenge considering the implications of a nationwide implementation of the service. However, this is beyond the scope of this research.

Bibliography

[1] DFT, "Transport Infrastructure Skills Strategy: Building sustainable skills," *Dep. Transp.*, 2016.

[2] A. Vittorio, V. Rosolino, I. Teresa, C. M. Vittoria, P. G. Vincenzo, and D. M. Francesco, "Automated Sensing System for Monitoring of Road Surface Quality by Mobile Devices," *Procedia - Soc. Behav. Sci.*, vol. 111, no. Supplement C, pp. 242–251, 2014.

[3] M. W. Sayers and S. M. Karamihas, "The little book of profiling," *Basic Inf. about Meas. Interpret. Road Profiles*, no. September, p. 100, 1998.

[4] ASTM E950/950M-09, "Standard Test Method for Measuring Longitudinal Profile of Traveled Surfaces with an Accelerometer Established Inertial Profiling Reference," vol. 9, no. Reapproved, pp. 1135–1140, 2009.

[5] ASTM E867-06(2012), "Standard Terminology Relating to Vehicle-Pavement Systems," 2012.

[6] ISO, "ISO 13473-5:2009 Characterization of pavement texture by use of surface profiles — Part 5: Determination of megatexture," p. 29, 2009.

[7] L. T.-G. Smith, K L, J.W. Hall, "Guide for Pavement Friction," *Contract. Final Rep. NCHRP Proj. 01-43*, no. February, pp. 1-65-91, 2009.

[8] ISO, "ISO 13473-2:2002 Characterization of pavement texture by use of surface profiles -- Part 2: Terminology and basic requirements related to pavement texture profile analysis," p. 17, 2002.

[9] S. M. Karamihas, T. D. Gillespie, S. D. Kohn, and R. W. Perera, "Guidelines for Longitudinal Pavement Profile Measurement," no. February, 1999.

[10] ASTM E1926 - 08(2015), "Standard Practice for Computing International Roughness Index of Roads from Longitudinal Profile Measurements," p. 16, 2015.

[11] G. Alessandroni, A. Carini, and L. Emanuele, "Sensing road roughness via mobile devices: A study on speed influence," *2015 9th Int. Symp. Image Signal Process. Anal.*, pp. 270–275, 2015.

[12] V. Douangphachanh and H. Oneyama, "Formulation of a simple model to

estimate road surface roughness condition from Android smartphone sensors," *2014 IEEE Ninth Int. Conf. Intell. Sensors, Sens. Networks Inf. Process.*, pp. 1–6, 2014.

[13] X. Zhu, Y. Chen, Y. Liu, and D. Liu, "Three-Dimensional Road Roughness Measurement System Development Based on Laser Rangefinder," *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*. pp. 1613–1616, 2015.

[14] Z. Yuan, X. Zhang, S. Liu, X. Han, and Y. Du, "Laser line recognition for autonomous road roughness measurement," *2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*. pp. 436–440, 2015.

[15] J. C. Tudón-Martínez, S. Fergani, O. Sename, J. J. Martinez, R. Morales-Menendez, and L. Dugard, "Adaptive Road Profile Estimation in Semiactive Car Suspensions," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 6. pp. 2293–2305, 2015.

[16] C. Banica, S. V Paturca, S. D. Grigorescu, and A. M. Stefan, "Data acquisition and image processing system for surface inspection," *2017 10th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*. pp. 28–33, 2017.

[17] S. Mathavan, K. Kamal, and M. Rahman, "A Review of Three-Dimensional Imaging Technologies for Pavement Distress Detection and Measurements," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5. pp. 2353–2362, 2015.

[18] R. Medina, J. Llamas, E. Zalama, and J. Gómez-García-Bermejo, "Enhanced automatic detection of road surface cracks by combining 2D/3D image processing techniques," *2014 IEEE International Conference on Image Processing (ICIP)*. pp. 778–782, 2014.

[19] T. Kage and K. Matsushima, "Method of rut detection using lasers and in-vehicle stereo camera," *2015 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*. pp. 48–53, 2015.

[20] T. N. Do, R. Liu, C. Yuen, M. Zhang, and U. X. Tan, "Personal Dead Reckoning Using IMU Mounted on Upper Torso and Inverted Pendulum Model," *IEEE Sensors Journal*, vol. 16, no. 21. pp. 7600–7608, 2016.

[21] A. Suprem, V. Deep, and T. Elarabi, "Orientation and Displacement Detection for Smartphone Device Based IMUs," *IEEE Access*, vol. 5. pp. 987–997, 2017.

[22] S. Zihajehzadeh, T. J. Lee, J. K. Lee, R. Hoskinson, and E. J. Park, "Integration of MEMS Inertial and Pressure Sensors for Vertical Trajectory Determination," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 3. pp. 804–814, 2015.

[23] N. Millor, P. Lecumberri, M. Gómez, A. Martínez-Ramírez, and M. Izquierdo, "Drift-Free Position Estimation for Periodic Movements Using Inertial Units," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 4. pp. 1131–1137, 2014.

[24] H. Zhao and Z. Wang, "Motion Measurement Using Inertial Sensors, Ultrasonic Sensors, and Magnetometers With Extended Kalman Filter for Data Fusion," *IEEE Sensors Journal*, vol. 12, no. 5. pp. 943–953, 2012.

[25] U. X. Tan, K. C. Veluvolu, W. T. Latt, C. Y. Shee, C. N. Riviere, and W. T. Ang, "Estimating Displacement of Periodic Motion With Inertial Sensors," *IEEE Sensors Journal*, vol. 8, no. 8. pp. 1385–1388, 2008.

[26] Y. Stebler, S. Guerrier, and J. Skaloud, "An Approach for Observing and Modeling Errors in MEMS-Based Inertial Sensors Under Vehicle Dynamic," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 11. pp. 2926–2936, 2015.

[27] H. Ahmed and M. Tahir, "Accurate Attitude Estimation of a Moving Land Vehicle Using Low-Cost MEMS IMU Sensors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 7. pp. 1723–1739, 2017.

[28] Z. W. Wu, M. L. Yao, H. G. Ma, and W. M. Jia, "De-noising MEMS inertial sensors for lowcost vehicular attitude estimation based on singular spectrum analysis and independent component analysis," *Electronics Letters*, vol. 49, no. 14. pp. 892–893, 2013.

[29] R. Ferrero, F. Gandino, M. Hemmatpour, B. Montrucchio, and M. Rebaudengo, "Exploiting Accelerometers to Estimate Displacement," *5th Mediterr. Conf. Embed. Comput. 2016*, pp. 206–210, 2016.

[30] ISO, "ISO 2041:2009 Mechanical vibration, shock and condition monitoring -- Vocabulary," 2009.

[31] J. Lyle Bagley, "Vibration Analysis: Methods and Applications," in *Handbook of Technical Diagnostics: Fundamentals and Application to Structures and Systems*, H. Czichos, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 121–136.

[32] C. Scheffer and P. Girdhar, *Practical Machinery Vibration Analysis and Predictive Maintenance*. 2004.

[33] K. Mehlhorn, *Algorithms and Data Structures*. Springer, 2013.

[34] K. R. Rao, D. N. Kim, and J. J. Hwang, *Fast Fourier Transform - Algorithms and Applications*. Springer Science & Business Media, 2011.

[35] K. Kido, *Digital Fourier Analysis: Fundamentals*. Springer Berlin Heidelberg, 2014.

[36] ISO, "ISO 1925:2001 Mechanical vibration - Balancing - Vocabulary," 2001.

[37] A. Boudiaf, A. Djebala, H. Bendjma, A. Balaska, and A. Dahane, "A Summary of Vibration Analysis Techniques for Fault Detection and Diagnosis in Bearing," no. 1, pp. 37–42, 2016.

[38] L. Szabó, D. Fodorean, and A. Vasilache, "Bearing Fault Detection of Electrical Machines Used in Automotive Applications," pp. 2186–2192, 2016.

[39] Y. K. Chaudhari, J. A. Gaikwad, and J. V. Kulkarni, "Vibration analysis for bearing fault detection in electrical motors," in *1st International Conference on Networks and Soft Computing, ICNSC 2014 - Proceedings*, 2014, pp. 146–150.

[40] ISO, "ISO 2954:2012 Mechanical vibration of rotating and reciprocating machinery -- Requirements for instruments for measuring vibration severity," 2012.

[41] ISO, "ISO 13373-3:2015 Condition monitoring and diagnostics of machines -- Vibration condition monitoring -- Part 3: Guidelines for vibration diagnosis," 2015.

[42] ISO, "ISO 10816-8:2014 Mechanical vibration -- Evaluation of machine vibration by measurements on non-rotating parts -- Part 8: Reciprocating compressor systems," 2014.

[43] J. Tsurushiro and T. Nagaosa, "Vehicle localization using its vibration caused by road surface roughness," pp. 146–151, 2015.

[44] R. Hostettler, W. Birk, and M. L. Nordenvaad, "Extended Kalman filter for vehicle tracking using road surface vibration measurements," *Proc. IEEE Conf. Decis. Control*, pp. 5643–5648, 2012.

[45] A. J. Dean, R. D. Martini, and S. N. Brennan, "Terrain-based road vehicle localization using particle filters," *2008 Am. Control Conf.*, vol. 49, no. July, pp. 236–241, 2008.

[46] M. Lindfors, G. Hendeby, F. Gustafsson, and R. Karlsson, "Vehicle speed tracking using chassis vibrations," *IEEE Intell. Veh. Symp. Proc.*, vol. 2016–Augus, no. Iv, pp. 214–219, 2016.

[47] A. Rivas Rodriguez, R. Wunderlich, and S. Heinen, "Road Vibrations as a Source to Detect the Presence and Speed of Vehicles," *IEEE Sens. J.*, vol. 17, no. c, pp. 1–1, 2016.

[48] L. A. Jerry, K. M. Ng, K. Salam, and M. S. A. M. Ali, "Estimating vehicle speed from road surface vibration using exponential regression," *Proc. - 2013 IEEE 9th Int. Colloq. Signal Process. its Appl. CSPA 2013*, pp. 299–303, 2013.

[49] Y. Decoster, P. Orlewski, and R. Ranta, "Detection of human presence in a vehicle by vibration analysis," *IET Intell. Transp. Syst.*, vol. 6, no. 4, pp. 413–

420, 2012.

[50] P. Robertson, W. B. Coney, and R. Bobrow, "Vehicle load estimation from observation of vibration response," *Proc. - Appl. Imag. Pattern Recognit. Work.*, 2010.

[51] M.-H. Bao, *Micro mechanical transducers : pressure sensors, accelerometers, and gyroscopes*. Amsterdam New York: Elsevier, 2000.

[52] C. Silva, *Vibration monitoring, testing, and instrumentation*. Boca Raton, FL: CRC Press/Taylor & Francis, 2007.

[53] Y. Y. Chen, R. S. Chang, K. W. Jwo, C. C. Hsu, and C. P. Tsao, "A non-contact pulse automatic positioning measurement system for traditional chinese medicine," *Sensors (Switzerland)*, vol. 15, no. 5, pp. 9899–9914, 2015.

[54] M. A. Zakaria, H. Zamzuri, R. Mamat, and S. A. Mazlan, "A path tracking algorithm using future prediction control with spike detection for an autonomous vehicle robot," *Int. J. Adv. Robot. Syst.*, vol. 10, 2013.

[55] S. Kumar, J. Paefgen, E. Wilhelm, and S. E. Sarma, "Integrating on-board diagnostics speed data with sparse GPS measurements for vehicle trajectory estimation," *The SICE Annual Conference 2013*. pp. 2302–2308, 2013.

[56] C. G. Prevost, A. Desbiens, and E. Gagnon, "Extended Kalman Filter for State Estimation and Trajectory Prediction of a Moving Object Detected by an Unmanned Aerial Vehicle," *2007 American Control Conference*. pp. 1805–1810, 2007.

[57] S. Kim, J. Lee, and J. Lee, "Trajectory estimation of a moving object using Kohonen networks," *Artif. Life Robot.*, vol. 9, no. 1, pp. 36–40, 2005.

[58] D. Yang, Y. Han, and X. Lian, "Research on rapid measurement of medium short wave longitudinal road profiles," *Proc. - Int. Conf. Electr. Control Eng. ICECE 2010*, vol. 1, pp. 1742–1745, 2010.

[59] HA, "Volume 1 Series 700 Road Pavements -General," 2016.

[60] H. S. Shin, C. Lee, and M. Lee, "Adaptive threshold method for the peak detection of photoplethysmographic waveform," *Comput. Biol. Med.*, vol. 39, no. 12, pp. 1145–1152, 2009.

[61] Mathworks, "Find Peaks in Data." [Online]. Available: https://uk.mathworks.com/help/signal/ug/find-peaks-in-data.html. [Accessed: 14-Nov-2017].

[62] S. L. Horowitz, "A Syntactic Algorithm for Peak Detection in Waveforms with Applications to Cardiography," *Commun. ACM*, vol. 18, no. 5, pp. 281–285, May 1975.

[63] N. Instrument, "Peak Detection Using LabVIEW and Measurement Studio." [Online]. Available: http://www.ni.com/white-paper/3770/en/. [Accessed: 14-Nov-2017].

[64] D. Ratna, "Epoxy Resins," in *Handbook of Thermoset Resins*, 2009, pp. 155–174.