
Victoria J. Hodge, Simon O’Keefe, Michael Weeks, and Anthony Moulds

Abstract—In recent years, the range of sensing technologies has expanded rapidly, whereas sensor devices have become cheaper. This has led to a rapid expansion in condition monitoring of systems, structures, vehicles, and machinery using sensors. Key factors are the recent advances in networking technologies such as wireless communication and mobile ad hoc networking coupled with the technology to integrate devices. Wireless sensor networks (WSNs) can be used for monitoring the railway infrastructure such as bridges, rail tracks, track beds, and track equipment along with vehicle health monitoring such as chassis, bogies, wheels, and wagons. Condition monitoring reduces human inspection requirements through automated monitoring, reduces maintenance through detecting faults before they escalate, and improves safety and reliability. This is vital for the development, upgrading, and expansion of railway networks. This paper surveys these wireless sensors network technology for monitoring in the railway industry for analyzing systems, structures, vehicles, and machinery. This paper focuses on practical engineering solutions, principally, which sensor devices are used and what they are used for; and the identification of sensor configurations and network topologies. It identifies their respective motivations and distinguishes their advantages and disadvantages in a comparative review.

Index Terms—Asset management, condition monitoring, decision support systems, event detection, maintenance engineering, preventive maintenance, railway engineering, wireless sensor networks (WSNs).

I. INTRODUCTION

E XPERTS estimate that the railway industry will receive US$300 billion worth of global investment for development, upgrading, and expansion over the five years from 2009 [43]. Ollier [98] noted that effective management of rail infrastructure will be vital to this development, upgrading, and expansion, particularly if coupled with a move to intelligent infrastructure [39]. A key part of the management will be condition monitoring. Condition monitoring detects and identifies deterioration in structures and infrastructure before the deterioration causes a failure or prevents rail operations. In simple condition monitoring, sensors monitor the condition of a structure or machinery. If the sensor readings reach a predetermined limit or fault condition, then an alarm is activated. However, this simplistic approach may lead to a large number of false alarms and missed failures [36]. It only provides local analysis but does not take advantage of the superior capabilities when the sensors are networked and their data processed collectively. Integrated data processing allows an overall picture of an asset’s condition to be achieved and overall condition trends to be determined [97]. In recent years, networking technologies such as wireless communication and mobile ad hoc networking coupled with the technology to integrate devices have rapidly developed. The new technologies allow vast numbers of distributed sensors to be networked [5], [6], [37], [45], [122] to constantly monitor machines, systems, and environments. Wireless sensor networks (WSNs) [5], [134] are wireless networks of spatially distributed and autonomous devices. They use sensors to cooperatively monitor infrastructure, structures, and machinery. A typical WSN for railway applications is shown in Fig. 1. Each sensor node generally has a radio transceiver, a small microcontroller, and an energy source, usually a battery (see Section II-C for more detail). WSNs and data analytics allow the railways to turn data into intelligence [43]. They provide decision support through continuous real-time data capture and analysis to identify faults [52]. The data from distributed systems such as sensor networks are constantly monitored using classification [56], [57], prediction [85], or anomaly detection [61] to determine the current and future status of the distributed network. Lopez-Higuera et al. [78] developed a staircase of structural health monitoring, where the higher the level, the
higher the complexity and functionality. The simplest Level 1 systems detect the presence of damage without locating it, whereas Level 2 provides location information. A Level 3 system is able to grade the degree of damage and a Level 4 system can estimate the consequences of the damage and remaining service life. Finally, a Level 5 system will comprise complex hardware, custom algorithms and software to allow the diagnosis and/or the prognosis and even to recommend the solution to a problem.

WSN monitoring provides continuous and near real-time data acquisition and autonomous data acquisition (no supervision is required); increased frequency of monitoring compared with manual inspection; improved data accessibility, data management, and data use compared with non-networked systems as all data can be collected and processed centrally; the ability to combine data from a wide variety of sensors; intelligent analysis of data to “predict and prevent” events using intelligent algorithms; the ability to turn data into information about the status of important structures, infrastructure and machinery; and, a global data view that allows trending information to be determined where degradation is happening slowly over a relatively long period of time.

WSN monitoring can be used to:
1) maintain process tolerances;
2) verify and protect machine, systems and process stability;
3) detect maintenance requirements;
4) minimize downtime;
5) prevent failures and save businesses money and time;
6) request maintenance based on the prediction of failure rather than maintenance running to a standard schedule or being requested following an actual failure.

There are a number of challenges with WSNs. They generate large amounts of data at rapid rates and often on an ad hoc basis. Data may be produced from multisources that have to be fused. The systems and structures monitored using sensors often exhibit complex behavior, which is difficult to understand and interpret [102]. Hence, the data must be carefully managed to provide a view of the system status. Sensor data are very noisy and sensors themselves can become defective wherever they are installed. Sensor data may contain errors, particularly where sensor devices may also return additional ambient measurements such as temperature and humidity, and sensor data can be combined with the measurements from other systems. Rabatel et al. [103] noted that the behavior of assets is often affected by external factors. These external factors are encapsulated by contextual data, which describe the ambient running conditions. It is important to take them into account and build models for condition monitoring that consider these ambient conditions. Hence, Elia et al. [44], Rabatel et al. [103], and others incorporated wider ranging contextual data with the sensor data. Elia et al. [44] included train speed, position along the line (longitude and latitude from GPS system), and wind speed and direction when analyzing bogie vibrations, whereas Rabatel et al. [103] included structural criteria (route) and environmental criteria (weather, travel characteristics, travel duration) when analyzing bogie temperatures.

II. SENSOR DESIGN

There are a multitude of sensors types used in railway condition monitoring for analyzing different aspects of structures, infrastructure, and machinery (see Table I in the Appendix). The devices in Table I are transducers. Transducers convert energy from one form to another. In the case of sensors, the device typically converts a measured mechanical signal into an electrical signal. Most railway sensors fall under the umbrella type microelectromechanical systems (MEMS). MEMS are small, integrated devices, or systems that combine electrical and mechanical components. They are cheap and efficient [70]. Sensor design requires a tradeoff between functionality and power consumption, with functionality often coming at the cost of power. Condition monitoring systems for railways are often deployed in remote or inaccessible locations, where there is no wired power supply available. Hence, the sensors must receive power from either batteries or local energy generation. MEMS sensors have the advantage that they can be produced to consume ultralow power [19].

A. Measurement Sensors

Table I provides an overview of the sensor types used in railway condition monitoring and the measurements produced.

Sensor devices may also return additional ambient measurements such as temperature and humidity, and sensor data can be combined with the measurements from other systems. Rabatel et al. [103] noted that the behavior of assets is often affected by external factors. These external factors are encapsulated by contextual data, which describe the ambient running conditions. It is important to take them into account and build models for condition monitoring that consider these ambient conditions. Hence, Elia et al. [44], Rabatel et al. [103], and others incorporated wider ranging contextual data with the sensor data. Elia et al. [44] included train speed, position along the line (longitude and latitude from GPS system), and wind speed and direction when analyzing bogie vibrations, whereas Rabatel et al. [103] included structural criteria (route) and environmental criteria (weather, travel characteristics, travel duration) when analyzing bogie temperatures.
These need to be reliable to allow the WSN to function. Hence, authors have considered alternative energy sources for regions where no access to batteries or wired connection is possible.

1.5 years for 4 AA batteries by only waking their sensors when monitoring the growth of a crack over a period of time, then the microcontroller, memory, and transceiver power can all be low. However, for frequent sampling such as monitoring the temperature of a mechanical bearing every second, then the microcontroller, memory, and transceiver must be more powerful. Lynch and Loh [80] provided a thorough overview of sensor boards for WSNs.

**B. Sensor Nodes**

Sensor devices are mounted on boards. The boards form a platform combining mobile computing and wireless communication (media access control (MAC), routing and error detection) with sensor devices as shown in Fig. 2. Many sensor nodes are moving toward enabling data interrogation and autonomous data processing to identify anomalies in situ. Hence, many data processing algorithms have also been embedded in wireless sensor nodes for autonomous execution [80]. The boards generally comprise one or more wireless sensors, a microcontroller, transceiver, data storage (memory), and a power source. If the data sampling rate is low, for example, when monitoring the growth of a crack over a period of time, then the microcontroller, memory, and transceiver power can all be low. However, for frequent sampling such as monitoring the temperature of a mechanical bearing every second, then the microcontroller, memory, and transceiver must be more powerful.

**C. Sensor Power**

Sensor nodes often use batteries as their energy source. Batteries have a finite life. The batteries in an accelerometer monitoring a train pantograph only lasted 20 days [97]. Chebrolu et al. [32] were able to extend battery life by 1.5 years for 4 AA batteries by only waking their sensors when required and sleeping otherwise. However, even with extended life, fitting and replacing batteries is not always possible in remote and inaccessible railway locations. The nodes may also be embedded in the monitored structure or mechanics of a train, where no access to batteries or wired connection is possible. Hence, authors have considered alternative energy sources for nodes. These need to be reliable to allow the WSN to function for long periods of time with no human involvement and able to generate sufficient power.

Nagode et al. [89], Nelson et al. [92], Nuffer and Bein [96], Pourghodrat [101], Tan et al. [120], and Wischke et al. [130] investigated piezoelectric vibration harvesters that derive energy from the vertical displacement or vibration of tracks, ties, and sleepers. Wischke et al. [130] were able to demonstrate that sufficient energy could be harvested to supply a microcontroller with a radio-frequency (RF) interface. Tan et al. [120] noted that their rate of energy harvesting (≈1 mW) was lower than the typical rate of energy consumption of sensor systems (≈10–100 mW). Hence, they stored the harvested energy using supercapacitors and the sensors only drew energy when sufficient stored energy was available. Nagode et al. [89], Nelson et al. [92], and Tan et al. [120] all used inductive coils for harvesting. Electromagnetic harvesters such as induction coils have a higher power density than piezoelectric [69]. Nelson et al. and Tan et al. used the vertical movement of the track to generate energy. Nagode et al. designed and built two prototypes of a vibration-based electromagnetic (inductive coil) system. They both harvested mechanical energy from the movement of a rail wagon’s suspension springs to power sensors mounted on the wagons. The first prototype used the movement of the springs to generate linear motion in magnets to produce a voltage. The second prototype converted the linear motion of the wagon’s suspension coils into rotary motion then magnified and rectified the motion to turn a generator. Pourghodrat [101] developed a hydraulic energy harvester, which generated energy from compression and relaxation of fluid driving a hydraulic motor. It used the vertical displacement of the track for energy generation. Hydraulic harvesters can generate energy from smaller track deflections compared with electromagnetic, but the devices are still in development [101].

A number of authors have investigated using solar energy harvesting to power sensors. Bischoff et al. [19] investigated a solar energy rechargeable battery-powered base station. Sekula and Kolakowski [111] powered their piezoelectric strain sensors using accumulators permanently charged by photovoltaic modules. Solar power is a proven way to generate energy. However, it has a number of practical drawbacks for WSNs. Solar panels used to generate energy need frequent cleaning for optimal performance [55]. Sunlight varies on a daily basis, and sunlight is very limited in winter months in many global regions and power generation is adversely affected by bad weather. Hence, Wolfs et al. [131] powered accelerometers and gyroscopes using solar power with battery backup and Waki [125] powered a temperature sensor by a solar cell with a double-layer capacitor power supply for back up. These double layer capacitors have a longer life than a battery.

If the sensor nodes can be put into “sleep” mode and only woken when they are required, then their energy usage can be minimized. A number of bridge monitoring systems use this wake-on-event two-layer event detection. It requires two types of sensors: one set have lower power consumption to perform event detection and wake the second set of sensors using interrupts, [19], [32], [46], [70], [73] and [107]. Bischoff et al. [19], Feltrin [46] and Kruger et al. [70] detected approaching trains using ultralow-power MEMS acceleration sensors. These measured the vibrations of the bridge and detected the increased vibrations when a train approached. Similarly, Ledečzí et al. [73]
used standard strain gauges to sense the approaching train and woke the system by detecting changes in stress on the bridge. Sala et al. [107] used activation sensors type not specified) to wake their system for monitoring steel truss bridges. In contrast, Townsend and Arms [123] used a similar technique but with only one set of track-mounted strain gauges, which varied their sampling rate depending whether a train is present. They operated in low sampling mode of approximately 6 Hz to detect the increase in strain when a train approached. Once detected, the sensors increased their sampling rate and generated a strain waveform for data analysis. The authors stated that this reduced the power from 30 mA for continuous mode to less than 1 mA for event detection. Chen et al. [35] introduced a prototype system that developed this further by detecting approaching trains, waking the measuring sensors on the bridge trusses and then downloading the measured data onto the passing train, which acts as a data mule. Another data muling approach is the BriMon system [32], where trains were fitted with a beacon. A bridge-mounted high-gain antenna detected the train’s beacon 30+ seconds before the train approached and woke the battery-operated accelerometers on the bridge. The accelerometers measured the vibrations of the passing train on the bridge and uploaded their data to the passing train. This data muling only requires local data transmission, which reduces the power usage further.

III. NETWORK DESIGN

WSNs enable continuous real-time capture of data. However, WSNs need to be able to handle the harshness of outdoor long-term condition monitoring; often in hostile environments and must minimize energy usage as the nodes are not attached to a wired power supply. They typically use low-power sensors powered by batteries although authors are investigating alternative power supplies such as local energy generation. Hence, the network to enable data capture has to be carefully designed to overcome these factors and prevent transmission errors, latency, network outages, missing data, or corrupted data.

A. Base Station

The base station controls the sensor nodes and acts as a gateway for data transmission to a remote server. The sensor nodes use short-range communication such as Wi-Fi or Bluetooth to transmit data to the base station. The base station uses long-range communication such as GPRS or satellite to transmit collated data back to a server at a control center. It has a more powerful processor and more memory than the sensor nodes to allow it to collate data from multiple sensors. Hence, the base station requires more power. For example, train monitoring systems have battery-powered sensor nodes in the carriages but place the base station in the locomotive, where more power is available from the train engine.

B. Relay Nodes

Sensor nodes are energy-constrained, thus only having a short transmission range. If they cannot reach the base station, then higher capability relay nodes may be used. This is a particular issue in restricted environments such as railway tunnels, along the length of trains or along railway tracks. Relays ensure connectivity by relaying the data signal from the sensor node to the base station via one or more relay nodes using wireless multihop paths. Network nodes can be multifunctional. Houl et al. [62] and Shafiuullah et al. [112] both used sensor nodes that also function as relay nodes to transmit data.

C. Network Model

WSNs for railway applications follow the open systems interconnection model [137]. A typical WSN has five layers in a protocol stack with three planes [5], [6] to transmit data from the sensor nodes to the base station. The five layers from lowest to highest are the physical layer defines how the sensors transmit their data to the network. The data link layer specifies the network topology and connects nodes to each other. It uses MAC to control data access and permissions and a logical link control to control error checking and data packet synchronization. This is a key layer in railway monitoring WSNs as they may be monitoring difficult environments such as tunnels or monitoring moving trains. The topology must be carefully designed. The network layer routes the data through the network as packets. Railway WSNs are energy constrained; thus, routing has to be carefully designed to work within the available energy supply. The transport layer controls the sending and receiving of data. Finally, the application layer allows application software to access the data.

The size of the transmission varies from bits in the lowest (physical) layer, bytes, and frames in the data link layer, packets in the network layer, segments in the transport layer, and data in the highest (application) layer.

The power management plane manages the power consumption of sensor nodes. The mobility management plane identifies and records the movement of sensor nodes to ensure that data can be routed from the node to the control center and so that sensor nodes are aware of their neighbors for communication. This is important for WSNs monitoring moving trains. The task management plane balances and coordinates network regions (groups of nodes) and ensures that they generate data according to their power level.

D. Sensor Network Topology (Data Linkage)

The topology is constrained by the requirements of the monitoring and by the physical environment. Sensor nodes can be arranged in either an ad hoc or a preplanned configuration [134]. In ad hoc arrangement, sensor nodes are randomly placed in the monitored area. In preplanned arrangement, sensor nodes are arranged in either a grid, optimal placement [100] (described next), 2-D or 3-D placement [134].

Determining the optimum node placement is a complex task and often requires a tradeoff. The network configuration can be optimized against a number of different constraints. A network may minimize relay nodes, may need to ensure a minimum level of service (include a certain level of redundancy), minimize energy usage to preserve battery life, or may...
Stajano et al. [116] degraded RF signals. Conversely, the regular, tube-like tunnels within a rail tunnel is different from propagation outside [77], [116]. The complex geometry and rough surfaces of rail environments can also place constraints on the network topology. Signal propagation cannot be solved for all situations. The railway environment can also mean that the reliability versus communication delay tradeoff cannot be bypassed using alternative routes.

The authors concede need to ensure accessibility of the nodes [116]. For example, Pascale et al. [100] optimized a WSN for signaling and control in railway stations to calculate the minimum number of nodes required. The authors had to optimize the tradeoff between reliability and scheduling optimization which allocates nodes while maintaining a maximum level of delay. The authors concede that the reliability versus communication delay tradeoff cannot be solved for all situations. The railway environment can also place constraints on the network topology. Signal propagation within a rail tunnel is different from propagation outside [77], [88], [116]. The complex geometry and rough surfaces of rail tunnels degrade RF signals. Conversely, the regular, tube-like shape can act as an “oversized waveguide” [88]. Therefore, Stajano et al. [116] and Liu et al. [77] developed models to predict the signal strength at each node in a tunnel. Node placement can then be optimized using an algorithm based on parameters such as transmission strength, battery life, and redundancy between the sensor nodes and the base station.

The natural topology choices available for WSNs in rail applications are star, tree, and mesh topologies (see Figs. 3 and 4). In a star topology (single point-to-multipoint), there is a single base station that can send and/or receive messages to/from a number of remote sensor nodes. The remote nodes can only communicate directly with the base station but not with each other. This topology is simple; the point-to-point connection will result in fewer transmissions and fewer collisions; and, it has low communication latency. The disadvantages are the base station must be within transmission range of all nodes, remote nodes must increase their transmission power to overcome signal attenuation and loss, and it is not robust due to its dependence on a single base station to manage the network.

A tree topology is a hierarchy of nodes, with the root node serving client nodes that, in turn, serve other lower level nodes. In a tree, the nodes can be grouped at each level. The network is scalable. Messages pass from the sensor nodes through the tree branches to the root. Sensor nodes can communicate with their parent or with other nodes in a group within transmission range.

In contrast, in a mesh topology, any node in the network can communicate with any other node in the network that is within transmission range. This enables relaying of messages (multihop communication) keeping the transmission power low, saving energy compared with transmitting directly from sensor nodes to base station in a single hop, and allowing nodes to communicate indirectly with nodes outside their transmission range. The network can be easily extended by adding new nodes and also incorporates fault tolerance as failed nodes can be bypassed using alternative routes.

E. Communications Medium

There are many communication techniques used in WSNs in railways [3], [108]. Intersensor communications and sensor to base station transmission are usually short range. The base station transmits gathered data back to the control center, and this requires long-range communication. WSNs can use technologies based on standard mobile telephony (Bluetooth, GSM, GPRS, and UMTS) or broadband techniques such as Wi-Fi (IEEE 802.11, 2007), wireless personal area networks (WPANs) (IEEE 802.15.4 and ZigBee [11]) or WiMax [2] (IEEE 802.16.2, 2004). ZigBee enhances IEEE 802.15.4 by including authentication, data encryption for security, and data routing and forwarding. Mobile telephony and WiFi are peer-to-peer only, whereas WPANs and WiMax can connect many devices in a mesh topology, where any node can communicate with any other node. In addition, WiFi, WPANs and WiMax allow communication at much higher speeds and bandwidths than mobile telephony and are cheaper to set up. However, mobile telephony has better coverage and range than WiFi and WPANs. WiMax can operate over long distances. However, there is a tradeoff: either high bitrates or long distances but not both. An important recent communication advance has been GSM-R (Global System for Mobile Communications-Railway), which is an adaptation of GSM telephony for railway applications. It is designed for information exchange between trains and control centers. Its main advantages are its low cost and worldwide availability. Another common technique for railways is the radio system Terrestrial Trunked Radio, which is a standard for data communication for closed user groups [108]. It is a private mobile telephone system and is easily deployed using a series of antennas at stations or control centers along the railway routes.

F. Transmission and Routing

Sensor nodes use the greatest amount of energy during data communication. The transceiver in the sensor node contains frequency synthesizers, mixers, voltage control oscillators,
phase-locked loops, and power amplifiers [6]. These all consume power. Thus, data transmission and routing has to be carefully designed to minimize the energy usage.

When messages are transmitted through tree or mesh networks, they have to be routed to ensure they pass from source to destination (sensor node to base station). Message latency increases as the number of hops required to transmit the message increases; thus, the routing protocol must be carefully designed for the particular WSN topology. A number of routing protocols have been proposed recently (see [4]). If the base station fails in any topology, then the routing also fails. Thus, when Mahlknecht and Madani [81] used a daisy chain (tree) topology, they ensured that the nodes in the chain could also directly communicate with the global communication network if no base station was accessible. Another important consideration of multihop communication is interference prevention. For train-mounted WSNs, if two trains, Train A and Train B, are within radio communications range of each other, then data from sensors on Train A must never be transmitted to Train B and vice versa [104]. Reason et al. [104] used identifiers to bind the sensors to the individual train’s network.

The sensor nodes need to transmit data reliably via the network despite having only a limited power supply available. Hence, energy efficiency and reliability are both vital for WSNs. One important disadvantage of multihop (tree or mesh), in general, is that the power consumption of the nodes near the base station is much higher as they are used in data transmission to the base station and hence suffer energy drain. A possible solution to minimize energy usage and maximize fault tolerance is to use multihoming. This is used in the Internet and provides multiple connections (multiple routes) between points thus introducing a degree of fault tolerance and ensuring that no extra energy is required to transmit data. Aboelela et al. [1] mapped every node in the network to two clusters. One set of nodes operated as guardians of the clusters and cluster nodes transmitted via two designated cluster guardians. Each cluster guardian thus acted as a relay node and forwarded data from one cluster to the other or to the base station for processing.

Another issue with the ad hoc nature of WSNs is transmission interference. The MAC protocol must minimize both power usage and data broadcasting collisions. WSNs use IEEE 802.15.4, which uses carrier sense multiple access with collision avoidance multiplexing. This allows multiple nodes to access the same channel at different times without interference. However, the 2.4-GHz ISM band used by IEEE 802.15.4 is crowded causing variable and unpredictable transmission. Therefore, Liu et al. [76] investigated frequency diversity (FD) using modified sensors to monitor an underground railway tunnel where the confined environment made data transmission difficult. In FD, the data are transmitted using several frequency channels. Liu et al. [76] found that applying FD improved transmission for WSNs in confined environments such as tunnels. Shafiuullah et al. combined energy efficient MAC protocols [114] with multifunctional nodes [112]. They developed energy-efficient protocols for IEEE 802.15.4/ZigBee multihop (daisy chain) networks mounted in train carriages. The nodes acted as both sensors and routers to both take measurements and forward data to a base station in the train locomotive. One MAC protocol (E-BMA) is optimal for low-medium traffic and the second (EA-TDMA) is optimal for medium-high traffic. In E-BMA, sensor nodes use piggybacking to reserve transmission slots rather than sending control messages to reduce transmission and, thus, reduce energy usage. E-BMA also waits for one extra time step to determine if there are multiple packets to transmit and sends them together. EA-TDMA divides the transmission channel into time slots and allocates slots to sensor nodes to avoid collisions. The nodes wake up in their slot, transmit if they have data or switch off immediately if they have no data, and sleep during the remaining time slots to conserve energy.

Salaberry et al. [108] developed an alternative multitransmission technique using GPRS. They used two phonecards belonging to different telephone providers on board each train to increase the network coverage. The hardware can then switch between networks depending on network availability.

IV. MONITORING SYSTEM

The data from sensors is treated as either a time series, where data are produced continuously or periodically, or a sequence of readings where data is generated ad hoc, for example, generated every time a train passes. The data can be monitored by searching for thresholds (triggers), known problem signatures (classification), identifying unknown events (short-term analysis using outlier detection), or identifying drift over a longer period of time (long-term outlier detection).

Condition monitoring can be performed continuously or periodically. Continuous monitoring should detect a problem straight away but it is often expensive; energy hungry, which is a problem for WSNs where the network components need power; and the sensor data are very noisy, which requires careful preprocessing to ensure accurate diagnostics. Periodic monitoring is cheaper, uses less energy, and allows time for data cleaning and filtering but a problem will only be diagnosed at the next processing run. This may be acceptable for some situations that change slowly such as cracks developing in bridges but for time critical scenarios, then continuous monitoring is necessary.

In basic condition monitoring, the system is only able to distinguish between normal and abnormal conditions (no fault or fault). Beyond this, systems exhibit increasing levels of monitoring sophistication as outlined by the staircase of structural health monitoring of Lopez-Higuera et al. [78] described in Section I. A Level 5 system will comprise complex hardware, custom algorithms, and software to allow the diagnosis and/or the prognosis and even the solutions.

The topology of WSNs often varies over time. One very important factor in this topological variation is the mobility of the sensor nodes. Lorestani et al. [79] subdivided the communications network for their WSN into two: the fixed network relates to sensor nodes in fixed locations such as bridges, tunnels, and special points, whereas the movable network relates to sensor nodes attached to locomotives or rail wagons. The data for the movable network are logged with accompanying GPS coordinates. Movable (on-board) sensors can monitor the whole track length travelled by the train but only monitor the
sections of the train where the sensor nodes are attached [84]. In contrast, track-mounted (fixed) sensor nodes can measure the whole train as it passes but only at specific points where the nodes are mounted on the track. This tradeoff needs to be considered when designing the node placements. The following sections analyze existing WSNs for condition monitoring on the railways and conducts analyzes from the perspective of fixed (static) and movable (dynamic) monitoring [122]. In each category, the systems are grouped according to the object(s) being monitored.

A. Fixed Monitoring

Many authors have investigated health monitoring of infrastructure, including rail bridges, tunnels, rail tracks and track beds, and other track infrastructure. Krüger et al. [70] noted that structural health monitoring is becoming ever more important as both train speeds and the axle loads exerted by trains increase. Until recently, inspection has been performed visually [70], but this only examined the structure superficially and intermittently. Additionally, the visual analysis needs to be interpreted by an expert, which can be subjective. Sensors are objective and can provide data from all of the structure (including internally) to allow the whole structure’s health to be assessed and to analyze its durability and remaining life time. Wired sensors can be used for monitoring. However, wired systems are expensive, inflexible, time consuming to install, and the trains have to be stopped during installation [70]. The early fixed WSNs were simple such as attaching sensor nodes to the rails to monitor the rail temperature or low-voltage warning sensors that monitor the power supply to motors at points [15]. These systems simply generated a binary output (high or low). For the temperature sensor, a rail expert with knowledge of the ambient (stress free) rail temperatures set the alarm levels for “amber” and “red” warnings. If an alarm level was reached, then the system sent an SMS alert. Modern WSNs provide semi-automatic or automatic analysis of the sensor data to examine structural changes and to improve the durability of structures. Hence, WSN monitoring should reduce the overall maintenance costs. An overview of fixed monitoring systems is given in Table II in the Appendix.

1) Bridges: Bridges suffer structural defects exacerbated by the constant strains and vibrations of passing trains. Human inspection of bridges is difficult and much of the structure may be inaccessible. WSNs enable constant monitoring of the whole structure, including the internal structure of the bridge. The monitoring systems often comprise two units, as detailed in Section II-C. The first unit is low power. It detects the approach of a train and wakes the second unit. This second unit generates the measurements, whereas the bridge is under load (a train is crossing) to assess the bridge’s health.

For an overall assessment, Krüger et al. [70] used a variety of MEMS to monitor structural changes in concrete including: piezo-based acoustic emission (AE) sensors for crack and fatigue detection, strain gauges to analyze static loads and stresses, accelerometers to analyze dynamic loads as trains pass, and temperature/humidity sensors to determine the ambient conditions. Grosse et al. [53], [54], Hay et al. [58], Hay [59], Ledeczi et al. [73], and Nair and Cai [90] used AE testing to detect fatigue cracks and other internal structural damage in bridges. The rate and intensity of AE from cracks defines the crack growth rate, which Hay et al. [58] categorized as inactive, active, or critically active. Ledeczi et al. [73] combined AE sensors to detect crack growth with strain gauges to identify changes in the forces exerted by trains both of which indicate structural problems. Bischoff et al. [19] and Feltrin [46] also analyzed force changes using resistance strain gauges to detect structural damage. Similarly, Sala et al. [106] and Sekula and Kolakowski [109] analyzed force changes using piezoelectric strain sensors mounted on the bridge trusses. Kolakowski et al. [67] used ultrasonic strain gauges to record elastic waves in bridge trusses due to vibrations caused by passing trains. Changes in vibrations can indicate structural problems similar to changes in forces. Chebrolu et al. [32] detected these changes in vibration using accelerometers. During monitoring, Kolakowski et al. [67], Sala et al. [107] and Sekula and Kolakowski [111] used an extra module: piezoelectric strain ("weigh in motion") sensors mounted on the track approaching the bridge, which weigh the train; thus, the train’s weight can be factored into the structural assessment.

Kerrouche et al. [65] incorporated optical sensors into the fabric of bridges during bridge repairs to allow structural analysis. They used fiber Bragg grating (FBG) strain sensors with grooves cut in the bridge fabric to house the optical fibers carrying the FBG sensors. The authors concluded that the FBG monitoring needs more thorough evaluation as it was difficult to distinguish the actual strain from the noise.

2) Tunnels: Tunnels can be damp leading to rapid corrosion damage and other defects; thus, monitoring is essential. Monitoring of tunnels has many similarities to bridge monitoring. However, rail tunnels are often difficult to access and inspect. An added complication of tunnel monitoring compared with bridge monitoring is data transmission in a confined environment [88]. Data generated by sensors inside the tunnel have to be transmitted using a suitable relay mechanism (often multiple hop using relay nodes) from the various sensor nodes to the outside. Bennett et al. [17], Cheekiralla [33], Hoult et al. [62], and Stajano et al. [116] described WSNs deployed in the London Underground tunnels. Bennett et al. [17], Hoult et al. [62], and Stajano et al. [116] detected distortion using inclinometers and measured movement using linear potentiometric displacement transducers. They coupled the sensors with four relay nodes in a mesh topology to transmit the sensor data to a receiver unit in the tunnel. There is no GPRS signal near the London Underground tunnels. The authors used repeater units placed along the tunnel to transmit the data from the gateway to a computer located outside the tunnel at the top of a ventilation shaft. This transmission computer logged the data and transmitted them to the offsite computer server using a mobile-phone signal. Bennett et al. also analyzed tunnels in the Prague metro using the same sensor network. Here, a GPRS signal is available at the tunnel mouth; thus, they connected the gateway node to the transmission computer directly using an Ethernet cable with no requirement for repeater units, thus, reducing the chances of data loss. Cheekiralla [33] monitored deformation in the tunnel when construction work was happening nearby using off-the-shelf battery-powered pressure transducers to measure.
vertical displacements and the transverse deformation along the most vulnerable zone of the tunnel. Cheekiralla coupled each pressure transducer with a microcontroller, RF transmitter, analog-to-digital converter, and power conditioning hardware. The base station was located near the tunnel with a RF receiver unit connected to a laptop. The laptop had a wireless modem to periodically send data to a server.

3) Rail Tracks: Track monitoring systems also play a vital role in maintaining the safety of the railways. Monitoring bridges and tunnels (discussed in the previous paragraphs) uses sensors to identify and analyze defects (cracking and displacement) in large structures. In contrast, track monitoring involves identifying and analyzing defects in long narrow metal rails. Tracks can crack and displace like bridges but also twist and tilt (incline). Hence, track monitoring can vary from detecting settlement and twist such as that caused by nearby tunneling or excavation [112]; to measuring the forces exerted by train wheels on the tracks; to monitoring the development of cracks and structural flaws as train passes and over the longer term.

Durham Geo Slope Indicator Company (DGISC) [112] and Aboelela et al. [1] analyzed the track movements (shift) over time particularly where nearby construction work may affect the track or where track passes through vulnerable areas such as regions prone to landslides. Aboelela et al. [1] mounted inclinometers parallel with the tracks directly on the sleepers or in the ballast with continuous tensioned rails to monitor tilt. Likewise, DGISC [112] mounted inclinometers parallel with the tracks to measure settlement. They also mounted inclinometers perpendicular to the track (on the ties) to monitor twist. Aboelela et al. [1] also plan to use inclinometers to monitor ties to detect broken ties or sabotage.

Other systems analyzed the effects of trains passing along the tracks using track-mounted strain gauges [10, 71, 94, 99], piezoelectric strain sensors [112], FBG strain sensors [41, 72, 118, 119, 127], AE sensors [7, 22, 121, 128, 135], and accelerometers [10, 18, 71], Current sensors readings can be compared with historical sensor measurements and unusual readings or readings that match known problem signatures can be detected [7, 22, 41, 72, 118, 121, 127]. Strain gauges monitor the forces exerted by trains (“weigh in motion”), and accelerometers measure vibrations (motion of the track [9]), whereas AE sensors detect and analyze cracks. Tam et al. [119] installed FBG strain sensors on each side of a track to detect imbalances on the two sides of train wheels. If there is a large variation between the left and right axle loading then there is the danger of train derailment.

Monitoring the impact of passing trains is important for two reasons: it can detect track damage, but it can also detect damaged trains. Brickle et al. [23] detailed state-of-the-art automated systems using track-mounted sensors for wheel condition monitoring. Excessive forces, particularly where the train has a defect such as out-of-round wheels, flat wheels, or unbalanced axle loading, can crack sleepers, damage the rail head, and cause failure of rail by either growth of fractures, or fatigue and fracture of in-track welds [10]. Filograno et al. [47] used FBG sensors for train identification, axle counting, speed and acceleration detection, wheel imperfection monitoring, and dynamic load calculation by placing them in different positions on the track. Monitoring the forces exerted using strain sensors can identify defective axles and wheels [118] or incorrectly loaded trains [41]. Analyzing the track vibration data produced by accelerometers [16] and the AE signal transferred through the rail [7, 22, 121] from AE sensors can also detect out of round wheels as trains pass. Belotti et al. [16] incorporated an inductive axle-counter block into their system for assessing the train speed. The speed affects the expected vibrations; thus, the speed can be built into the wheel analysis model. Detecting wheel faults can also pinpoint trains that are at risk of causing a derailment. In [10], [71], the risk of derailment is represented by the ratio between the lateral and the vertical force (L/V ratio), which quantifies the relationship between the applied load and the deflection of the track [10].

Condition monitoring protects both the trains and the track, increases the track and train reliability and allows repair to be scheduled. If the track has been marked as defective and awaiting repair but is still usable, AE sensors can be attached to monitor the defects and ensure that they are not getting worse [135]. Track awaiting repair usually has a speed restriction enforced so, if it is monitored, this may not be required.

4) Rail Track Beds: As well as monitoring the track, it is important to analyze the rail bed [8]. The pore-pressure of water in the subsoil of rail beds increases as trains pass. Repeated pressuring can reduce the shear strength of the soil and increase the potential for bed failure. Wong et al. [132] used vibrating wire piezometers to measure this pore pressure. Vibrating wire piezometers are accurate and reliable. However, they have a limited reading frequency due to the vibrating wire. Wong et al. [132] generated data at a rate of 0.1 Hz and Hendry [60] notes that this low rate of data acquisition may miss the peak pore pressures under axle loads which occur at frequencies of >1 Hz. Konrad et al. [68] measured pore pressure using piezometers and wire potentiometers which generate data at > 100 Hz. This higher frequency allows the analysis of pore pressure during the passage of individual trains. Other systems use a variety of sensors for monitoring. Aw [8] analyzed a railway site experiencing mud pumping problems that requires frequent (bimonthly) track maintenance. Aw used accelerometers to monitor the track vibrations, settlement probes to monitor track settlement, piezometers to measure ground water pressure, and temperature sensors. Hendry [60] used piezometers to measure the pore pressure and detect changes; extensometers to measure the relative vertical motion at various depths; and, an arm comprising jointed segments with accelerometers attached. The movement of the segments and overall shape of this arm is used to analyze the horizontal deformation of the foundation material during the passage of a train. Baum et al. [14] developed a WSN early warning system for rail landslides. It combined data from rain gauges to determine the rainfall level and likely saturation, tensiometers to measure the pore pressure and reflectometers to measure the soil water content. Tensiometers are able to detect changes in pore pressure rapidly to warn of slope instability, whereas reflectometers detect water content changes over time to warn of wet soil slippages.

5) Other Track Infrastructure: A variety of infrastructure is used to operate the tracks. Monitoring the condition of this infrastructure is an important part of railway condition...
monitoring. Fishplates are metal plates attached to the end of track sections to join and align them [51]. Fishplates can be tampered with, incorrectly maintained or the nuts and bolts may work loose. This may cause track sections to become misaligned potentially resulting in a derailment. Buggy et al. [27] and Ghosh et al. [51] proposed WSNs to monitor fishplates using FBG stain sensors [27] and piezo-resistive pressure sensors [51], respectively. If the nuts and bolts of a fishplate are loosened, then the strain and pressure on them change. The WSN can send an alert to the nearest control room.

The railways use track circuits which detect approaching trains and ensure correct train and signal operation [36], [49], [133]. Track circuits transmit a signal to a receiver via the two running rails of a track section [36]. If the signal reaches a receiver at the far end of the track section, then no trains are present. If a train is present on the section, then the signal is short circuited by the wheels, the signal does not reach the receiver and the track relay is not energized. Fukuta [49] noted that this receiver can be a wireless MEMS. Detecting current flows along the track can also pinpoint broken rails or disconnected wires. Railway points switch (move) the track between two alternative routes to ensure the train takes the correct route [63], [83], [86], [136]. The points work in conjunction with track circuits to detect and route trains. Hence, both the track circuits and the points must be fully functioning to accurately detect and route trains to ensure safe and reliable operation. Points monitoring systems often use a variety of sensors as they require multifaceted monitoring. Zhou et al. [136] monitored: distance, motor driving force, driving current and voltage, electrical noise, temperature and state changes, whereas Marquez et al. [83] measured a similar set of variables: voltage, current drawn by the electric motor, current drawn by the system as a whole, and the force in the drive bar measured with a load pin. Igarashi and Siomi [63] used magnetic lock-warp sensors to detect maladjusted switches and points irregularities by measuring displacement through temperature variations.

B. Movable Monitoring

WSNs are able to monitor the condition of a range of mechanics, systems, and environments using on-board sensors to measure parameters such as temperature, shocks, tilts, and humidity [138]. Thus, WSNs demonstrate a promising platform for real-time movable condition monitoring of transport systems, allowing the early detection and diagnosis of problems. Bruni et al. [24], Ngigi et al. [95] and Ward et al. [126] provided overviews of the opportunities available for on-board vehicle monitoring systems, which analyzed the mechanics. Measurements can all be taken during normal train service that allows in-service condition monitoring of the train chassis and mechanics and the rail track that the train is running on. The sensor nodes used for movable monitoring are often mounted on the train bogies or carriages. These locations are susceptible to high levels of vibration; thus, the sensor nodes have to be rugged or protected. Authors use strong adhesives to bond the nodes to the object being monitored and sealed protective cases to shield the sensor nodes. For example, Grudén et al. [55] used sensors to monitor train bogies. They used shock absorbing material around circuit boards to protect them, separated the battery from the circuit board, mounted the antenna separately from the circuit board, and used flexible cables.

One of the key issues with movable sensor monitoring is communication, due to the mobility of the sensors. Movable rail monitoring can transmit the data from the movable nodes to a static node when the static node is in range. Transferring data from the movable nodes to a static node often happens when a train stops in a station. Alternatively, the movable nodes can form a network with a base station within the network. If the nodes form a network on the train, then the base station is often mounted on the locomotive, where power is available to power the data transmission. This transmits the data over a suitable mechanism, such as satellite or GSM. Despite the issues with movable monitoring, mobile sensor nodes can achieve a higher degree of coverage and connectivity compared with static sensor nodes. As they move to a range of locations.

Object mobility (such as the movement of sensors) can be classified into uncontrolled and controlled mobility [13]. With uncontrolled mobility, a sensor node can move randomly and freely. In contrast, a node only moves along a predetermined route with controlled mobility. Vehicles on the railways can only move in limited trajectories, each train follows a planned route along tracks, which are set by points and signals and the movements are dictated by the rail timetable. Establishing WSNs for controlled mobility is easier compared with WSNs for uncontrolled mobility. If the routes are known and fixed, then the nodes and receivers can be arranged for maximum efficiency as wireless sensor nodes have limited transmission ranges. However, Chen et al. [34] noted that even with controlled mobility, trains can be over a mile long. If these trains are fitted with sensor nodes, then the off-train receivers have to be carefully designed and positioned to cope. An overview of movable monitoring systems is given in Table III in the Appendix.

1) Train Engines, Passenger Carriages, and Freight Wagons: Wagon or container failure could lead to damage of the rail infrastructure or the environment and could even result in loss of life; thus, it is important that the wagons, containers, and their respective contents are also monitored. Additionally, freight trains use unpowered and unwired railroad wagons, thus an on-board WSN can analyze the condition of the freight wagons [104] without power from the train. A number of authors have developed systems for analyzing train carriages, wagons, and engines. This can involve analyzing the mechanics of the vehicles or analyzing the contents where safety-critical material is being transported. Many of these systems form a mobile network in a daisy chain topology between adjacent carriages and with the engine acting as the base station [50], [81], [104], [112], [131]. The sensor data can also classify the rail wagon’s response to the track and even detect potential derailments. Ward et al. [126] noted that on-board sensors can identify track defects including small vertical irregularities (dipped joints or voids), small lateral irregularities (kinks), larger vertical irregularities, larger lateral irregularities (misalignment), and track twist.

A number of authors analyzed the mechanics of carriages and engines using parameters such as the temperatures, strains,
and forces at critical locations on the rail wagon shell and mechanics. Gao et al. [50], Nejikovsky, and Keller [93] and Tam et al. [119] used WSNs to monitor the dynamics and structural condition of both locomotives and passenger carriages, whereas Shafiuullah et al. [112] and Wolfs et al. [131] monitored carriages. In [93], engine sensors analyzed the health of locomotive engines by measuring parameters, including water temperature, oil temperature, engine temperature, switch statuses, fans statuses, throttle positions, rotations per minute, and traction motor currents. Similarly, Tam et al. [117] used FBGs to analyze the strain and temperature of train body shells in critical locations, which include the welding joints, cross beams, and sole bars. Accelerometers can analyze different properties of the wagons depending where they are mounted [50], [93], [131]. Wolfs et al. [131] placed them at the edges of rail carriages; Gao et al. [50] placed them on the floor of the locomotive and the chassis of carriages; and Nejikovsky and Keller [93] mounted them on the body of passenger carriages of tilting trains to monitor rail wagon body motions. These train-mounted accelerometers can measure vertical and horizontal vibrations [50], [131]. They can also monitor the lateral accelerations when travelling around curves to ensure the lateral forces do not exceed government safety levels [93]. Gyroscopes [131] can measure the longitudinal accelerations, as well as pitch, roll, and yaw of carriages to analyze motion.

Casola et al. [31], Flammini et al. [48] and Maly et al. [82] proposed to integrate data from a range of heterogeneous sensors for more holistic condition monitoring. Maly et al. [82] performed hot box detection, flat wheel detection, dynamic scaling, and loading gauge detection using commercially available sensors, whereas Casola et al. [31] analyzed temperature, humidity, and vibrations among others (the specific sensor types are not stated). The sensors are grouped into units, which operate independently. If a unit detects a faulty train, then it alerts the control center and provides detailed information about the detected irregularity.

Wagons containing freight are susceptible to tampering and intrusion. Authors have combined accelerometer data with light, shock, humidity, and temperature readings to provide security monitoring [81], [104], [112]. Mahlknecht and Masani [81] used vibration shock sensors; Mahlknecht and Masani [81] and Shafiuullah et al. [112] monitored the humidity to detect changes; Mahlknecht and Masani [81] and Reason et al. [104] used light sensors to detect wagons that are open, and Mahlknecht and Masani [81] used motion detectors for analyzing movement and detecting irregularities.

2) Train Bogies: The train bogies perform a number of tasks, mainly to guide the train on both straight and curved tracks, to safeguard the train’s dynamic stability, and to provide a comfortable ride for the passengers [126]. Most commercial condition monitoring systems for rail vehicles analyze the bogie system. Ngigi et al. [95] noted that this is because the bogie’s important components undergo rapid changes during journeys and can pose safety related issues. Train axles and bogies run hot during train journeys. However, excess heat is generated as the axles and bogies wear. It is imperative to detect this excess heat to allow early detection of wear and deterioration, schedule maintenance and thus prevent vehicle failures and derailments. Hot axle bearing detectors (HABDs) have been in use for a number of years [98]. The early HABDs could detect defects but were only capable of generating an alarm once a certain threshold was reached [98]. In contrast, new systems generate a significant amount of data, both as alarms and as time-series graphs of the bearing temperatures. Different heat types have different time-series signatures [98]. Hence, it is possible to identify faults such as hot wheels caused by brake drag or extraneous heat sources such as leaking engine exhaust fumes. Previously, when only a simple threshold was used, alarms caused by these other sources of heat may have been misinterpreted as false positives, due to the lack of detailed information [98]. Rabatel et al. [102], [103] and Carrascal et al. [30] used MEMS temperature sensors; Kim et al. [66] used surface acoustic wave (SAW) sensors mounted on the train’s bogies to detect overheated wheel bearings and Grudén et al. [55] used temperature sensors to analyze wheel performance. Grudén et al. [55] used three sensors to measure bogie temperatures and a fourth sensor to measure the air temperature to establish the ambient conditions. The ambient conditions affect the train mechanics, thus the need to be incorporated into the condition monitoring to prevent false positives and false negatives.

Bogies can also be analyzed for vibrations. Elia et al. [44], Mei and Li [87] and Tsunashima et al. [124] proposed systems for the measurement of bogie vibrations. Elia et al. and Tsunashima et al. mounted accelerometers on the bogies and Elia et al. also placed them on axle boxes of train carriages to measure the lateral acceleration. Mei and Li used robust inertia sensors mounted onto the bogie frame to monitor vibrations and ground speed. The papers also demonstrated that rail/wheel conicity caused by incorrect track gauge or inclination [44] and track irregularities [87], [124] can be identified from the respective bogie-based sensors. Tsunashima et al. [124] were able to analyze the sensor data onboard the train at source to provide real-time track analysis.

3) Train Axles: As well as generating excess heat, excess strains (such as torsional and bending stress) may be exerted on axles causing deterioration. The degree of torsional stress is affected by the train’s traction control system, mechanical brake characteristics, the wheel–rail interface and the track adhesion conditions. This number of factors affecting the stress means that establishing a baseline to create a model necessitates measuring the torsional axle stress levels over a long term, ideally over the four seasons of a year [42]. Similarly, the axle bending stress levels vary due to the track quality, vehicle loading, braking and traction effects and curving speeds. Deuce and Rosinski [42] also observed strong electromagnetic interference (EMI) when measuring strains on powered axles. This required the development of bespoke instruments, including strain gauge amplifiers, telemetry systems, solid state data loggers, and signal processing units equipped with wireless data communication [42].

4) Train Wheels: A number of the systems already discussed for track and train vehicle monitoring are able to detect wheel irregularities. Train wheels wear over time and can also suffer flat wheels and out-of-round wheels. The previous section (Section IV-A fixed monitoring) discussed how WSNs
attached to the track detected faulty wheels by analyzing vibrations and stresses exerted by the wheels on the tracks. Wheel problems can also be detected by sensors on board the train analyzing the wheel surfaces or analyzing forces on the train bogies and wheels. The wheel-sets and bogie each have two degrees of freedom (lateral and yaw) [75]. Wired sensors could be used for analysis but require holes in the train axle for the power leads, which weaken the axle, whereas wireless sensors are not very durable in the harsh conditions [84]; thus, this lack of durability has to be considered.

Nuffer and Bein [96] introduced a wear detection system for train wheels. It detected changes in the vibration behavior of the entire wheel caused by surface changes on the rolling contact area. The system positioned piezoelectric sensors on distinct areas of the wheel. The sensors represented the surface as an electrical signal. Changes in the surface can then be detected by changes in the electrical signal. However, it proved difficult to adequately define the correlation between the behavior of the wheel and the measured signals; thus, the system was sensitive to small changes in model parameters. Li et al. [75] noted that the behavior models for rail monitoring are very complex with strong interdependencies between the measured variables and between different degrees of freedom. Thus, a number of variables need to be measured. They mounted sensor nodes on the bogie frame to monitor the dynamic behavior of wheel profile (conicity) and dampers. Two accelerometers mounted on the front and rear wheel set measure the lateral accelerations of the wheel set. An accelerometer and a gyrometer mounted on the bogie measured the lateral acceleration and yaw velocity of the bogie. Finally, a vehicle body accelerometer measured the lateral acceleration of the vehicle body.

It is also important to monitor the wheel/rail contact to detect skidding and wear. Matsumoto et al. [84] applied a multifaceted approach. They fitted noncontact gap sensors on nonrotating parts of a bogie to measure the bending deflection (distortion) of the wheel by the lateral contact force. They also used magnetostrictive displacement sensors to measure the deflection of the primary suspension by the vertical contact force. These two measurement combined give the wheel/rail contact forces. Matsumoto et al. [84] used the same derailment coefficient as [10] and [71] (discussed in the fixed monitoring section): the ratio between the lateral and the vertical force (L/V ratio). Balas and Jain [9] fused data from speed sensors and piezoelectric accelerometers attached to train wheels. Using data modeling, they were able to detect skidding if the integral of the acceleration (speed of the wheel) measured by the accelerometer differed from the speed of the carriage. They also calculated the wheel adherence as a function of the carriage weight, carriage velocity, brake cylinder pressure, and the skid amount from the skid model.

5) Train Brakes: Heavy haul transportation involves carrying heavyweight railway freight. The trains have multiple locomotives, which must coordinate their braking. Song et al. [115] developed a WSN for brake coordination and monitoring to reduce the maintenance and inspection requirements. They attached air pressure sensors to the locomotives and train rear, installed relay nodes on the carriage roofs, and placed accelerometers and ultrasonic sensors along the track along with railway relay nodes. When the front locomotive brakes, the sensors in it detect the change of air pressure, and the relay node on the locomotive transmits a message to the other locomotives via the relay nodes on either the train or the track using a wireless radio link. The other locomotives can then apply their brakes according to the message from the lead locomotive. The track relay nodes are used where transmission via train carriages is weak such as bends and tunnels. The data regarding the railway infrastructure from the accelerometers and ultrasonic sensors along the track can be collected and analyzed for condition monitoring.

6) Train Pantographs: In electrified railways, the overhead power lines and the pantographs on the trains require monitoring. The pantograph ensures that the train contacts the overhead power line to collect current and power the train. Ensuring the integrity of the pantograph, the pantograph/overhead line interface and the overhead power lines is important to ensure the power supply to the train, prevent breakdowns and identify sections of power line requiring maintenance. Thus, monitoring the force exerted between the contact and the wire is essential to detect and prevent problems.

Boffi et al. [21], Bruno et al. [25], Elia et al. [44], Waki [125], and Willsch et al. [129] all performed condition monitoring of the pantograph itself. Elia et al. also monitored the overhead power lines for irregularities and loss of tension. Optical sensors are immune to EMI making them ideal to use near power lines where EMI is high. Thus, Boffi et al. and Willsch et al. measured the strain exerted between the pantograph’s current collector and the overhead wire using interferometric fiber [21] and FBG [129] sensors. Elia et al. and Waki used a variety of sensors to analyze different aspects of the pantograph. Both used accelerometers that can detect impact accelerations that occur when passing a contact failure and both integrated sensors in the current collector. Elia et al. used load cells to measure the reaction forces and displacement transducers to measure the motion of the current collector and articulated frame. Waki along with Bruno et al. attached ultraviolet (UV) sensors to the current collector to detect arcing. Arcing occurs when the contact between the pantograph and the overhead wire is lost and causes power transmission loss and overheating. To detect overheating further, Waki measured the pantograph’s temperature using a standard temperature sensor. Waki used a video camera to record overhead line conditions and pantograph movement. Leung et al. [74] used passive magnetoelectric current sensors to analyze the current in the transmission cable running from the pantograph to the train’s power unit. Changes in the current flow can indicate power transmission loss and need to be detected.

V. CONCLUSION

This paper has reviewed the range of WSNs used for condition monitoring in the railway industry. The emphasis is on the practical engineering solutions, principally which sensors devices are used and what they are used for; and, identification of sensor node configurations and network topologies.

Until recently, railway inspection has been visually performed [70], but this only examines objects superficially and
intermittently, and the analysis needs to be interpreted by an expert, who can be subjective. Sensors are objective and can provide data from the entire object (including internally) to allow the whole object’s health to be fully assessed and to analyze its durability and remaining life time. A broad range of sensors are used in railway monitoring to provide an extensive range of data and allow monitoring of different structures, vehicles and machinery. The main challenge for WSNs in railway applications is determining the best measurement technologies to use. The WSN must be reliable and accurate to enable effective condition monitoring in harsh and inaccessible environments but must also be cost effective. It must be possible to translate the sensor data from the WSN into relevant and clear information to enable decision support in the railway infrastructure maintenance lifecycle.

The paper divides railway condition monitoring into fixed monitoring for immobile infrastructures such as bridges, tunnels, tracks and associated equipment, and movable monitoring for vehicles and their mechanics. Fixed monitoring uses sensors to monitor vibrations, stresses and sound waves passed through structures (acoustics) caused by passing trains (short-term monitoring) and also changes in stresses, pressures and sound waves passed through structures over the longer term (long-term monitoring).

One of the key issues for fixed monitoring is network topology. The topology is constrained by the requirements of the monitoring and by the physical environment. Sensor nodes can be arranged in either an ad hoc or a preplanned configuration [134]. Determining the optimum node placement is a complex task and often requires a tradeoff. The network configuration can be optimized against a number of different constraints. A network may minimize relay nodes, may need to ensure a minimum level of service (include a certain level of redundancy), minimize energy usage to preserve battery life, or may need to ensure accessibility of the nodes [116]. The communication mechanisms, for example, Bluetooth, Wi-Fi, GSM, or satellite also need to be evaluated to ensure coverage and reliability and the routing protocol to ensure data is successfully transmitted from sensor node to base station to control center. Another issue is powering the sensors as fixed monitoring often requires placing sensor nodes in inaccessible locations, for example, in tunnels, on bridge trusses, or in rail track beds. Many sensors use batteries but replacing batteries in inaccessible locations may not be possible. Authors have considered ambient energy harvesting such as converting vibrations caused by passing trains into energy or using solar power. These energy harvesters may be accompanied with energy storage such as capacitors. Authors have also reduced sensor energy usage using event detection where a single cheap sensor detects approaching trains and wakes a larger sensor network to commence taking measurements of a bridge structure.

Movable monitoring uses sensors to monitor temperatures, vibrations, stresses, forces, currents, and sound waves in vehicles and their mechanics particularly bogies and chassis. These locations are susceptible to high levels of vibration and EM; thus, the sensors have to be rugged or protected. One of the key issues with movable sensor monitoring is communication due to the mobility of the sensors. Movable rail monitoring either transmits the data from the movable nodes to a static node when the static node is in range, or the movable nodes form a network with a base station within the network which transmits the data over a suitable mechanism, such as satellite or GSM.

This paper focuses on the sensor technology used to generate condition monitoring data to enable practical condition monitoring systems. These data must be managed and turned into useful information to generate useful information. Condition monitoring systems must store large quantities of data to build models for analysis. The data must be validated first to ensure that they are correct and error-free (sensor faults, noise, null values, communication errors, etc.). This process may even be performed in the sensor node’s microcontroller; thus, only valid data are transmitted thus minimizing the transmission load. These data can then be processed in a number of different ways to generate information. Once data are collated, they can be analyzed using robust algorithms to identify faults in near real-time. Algorithms need to be robust as WSN data is noisy, can be intermittent, may contain errors, has many interdependencies and the data volume is very high. The WSN data can also be stored and analyzed over longer time periods to identify long-term progressive faults such as a slowly developing crack.

Some systems incorporate contextual data that describe the ambient conditions, which will affect the object monitored. These contextual data can be built into models to improve the coverage and accuracy of the model and to help provide explanation of condition monitoring decisions. Elia et al. [44] included: train speed, position along the line (longitude and latitude from GPS system) and wind speed and direction when analyzing bogie vibrations, whereas Rabatel et al. [102], [103] included structural criteria (route) and environmental criteria (weather, travel characteristics, travel duration) when analyzing bogie temperatures.

VI. Future Research

There are a number of promising directions for further research in condition monitoring in the railways. One future direction is a move toward holistic integrated systems [31], [48], [82] which provide near real-time information and alerts. These approaches will integrate data from different sensor systems using sophisticated modeling techniques. They will also incorporate more contextual data into the modeling, including the ambient conditions, the route, the journey time, the weight of the train, and more. The models will use multistage data fusion. This fuses measurements from a variety of sources (sensors and contextual data) and over a range of time epochs to generate a consolidated state history of the object being monitored [38]. The models will also mitigate data dependencies (physical dependencies) across the physical objects being monitored. An integrated multifaceted approach should improve prediction quality. For example, train monitoring data can be combined with route data and GPS data. If multiple trains detect a vibration fault at exactly the same GPS location, then the fault is more than likely in the rail infrastructure such as uneven track. However, if only one train identifies a vibration anomaly at a particular GPS location, then the fault is more likely to be in the rolling stock [64].
As sensors become cheaper and easier to fabricate, their availability and use will increase. Bukova and Svecova [28] proposed a WSN comprising many tiny wireless MEMS (smart dust), which are able to collect, process, and transmit data from light to vibrations. Vast numbers of these sensors can be placed inside a wagon, for example, where they can...
measure temperature, humidity, or vibrations. Bukova and Svecova posited that recent developments in sensor fabrication techniques will produce MEMS that are the size of a grain of sand yet contain computing circuits, bidirectional wireless communications technology and a power supply.

The number of sensors will increase in the future, and the number and distribution of monitored objects will expand. This will require further development of energy efficient routing and transmission mechanisms to ensure reliable and consistent communication. Condition monitoring will also tend to parallel and distributed processing across multiple nodes to speed execution. This could be processing the data in parallel at the same geographical location using a single machine with multiple processing cores or at the same geographical location using multiple compute nodes or even distributed processing, where the data are processed at multiple geographical locations, and the results are assimilated at a central location.

As WSNs expand and become more pervasive, security against intrusion, eavesdropping, data tampering and unauthorized control will be ever more important [116]. WSNs pose unique problems for security. WSNs broadcast data between nodes making them vulnerable. The sensors are energy constrained with limited processing capability; thus, any security software needs to be carefully designed. Network administrators will have to identify which WSNs need to be secured, how they need to be secured, and what level of security is required [117]. The type of security is important. Does the WSN need to be secured for access? How often and to what level should the WSN be monitored for intrusion detection? Does the data need protecting from data tampering? The level of security will be dictated by the value of the asset, the type of asset and the consequences of a failure. WSNs will become more interlinked; thus, it is imperative that WSN security cannot be compromised through access from other networks.

In the future, ever more data will be collected from railway infrastructure and vehicles. This inevitably leads to a requirement for the railway industry to develop standards for data collection [126] and processing. West et al. [110] proposed developing Web Ontology Language standardized ontologies to encapsulate railway data and ensure consistency for transmission and processing. Different systems and different companies can then share data using a consistent interface.

As more and varied condition monitoring systems are developed, then there is a further requirement of standardization of presentation of decisions and information across these systems for consistency and to allow integration of multiple systems. Equally, many condition monitoring algorithms are “black boxes” providing little or no explanation of decisions [110]. In the staircase of structural health monitoring of Lopez-Higuera et al. [78], a Level 4 system can estimate

<table>
<thead>
<tr>
<th>Object monitored</th>
<th>Measurement</th>
<th>Sensor</th>
<th>Citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bridge</td>
<td>Crack/Fatigue Detection</td>
<td>Acoustic Emission</td>
<td>[53] [54] [58] [59] [70] [73]</td>
</tr>
<tr>
<td></td>
<td>Stresses</td>
<td>Strain Gauge</td>
<td>[19] [46] [70] [123]</td>
</tr>
<tr>
<td></td>
<td>Piezoelectric Strain Gauge</td>
<td>[107] [111]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fiber Bragg Strain Gauge</td>
<td>[65]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ultrasonic Strain</td>
<td>[67]</td>
<td></td>
</tr>
<tr>
<td>Weight (of Train)</td>
<td>Strain Gauge</td>
<td>[106]</td>
<td></td>
</tr>
<tr>
<td>Vibrations</td>
<td>Dynamic Load</td>
<td>Accelerometer</td>
<td>[32] [70] [107]</td>
</tr>
<tr>
<td>Tunnel</td>
<td>Structure Distortion</td>
<td>Inclinometers</td>
<td>[17] [62] [116]</td>
</tr>
<tr>
<td></td>
<td>Structure Movement</td>
<td>Displacement Transducers</td>
<td>[17] [62] [116]</td>
</tr>
<tr>
<td></td>
<td>Vertical Deformation</td>
<td>Pressure Transducer</td>
<td>[33]</td>
</tr>
<tr>
<td></td>
<td>Transverse Deformation</td>
<td>Pressure Transducer</td>
<td>[33]</td>
</tr>
<tr>
<td>Track</td>
<td>Crack/Fatigue Detection</td>
<td>Acoustic Emission</td>
<td>[7] [22] [121] [128] [135]</td>
</tr>
<tr>
<td></td>
<td>Out of Round Wheel</td>
<td>Acoustic Emission</td>
<td>[7] [22] [121]</td>
</tr>
<tr>
<td></td>
<td>Accelerometer</td>
<td>[16]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stresses</td>
<td>Strain Gauge</td>
<td>[9] [67] [71] [94] [99] [111]</td>
</tr>
<tr>
<td></td>
<td>Piezoelectric Strain Gauge</td>
<td>[67] [109]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fiber Bragg Strain Gauge</td>
<td>[41] [47] [72] [118] [119] [127]</td>
<td></td>
</tr>
<tr>
<td>Vibrations</td>
<td>Dynamic Load</td>
<td>Accelerometer</td>
<td>[10] [18] [17] [71]</td>
</tr>
<tr>
<td></td>
<td>Settlement and Twist</td>
<td>Inclinometer</td>
<td>[112]</td>
</tr>
<tr>
<td></td>
<td>Incline</td>
<td>Inclinometer</td>
<td>[1]</td>
</tr>
<tr>
<td>Rail bed</td>
<td>Dynamic Acceleration (Track)</td>
<td>Accelerometers</td>
<td>[8]</td>
</tr>
<tr>
<td></td>
<td>Pore Pressure (Ground Water)</td>
<td>Piezometers</td>
<td>[8] [60] [68] [132]</td>
</tr>
<tr>
<td></td>
<td>Pore Pressure (Ground Water)</td>
<td>Tensiometers</td>
<td>[14]</td>
</tr>
<tr>
<td></td>
<td>Pore Pressure</td>
<td>Wire Potentiometers</td>
<td>[68]</td>
</tr>
<tr>
<td></td>
<td>Long Term Settlement (Track)</td>
<td>Settlement Probes</td>
<td>[8]</td>
</tr>
<tr>
<td></td>
<td>Temperature</td>
<td>Temperature Sensors (Thermistor)</td>
<td>[8]</td>
</tr>
<tr>
<td></td>
<td>Vertical Motion</td>
<td>Extensometer</td>
<td>[60]</td>
</tr>
<tr>
<td></td>
<td>Water Content</td>
<td>Reflectometer</td>
<td>[14]</td>
</tr>
<tr>
<td></td>
<td>Movement / shape (of rail bed)</td>
<td>Accelerometers mounted in arm</td>
<td>[60]</td>
</tr>
<tr>
<td>Track infrastructure</td>
<td>Pressure</td>
<td>Piezoresistive Pressure Sensors</td>
<td>[51]</td>
</tr>
<tr>
<td></td>
<td>Strain</td>
<td>Fiber Bragg Strain Gauge</td>
<td>[27]</td>
</tr>
<tr>
<td></td>
<td>Displacement</td>
<td>Magnetic</td>
<td>[63]</td>
</tr>
</tbody>
</table>
The consequences of the damage and remaining service life. However, only a Level 5 system will allow the diagnosis and/or the prognosis and recommendation of a solution to the problem. Future systems will need to provide clearer and more detailed explanation of their recommendations to increase usability.

In the future, condition monitoring will be able to exploit cheaper, more robust and more pervasive hardware. New data processing techniques will generate more accurate, robust, and reliable models from existing and new sensor data. Data and processing will be standardized. Systems will be secure and provide clear and detailed decisions and recommendations.

APPENDIX

See Tables I–III.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their helpful suggestions, which improved the quality of this review.

REFERENCES


Michael Weeks received the B.Eng. (Hons) degree in computer engineering and the M.Phil. degree in electronics from University of Teesside, Middlesbrough, U.K., in 1997 and 2006, respectively. Since 2000 he has been a Research Associate with the Department of Computer Science, University of York, York, U.K., working on several research projects related to neural networks. He is currently working on two projects: a web-based infrastructure for condition monitoring across sensor networks and an internationally recognized web-based virtual laboratory for neuroscientists. Between 1997 and 2000 he was a Research Assistant with University of Teesside in the area of media server architectures. Mr. Weeks received the Sony Prize for Electronics while studying at University of Teesside. He was also a member of the research group at York that won the Outstanding Engineering Research Team of the Year award at the Times Higher Education Awards 2011.

Anthony Moulds received the B.Sc. (Hons) degree in electronics engineering from University of Kent, Kent, U.K., in 1984 and the M.Sc. degree in information technology from University of York, York, U.K., in 1990. After graduating in 1984, he spent four years as an Electronics Design Engineer with a subsidiary company of Westland Helicopters, U.K., and later was with the Department of Computer Science, University of York, in 1990 as an Experimental Officer. He is currently a Senior Experimental Officer within the Department. He has particular expertise in the design and development of novel research hardware with a focus on high-speed circuit design and field-programmable gate array implementation. Mr. Moulds is a member of the Institution of Engineering and Technology, U.K.