Original Article

Structural Health Monitoring
2021, Vol. 20(6) 3252–3264

DOI: 10.1177/1475921720982885
journals.sagepub.com/home/shm

Nonlinear modulation with low-power sensor networks using undersampling

Peter Oppermann[1], Lennart Dorendorf[2], Marcus Rutner[2] and Christian Renner[1,3]

Abstract
Nonlinear modulation is a promising technique for ultrasonic non-destructive damage identification. A wireless sensor network is ideally suited to monitor large structures using nonlinear modulation in a cost-efficient manner. However, existing approaches rely on high sampling rates and resource-demanding computations that are not feasible on low-cost and low-power sensor network devices. We present a new damage indicator that uses the short-time Fourier transform to derive amplitude and phase modulation with less computational effort and memory usage. Evaluation of the proposed method using real experiment data exhibits performance and reliability similar to the conventionally used modulation index. Undersampling is demonstrated, which reduces the memory demand in a test scenario by more than 100 times, and the required energy for sampling and processing more than four times. The loss of accuracy introduced by undersampling is shown to be negligible.

Keywords
Vibro-acoustics, nonlinear modulation, structural health monitoring, wireless sensor network, low power

Introduction

Nonlinear acoustic methods for non-destructive testing exhibit higher sensitivity to damage than linear methods. Among the nonlinear techniques, vibro-acoustic modulation (VAM) is a promising candidate. Since the 1990s, several research groups have exploited VAM to detect fatigue damage before visible cracks appeared.[1–4]

However, several hurdles still exist before VAM can be deployed to monitor real structures outside the laboratory. One of these is the reliability of the method: the underlying mechanisms that produce the nonlinear modulation (NM) are not yet fully understood, and the soundness of the damage assessment is not yet ready for safety-critical infrastructure. A second problem is the practical implementation of the method. On complex structures, such as bridges, measurements have to be carried out at many locations to cover all fundamental structural elements.

A sensor network implementing VAM could hence leverage the potential of the method. However, using cables for power supply and communication increases the deployment cost of such a network considerably. If batteries are used to power the sensor nodes, power consumption must be kept as low as possible to reduce the battery replacement interval. Not relying on batteries at all is an even preferable option. Energy harvesting allows sensor nodes to generate the required power themselves from ambient energy such as vibration, wind, or sunlight. Such self-powered wireless sensor networks (WSNs) have been demonstrated successfully.[5–7] However, the authors also stress that available computation and communication resources are minimal.

Several studies investigate the potential of vibrational energy harvesting in detail[8,9]. The amount of energy that a sensor node can expect to harvest at a typical bridge is in the order of a few milliwatts, which only allows for very simple signal processing and a few hundred bytes of data transmission with low-power hardware.

[1]Institute smartPORT, Hamburg University of Technology, Hamburg, Germany
[2]Institute for Metal and Composite Structures, Hamburg University of Technology, Hamburg, Germany
[3]Institute of Computer Science, Research Group Practical Computer Science, University of Koblenz-Landau, Koblenz, Germany

Corresponding author:
Peter Oppermann, Institute smartPORT, Hamburg University of Technology, Am Irrgarten 3-9, Hamburg 21073, Germany.
Email: peter.oppermann@tuhh.de

Table 1. Typical characteristics of low-power wireless sensor nodes compared with modern desktop computers.

	WSN node	Desktop computer
Memory	0.1–10 kB	4–16 GB
CPU frequency	1–16 MHz	2–4 GHz
Power requirement	0.1–10 mW	100–400 W
Network speed	0.1–50 kbits^{-1}	100–1000 MBs^{-1}

WSN: wireless sensor network.

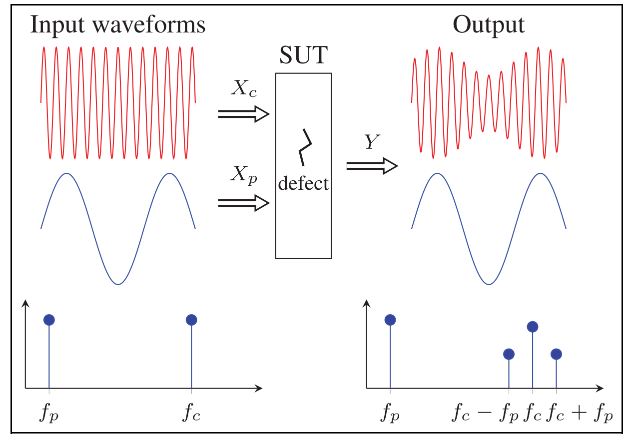

Figure 1. Nonlinear modulation principle as proposed by Donskoy,[12] and Lim and Sohn.[13]

Wireless sensor nodes performing the NM technique were presented by Liu et al.[10] and Yang et al.[11] Both used expensive and energy-demanding hardware and bulky batteries to realize long-term monitoring. The former uses a high-performance microcontroller with 216 MHz and an external 256 Mbit RAM to store the samples for processing. Energy harvesting is used to prolong the lifetime up to 1.5 years, with one measurement every 3 weeks. The latter study is similar, but additionally employs a field-programmable gate array (FPGA) to reduce the power demand during processing.

Further reduction of the method's computational complexity and memory demand is necessary to enable the use of simpler, cheaper hardware, and battery-free devices running purely on harvested energy. Simultaneously, lower energy demand reduces the interval between measurements and yields a higher temporal resolution. Typical properties of low-power devices suitable for self-powered WSNs are compared to desktop computers in Table 1.

The sampling rates and signal durations required to calculate the commonly used modulation index (MI) rely on fast processors and large amounts of memory. Neither sampling at this rate nor transmitting the acquired samples for remote processing is feasible on a low-power sensor node, given their constrained resources.

Recent research suggests that splitting the VAM signals into amplitude modulation (AM) and frequency modulation (FM) might increase sensitivity to damage. Different techniques have been used to achieve the separation of both modulation types: the Hilbert transform (HT),[3] the Hilbert–Huang transform,[1] and the in-phase/quadrature homodyne separation (IQHS) algorithm.[2] However, all three methods require even higher computational power than the conventional MI. In this work, we present techniques to implement VAM capable of running with significantly lower energy usage and smaller memory footprint, but still produce comparable results. Our contributions are as follows:

- investigation of the short-time Fourier transform (STFT) as an alternative to HT for computing AM and phase modulation (PM) separately;
- a new damage index that can be computed from PM and AM and performs similar to the established MI; and
- the demonstration of heavy undersampling to drastically reduce required sampling rates and consequentially also reduce computational complexity, hence rendering the implementation of VAM in a low-power WSN feasible.

In the remainder of this work, we first revisit VAM and explain the conventionally used MI. We then describe how to separately assess AM and PM using the STFT and combine them in a new damage index. Then, we demonstrate how the narrow bandwidth of the VAM signal can be leveraged to recover AM and PM with lower than usual sampling rates. This is followed by a comparison of the performance of the proposed damage index with the established MI on real data and an investigation of the errors introduced by undersampling. Finally, we present a hypothesis indicating that AM and PM are not exclusively caused by different physical mechanisms, but may be two different effects of the same underlying mechanism.

VAM and the MI

During an NM measurement, two sinusoidal acoustic waves, X_p and X_c, with distinct frequencies f_c and f_p are introduced into a specimen under test, where $f_p \ll f_c$. The lower frequency f_p is often referred to as the *pumping frequency*, while the higher frequency f_c is referred to as the *carrier* or *probing frequency*. Then, the resulting ultrasonic wave, Y, is measured at a different location of the specimen. The process is depicted in Figure 1.

Many studies show that the received signal Y is becoming increasingly modulated by the pumping signal when defects in the material accumulate. The exact physical mechanisms causing modulation have not yet been fully understood. However, existing studies suggest that fatigue damage can be detected before it grows into an open crack.

The modulation occurs in both the form of AM and PM/FM. In the amplitude spectrum of Y, AM manifests as two sidebands appearing at frequencies $f_c \pm f_p$. PM also appears as sidebands, but can produce a potentially unlimited amount of sidebands at $f_c \pm k \cdot f_p$, where $k \in \mathbb{N}^+$. In a signal containing both PM and AM, the two forms cannot be separated easily in the frequency domain alone.

The intensity of the modulation has to be tracked over the lifetime of the structure under test. A standard approach to measure this modulation intensity is to compute the amplitude spectrum of Y using the Fourier transform. Then, the amplitude of the first sidebands at frequency components $f_c \pm f_p$ and the amplitude of the carrier at f_c are compared. The modulation is often assessed using the MI, in decibel, defined as

$$\mathrm{MI} = 20 \cdot \log_{10}\left(\frac{A_{f_c - f_p} + A_{f_c + f_p}}{2A_{f_c}}\right) \tag{1}$$

where A_f is the amplitude in the spectrum at frequency f. The MI and other very similar metrics have been used in many studies.[4,12,14]

While this is sufficient to represent pure AM intensity accurately, the MI cannot assess AM and PM individually. It does not precisely represent a total modulation since it neglects the information given in the additional sidebands produced by PM.

Note that the resulting waveform can take different shapes than the one shown in Figure 1 depending on the specific situation. For example, Lee et al.[15] show that—for a large crack—no contact between the crack's surfaces is given under tension, and hence, no energy can be transmitted during part of the low-frequency period. However, even in this situation, similar sidebands appear in the frequency domain.

Required resources for MI computation

The MI calculation using the frequency spectrum requires high sampling rates and considerable processing power. As an example, assume $f_c = 200\,\text{kHz}$ and $f_p = 10\,\text{Hz}$. Frequencies in this range are used in several studies.[2] The minimal required sampling frequency is then 400 kHz.

The frequency components of the carrier and the sidebands must be clearly distinguishable in the spectrum. That requires a high resolution of the spectrum.

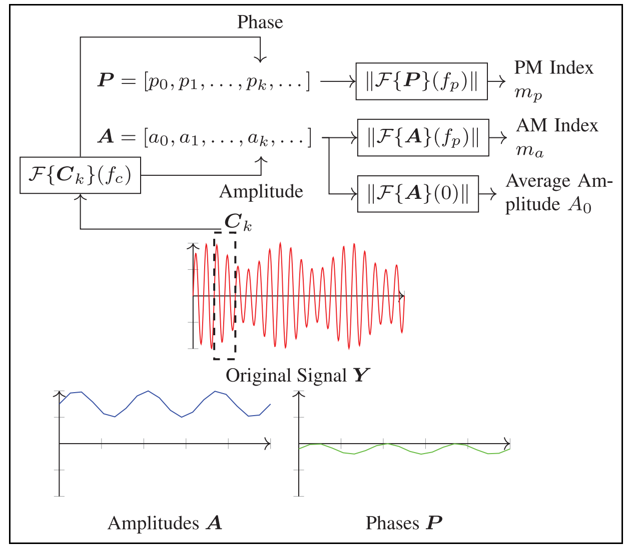

Figure 2. Schematic description of the separation of amplitude and phase modulation from the original signal with the short-time Fourier transform.

When using the discrete Fourier transform (DFT), the frequency resolution Δf is given by

$$\Delta f = \frac{f_s}{n_s} \tag{2}$$

where n_s is the number of recorded samples. Hence, to achieve a certain frequency resolution, say, 2 Hz, where the sidebands can be clearly distinguished, using the minimum sampling rate of $f_s = 400\,\text{kHz}$, at least $n_s = 200,000$ samples need to be recorded. The typical analog-to-digital converter (ADC) samples the voltage with 14- to 16-bit resolution. Hence, the required amount of memory to store all samples of a single measurement is roughly 400 kB. This drastically exceeds the available RAM on WSN devices (Table 1).

Even if abundant memory is available, the execution time of a fast Fourier transform (FFT) with that many samples will be in the range of minutes. Performing intense computations directly on the device is not possible on an energy budget given by batteries or even energy harvesting. Furthermore, this amount of data cannot be transmitted for remote computation since radio transmission also requires considerable energy. Hence, more manageable ways to compute the modulation intensity have to be employed to perform VAM in a WSN.

Separation of AM and PM

We propose a modulation intensity classifier that calculates AM and PM individually using the STFT and then combines both to yield a single damage indicator, comparable to the MI (Figure 2). This new damage indicator is a precondition for undersampling with very low sampling rates, which will be presented in the next section.

PM and FM

FM and PM are very similar. If the modulating signal consists of just a single frequency component, pure FM and pure PM are indistinguishable.[16] With a modulating signal

$$s(t) = \cos(\omega_p t) \tag{3}$$

a phase-modulated sinusoid is given as

$$S_{\mathrm{PM}} = A_0 \sin\left(\omega_c t + m_p \cos(\omega_p t) + \phi_0\right) \tag{4}$$

with PM index m_p. A frequency-modulated sinusoid is given as

$$S_{\mathrm{FM}} = A_0 \sin\left(\omega_c t + \omega_c m_f \cdot \int_0^t s(\tau)\mathrm{d}\tau + \phi_0\right) \tag{5}$$

In the given case, that $s(t)$ is again a sinusoid, equation (5) can be rewritten as

$$S_{\mathrm{FM}} = A_0 \sin\left(\omega_c t + \omega_c \frac{m_f}{\omega_p} \cdot \sin(\omega_p t) + \phi_0\right) \tag{6}$$

which differs from the definition of the phase-modulated signal from equation (4) only in two ways: the intensity of the PM is now dependent on the pumping and probing frequency, and the modulation signal is phase shifted by $\pi/2$. For practical purposes in VAM, however, we observe the evolution of FM/PM over the specimen's lifetime at a single pumping frequency. Hence, the change to the signal's phase either originates from PM or FM, or a mixture of both.

Previous work has regarded any modulation to the phase of the signal as pure FM. However, the methods, such as the HT or the IQHS algorithm, cannot distinguish between FM and PM. We will consider all modulations to the phase as pure PM and all modulations to the amplitude as pure AM in the remainder of this work. To motivate this choice, we provide a wave propagation model in the later section, which shows that PM must be expected in the signal.

PM creates several sidebands in the power spectrum. The conventional MI considers only the first sideband, where PM cannot easily be distinguished from AM. We choose to use the STFT, in which the whole signal is split into shorter chunks

$$\boldsymbol{C}_k = \left[y_{kn_c}, y_{kn_c+1}, \ldots, y_{(k+1)n_c-1}\right], \quad k \in [0, K] \tag{7}$$

where n_c is the number of samples contained in each of the K chunks. If n_c is chosen small compared to the pumping signal's period, the content of each chunk is approximately a pure sinusoid with the carrier frequency f_c—the modulation within the chunk is negligible. Hence, we can safely assume that the carrier frequency is the dominating frequency component within the chunk.

The phase of the carrier in each chunk can then be calculated using the DFT. This yields a new signal $\boldsymbol{P} = [p_0, p_1, \ldots, p_K]$, where

$$p_K = \arg(\mathcal{F}\{\boldsymbol{C}_k\}(f_c)) \tag{8}$$

is the phase calculated from chunk $\boldsymbol{C}_k$. When the carrier's phase offset changes over time, this can be observed in $\boldsymbol{P}$. Finally, the Fourier transform of $\boldsymbol{P}$ is used to retrieve the amplitude of the PM in the spectrum at the pumping frequency f_p

$$m_p = \left\|\mathcal{F}\{\boldsymbol{P}\}(f_p)\right\| \tag{9}$$

Now, m_p approximates the PM index in $\boldsymbol{Y}$.

Note that the chunk size n_c has to be chosen appropriately so that every chunk contains an integer number of periods of the carrier signal. Otherwise, each chunk's phase will be shifted by a constant amount from each chunk to the next. Given the sampling rate of the original signal f_s and the carrier frequency f_c, the chunk size must fulfill

$$n_c = l\frac{f_s}{f_c} \in \mathbb{N}, \quad \text{where } l \in \mathbb{N} \tag{10}$$

The smallest possible chunk size is given when $l = 1$. If the sampling frequency is not divisible by the carrier frequency, then a bigger l must be chosen to accommodate an integer number of periods in each chunk

$$l = \frac{f_c}{\gcd(f_s, f_c)} \tag{11}$$

where $\gcd(a, b)$ denotes the greatest common divisor of two numbers a and b. Hence

$$n_c^{\min} = \frac{f_s}{\gcd(f_s, f_c)} \tag{12}$$

Any multiple of $n_c^{\min}$ is, of course, possible. However, larger chunks will lead to fewer samples of the phase measurements $\boldsymbol{P}$. Finally, the sampling rate of phase measurements $f_s^p = f_s/n_c$ has to be high enough that we can compute the modulation caused by the pumping frequency f_p. According to the Nyquist theorem, $f_s^p \geqslant 2f_p$. Hence

$$\frac{f_s}{\gcd(f_s,f_c)} \leqslant n_c \leqslant \frac{f_s}{2f_p} \tag{13}$$

Amplitude Modulation (AM)

An amplitude-modulated signal can be written as

$$S_{\mathrm{AM}} = \left[A_0 + m_a \sin\left(2\pi f_p t\right)\right] \sin(2\pi f_c t + \phi_0) \tag{14}$$

where m_a is the amplitude modulation intensity. The amplitude of a signal over time is referred to as the *envelope* of a signal. In the previous section, the instantaneous phase (IP) was approximated using the STFT. The envelope can also be recovered in the same way as the phase.

Using the same chunks $\boldsymbol{C}$ as in the previous section, the vector $\boldsymbol{A} = [a_0, a_1, \ldots, a_K]$ of amplitudes per chunk $\boldsymbol{C}_k$ can be calculated with

$$a_k = \|\mathcal{F}\{\boldsymbol{C}_k\}(f_c)\| \tag{15}$$

The sampling rate required for the envelope only depends on the pumping frequency. According to the Nyquist–Shannon sampling theorem, the minimum required sampling rate would be at least $2f_p$. Otherwise, the same restrictions on the chunk size n_c apply as for the PM.

We can now estimate the intensity of the AM with pumping frequency f_p by computing the amplitude of the frequency spectrum of the envelope at f_p

$$m_a = \left\|\mathcal{F}\{\boldsymbol{A}\}\left(f_p\right)\right\| \tag{16}$$

In the conventional MI, the modulation intensity is normalized by the signal strength of the carrier. We assume the carrier amplitude A_0 to be the average of the amplitude vector $\boldsymbol{A}$. The amplitude of the Fourier transform at frequency zero yields the constant offset, and hence the average, as

$$A_0 = \|\mathcal{F}\{\boldsymbol{A}\}(0)\| \tag{17}$$

Combining AM and PM

With the mechanisms from the previous sections, we can derive the PM and AM independently. However, experiments on real specimens, that we present in section "Evaluation," show that if only PM or only AM is inspected over the lifetime of a specimen, their trend is not conclusive. Therefore, no clear damage assessment can be drawn from each of them individually.

However, we can combine the independent measurements of PM and AM to create a single indicator tracking the strength of combined modulation over time. The intensity of AM is measured as the amplitude of the fluctuation of carrier amplitude. PM is measured in the amplitude of the fluctuation of the carrier's phase (in radians). We can construct a meaningful combination of both geometrically.

A pure sinusoidal signal can be represented mathematically as

$$S_{\mathrm{pure}} = A_0 \cdot e^{j\cdot(2\pi f_c t + \phi_0)} \tag{18}$$

where j is the imaginary unit, A_0 is the amplitude of the sinusoidal wave, f_c is its frequency, and ϕ_0 is its phase offset. In the modulated signal, both the amplitude of the sinusoidal signal (AM) and its phase (PM) change over time

$$S_{\mathrm{mod}} = \left(A_0 + m_a \sin\left(2\pi f_p t\right)\right) \cdot e^{j\cdot\left(2\pi f_c t + \phi_0 + m_p \cdot \sin\left(2\pi f_p t\right)\right)} \tag{19}$$

where m_p is the PM amplitude in radians and m_a is the AM amplitude in volts.

We define the combined modulation intensity m_c as the maximum distance between the modulated and the hypothetical unmodulated carrier, normalized by the amplitude of the unmodulated carrier

$$m_c = \max_t \frac{\left\|S_{\mathrm{pure}} - S_{\mathrm{mod}}\right\|}{\left\|S_{\mathrm{pure}}\right\|} \tag{20}$$

which equals

$$m_c = \frac{\|(A_0 + m_a)\cdot e^{jm_p} - A_0\|}{A_0} \tag{21}$$

Practically, we estimate A_0 from the average amplitude of the modulated carrier. Figure 3 visualizes this relation geometrically in the complex plane.

Comparison of STFT with the HT

The standard approach for splitting a signal into instantaneous amplitude (IA) and IP is the HT, which has been used in some recent work on VAM.[1,3] The STFT

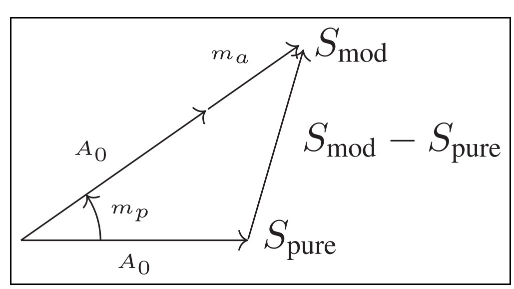

Figure 3. Visualization of the combination of AM and PM in the complex plane. The modulation strength is defined as the maximum difference of a modulated carrier S_{mod} and an unmodulated carrier S_{pure}.

is not equivalent to HT, but it is an approximation that is sufficient for VAM, where the pumping frequency is much smaller than the probing frequency, that is, the modulation happening within a single chunk during STFT is negligible.

We have verified that the STFT-based method produces comparable results to the IA and IP computed with HT, both with simulated and experimental data. The vectors $\boldsymbol{A}$ and $\boldsymbol{P}$ computed by the STFT correspond to the envelope and the IP, respectively. However, it is computationally more advantageous to use the STFT for several reasons.

We use the STFT to reveal the amplitudes and phases in the carrier's frequency bin. The HT, however, reveals all changes to phase and amplitude. Intense filtering has to be applied to mitigate the effect of noise or ambient vibration. Alternatively, the Hilbert–Huang transform is sometimes used, which first splits the signal into empirical modes and then applies the HT to them individually.[1]

The fastest algorithm to compute the discrete HT is applied in the frequency domain by first using an FFT to compute the complex frequency spectrum, perform a simple element-wise vector multiplication in the frequency domain, and then transfer the result back into time domain using the inverse FFT.[17] With the STFT, however, no inverse transformation is necessary.

The HT produces IP and envelope vectors with a sampling rate equal to the original signal. Since the pumping frequency is much lower than the carrier frequency, such high sampling rates are not required to derive the low FM and unnecessarily increase memory usage. With the STFT, the number of elements in the vectors $\boldsymbol{A}$ and $\boldsymbol{P}$ is reduced significantly depending on the chunk size.

When the STFT is used in the case of VAM, we are only interested in very few frequency components. Therefore, not even a full FFT has to be computed in our approach. Instead, the Goertzel algorithm[18] can be used to derive the required frequency components individually, which greatly reduces the computational effort.

Moreover, the STFT processes incoming data in chunks. Each chunk can be processed in a pipeline already during the measurements. The full number of samples never has to be kept in memory simultaneously.

Undersampling

The Nyquist–Shannon theorem constructs a relationship between the bandwidth of a signal and the required sampling frequency to reconstruct this signal. In many applications, the bandwidth is assumed to be equal to the maximum frequency in the signal. This assumption leads to the often formulated requirement: the sampling frequency must be at least twice the maximum frequency in the signal.

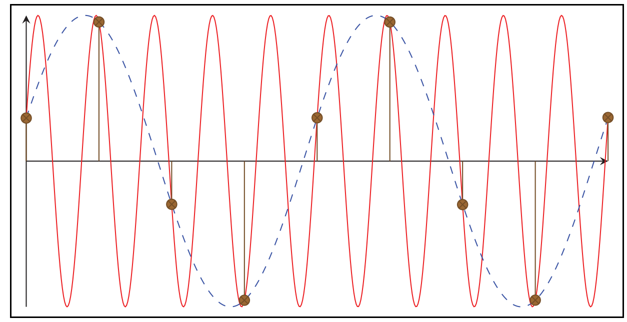

Figure 4. The red, solid line shows a sinusoidal signal. If it is sampled with a frequency below the Nyquist rate, the resulting samples appear to come from a sinusoid with a lower frequency. This alias is shown as blue, dashed line.

However, in the case of VAM, the bandwidth of the signal is actually very small: the frequencies of interest cover just a small range around the carrier frequency. Hence, even with much lower sampling rates, precise reconstruction of the original signal is possible.

Practically, if a signal is sampled with a sampling rate f_s, the DFT can only contain frequency contributions up to $f_s/2$. This range is called the baseband. If a signal contains a frequency component $f > f_s/2$, the DFT of that signal contains a component f' in its baseband, that is an alias of the higher frequency f. The amplitude and phase of that alias component are equivalent to the phase and amplitude of the original signal. Figure 4 illustrates how an alias appears from a higher frequency.

Unfortunately, several different frequencies f can map to the same f' in an undersampled signal. Therefore, an essential requirement for correctness in undersampling is that the signal does not contain more than one frequency component, that maps to the same alias f'. A necessary condition for this is that the bandwidth of the signal is smaller than half the sampling frequency. Figure 5 depicts the mapping of the frequency band in the frequency domain.

The narrow band around f_c that is used during a VAM measurement allows for drastic undersampling. However, a bandpass filter should be used to reduce the amount of noise and higher-order harmonics. Appropriate bandpass filtering guarantees that no two frequency components in the signal can map to the same alias frequency.

Using undersampling theorems for unsymmetric spectra,[19] we can calculate the necessary sampling frequency to map the carrier to a known frequency bin in

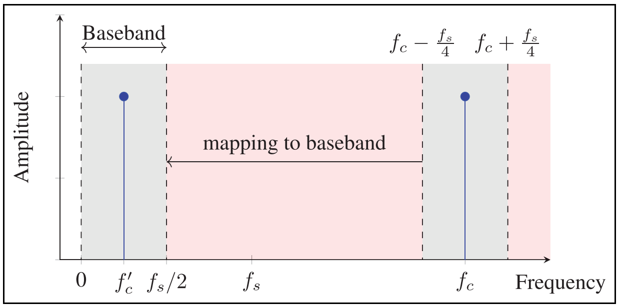

Figure 5. Undersampling a bandlimited signal is equivalent to mapping the frequency band into the baseband. Using the described constraints for the sampling frequency, the carrier frequency alias will appear at the center of the baseband. The signal must not contain any frequency component exceeding the band around f_c to guarantee an unambiguous mapping.

the baseband. The carrier frequency f_c is mapped to $f_c' = f_s/4$ in the baseband, if the sampling frequency f_s is chosen as

$$f_s = \frac{4f_c}{4r+1}, \quad \text{where } r \in \mathbb{N} \tag{22}$$

Note that undersampling can be performed with sampling frequencies that do not match equation (22); however, matching this equation simplifies further processing, because the alias always appears in the center of the baseband at $f_s/4$.

The higher r is chosen, the lower is the resulting sampling frequency. At the same time, the width of the frequency band B, that can accurately be reconstructed, shrinks with higher r. It follows from the Nyquist–Shannon that

$$r \leqslant \frac{f_c}{2B} - \frac{1}{4} \tag{23}$$

For practical purposes, the r-value is also bounded by the data acquisition system. Small jitters in timing or inaccuracy in sampling frequency can lead to large errors in the reconstructed signal when the ratio of undersampling is large.

The same methods for computing the phase vector $\boldsymbol{P}$, the envelope vector $\boldsymbol{A}$, and the resulting damage index m_c can be used as described earlier using the undersampled signal. Also, the same limitations on chunk size and the minimum required number of chunks apply. The only difference is that the carrier frequency f_c now has to be replaced by its alias at $f_s/4$ in all computations.

Evaluation

In this section, we will first show that the proposed damage indicator m_c performs comparably to the MI with real experimental data. In the second step, we investigate how much error is caused by applying undersampling before calculating m_c.

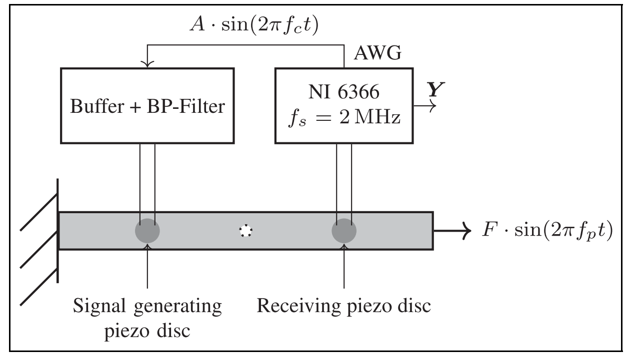

Figure 6. Schematic experiment setup. The pumping frequency is introduced with a tensile testing machine, while the high frequency is generated with a piezo disk.

AM, PM, and the combined damage indicator

To evaluate the performance of the proposed combined modulation intensity m_c, we re-evaluated data from existing experiments with real specimens. We then compare the modulation over the specimen's lifetime measured with the conventional MI and with the proposed m_c. We also inspect the AM and PM individually. In this step, no undersampling is applied. Equivalent experiment setups have been used in existing studies, and the MI has shown to be sensitive to fatigue damage accumulation, showing an increase after 70%–80% of the fatigue lifetime.[20] Strong increases of the MI occur in the presence of microcracks in the range of tens to hundreds of μm, which have been verified with electron microscopes.[12] In this work, we do not use additional measurement techniques to verify damage presence or size, but we compare the results from the proposed method to the MI.

The experiment was conducted with three aluminum specimens with dimensions 300 mm$\times$20 mm$\times$3 mm as shown in Figure 6. A 4.5-mm notch in the middle of the specimen was used to predetermine the fatigue failure's cross-section. Two disk-shaped piezoceramic transducers (PZTs) were applied with epoxy to generate the ultrasonic signal and measure it, respectively. At one piezo, the probing signal is introduced using the arbitrary waveform generator of an NI 6366 from the National Instruments. To drive the piezos with sufficient power, a custom designed buffer and bandpass filter was used. The signal on the other piezo is measured using an input channel of the NI 6366.

The specimen was mounted in a tensile testing machine, which mechanically generated the pumping

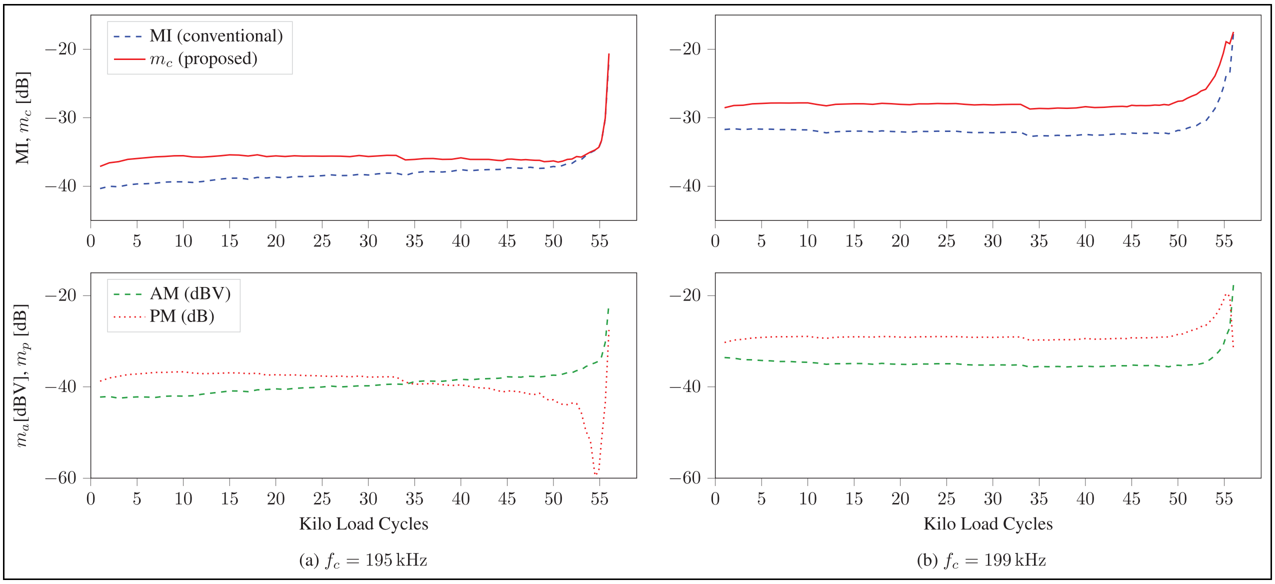

Figure 7. Comparison of different modulation intensity indicators. The graphs show the evolution of the damage indicators over the specimen's lifetime for two different carrier frequencies. The proposed m_c is very similar to the conventional MI despite the lower complexity in calculation. Amplitude and phase modulation intensities alone are not conclusive. For better comparability, the phase modulation index m_p is plotted in dB relative to 1 radian.

signal. At the same time, one of the piezos generated the probing signal. For the probing signal, 10 different frequencies were chosen, ranging from 190 to 199 kHz. This range is within the effective working range of the piezos and provided large amplitudes at the receiving piezo.

We chose 10 Hz as pumping frequency, since the resonant vibration modes of most target structures, such as bridges, are similarly low. For example, we conducted vibration measurements on the Köhlbrand Bridge in Hamburg, which reveal peak vibration around 14 Hz. Studies on other bridges report similar resonant vibration modes at 8 Hz[5] or even 3 Hz.[9] Such ambient vibration induced to the structure by traffic and wind can be leveraged to avoid the artificial generation of the pumping frequency. By choosing a tensile fatigue load as low-frequency excitation, it is ensured that the resulting stresses always act orthogonally on the surfaces of the fatigue crack. In addition, the same pumping frequency has been chosen in many experiments by Donskoy and Ramezani.[2]

The tensile testing machine applied a periodic force on the specimen to introduce fatigue damage in between the measurements. The signal on the receiving PZT was sampled with 2 MSample s^{-1}. The experiment was conducted for each specimen until the specimen broke.

We computed the conventional MI using the sidebands in the frequency spectrum as in equation (1), and the proposed damage index m_c as well as AM and PM individually. Figure 7 shows the different modulation indicators over load-cycles for two different frequencies. For both frequencies, the MI exhibits the typically observed strong increase in modulation intensity, starting at roughly 85%–90% of the lifetime.

Inspecting AM and PM at 195 kHz reveals that during the experiment, m_a increases steadily, while m_p decreases significantly. However, the combined modulation intensity remains relatively constant until the exponential increase starts toward the end of the lifetime, and correlates well with the MI.

At 199 kHz, m_c and the MI also correlate strongly. However, AM and PM do not exhibit the same tendency as at the previously discussed frequency. The same observations apply to all frequencies that we have inspected on all three specimens: the combined MI m_c correlates well with MI, but m_a and m_p seem not to follow any consistent trend. For some frequencies, m_a rises and m_p falls; for others, it is the other way round and sometimes both develop similarly over the specimen's lifetime. This unpredictable behavior suggests that neither AM nor PM is a reliable indicator of damage. Instead, a combination of both, such as m_c, needs to be inspected to make robust predictions from VAM measurements.

Undersampled signals

We evaluate the error that is introduced by undersampling the signal with the same experimental data we

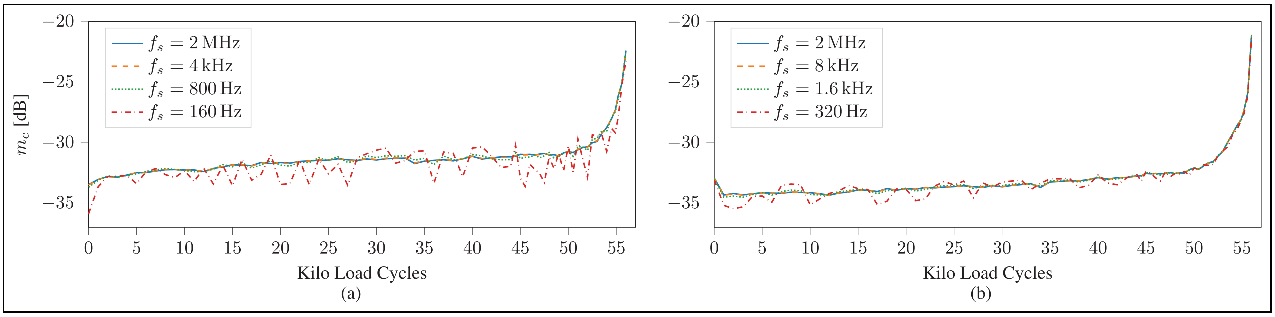

Figure 8. Comparison of damage index values computed with different rates of undersampling for a carrier frequency of (a) 193 kHz and (b) 194 kHz. The chunk size for the STFT was only $n_c = 4$.

have already used in the previous section. The original data were sampled with $f_s = 2\,\mathrm{MSample\ s^{-1}}$, and the damage indicators computed from the original data will be taken as ground truth for the evaluation.

Before undersampling, we apply a Butterworth bandpass filter to the original signal to eliminate unwanted aliasing effects. The filter's cut-off frequencies depend on the carrier frequency and are chosen at $0.9 * f_c$ and $1.1 * f_c$.

We pick every nth sample from the original time series to generate undersampled data, where n is the ratio of the original sampling frequency and the sampling frequency of the resulting time series. However, the choice of undersampling frequencies $f_s^{'}$ is limited by equation (22) and by the fact that the original sampling frequency f_s must be divisible by $f_s^{'}$. Note, however, that the latter condition is purely artificial since we are using existing, presampled data for evaluation.

Figure 8 shows the computed damage indicator m_c over the lifetime of the specimen evaluated with different sampling frequencies for two different carrier frequencies. The lowest used sampling frequencies are as low as 160 Hz. In the case of the lowest evaluated sampling frequencies, the damage indicator curve deviates considerably from the curve computed from the original data. Nevertheless, even in this case, the exponential rise of the curve toward the end of the specimen's lifetime is clearly visible.

With a sampling rate of 800 Hz in the case of Figure 8(a), or 1.6 kHz for Figure 8(b), the damage indicator computed from the undersampled data yields very similar results to the original one. Although the sampling rate is roughly 500 times lower than the Nyquist frequency, the differences between the curves are marginal compared to the curve's exponential trend over time.

In addition to the sampling frequency, the chunk size n_c is also a variable. With bigger n_c, the STFT produces less phase and amplitude measurements from the time series, but every individual phase and amplitude measurement will rely on more samples and hence be more accurate. We also compare the average deviation of the results from undersampled signals to the original signal in Table 2.

Table 2. Average deviation of m_c computed from undersampled signals and the original signal with different chunk sizes n_c in dB.

	Chunk size n_c			
f_s	4	8	16	32
		$f_c = 193$		
4 kHz	0.0296	0.0211	0.0240	0.0887
800 Hz	0.1668	0.1557	0.5572	2.4050
160 Hz	0.9286	–	–	–
		$f_c = 194$		
8 kHz	0.0107	0.0097	0.0098	0.0225
1600 Hz	0.0752	0.0732	0.1365	0.5679
320 Hz	0.3657	0.7979	–	–

The results show that the average deviations introduced by undersampling are, in many cases, less than a tenth of a dB. Even with severe undersampling, only 160 Hz, the average error is still less than a dB. Considering that during the specimen's lifetime, m_c often increases more than 10 dB, and these average errors seem tolerable.

Figure 9 shows the deviations of undersampled signals from those with full sampling rate for all available data sets with different combinations of probing frequency, sampling frequency, and chunk size. The errors depend strongly on the relation between f_c and n_c. If f_s/n_c is chosen bigger than 200, the m_c computed from the undersampled signal differs less than 1 dB from the ground truth for all data sets. A sampling rate above 800 Hz with a chunk size of four samples is sufficient to achieve this performance.

The results also show that the chunk size does play an important role. In our experiment, small chunk sizes of just four samples per chunk do not lead to significantly worse results, that is, the deviation from the original signal is not significantly higher than with larger

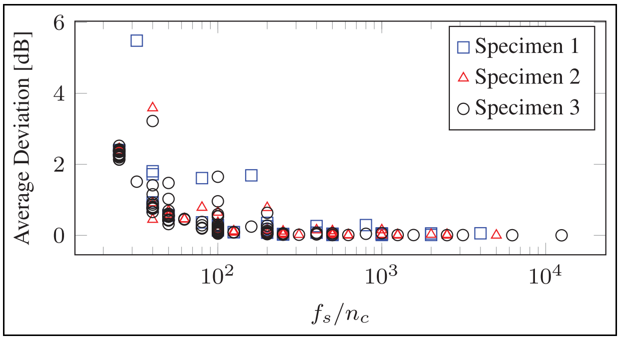

Figure 9. Overview of the deviation per data set. Every point represents a measurement for a probing frequency undersampled with a specific sampling frequency and chunk size. The plot shows the deviation of m_c from the result computed from the full 2 MHz signal. For each set of parameters, the deviations are averaged over all load cycles. The plot reveals that the errors depend strongly on the relation of f_s and n_c.

chunk sizes. On the contrary, if chunk sizes become too big, the deviations grow. The reason for this effect is that with a larger n_c, the vectors $\boldsymbol{A}$ and $\boldsymbol{P}$ have fewer samples and the modulation caused by the pumping frequency cannot be determined as accurately.

Memory and computation time

Undersampling enables a potentially huge reduction in memory demand and computational time. To quantify this improvement, we implemented the conventional MI algorithm, the HT-based determination of AM and PM, and the new STFT-based and undersampled algorithm on an ARM Cortex M4 microcontroller (STM32F446RE). Having 128 kB internal static random-access memory (SRAM) and running with 180 MHz clock frequency, this microcontroller unit (MCU) is much stronger than typical low-power MCUs characterized in Table 1. However, these resources are required for signal processing the samples for MI or HT methods.

Our test scenario evaluated a phase- and amplitude-modulated sine wave with carrier frequency 49 and 199 kHz. For the conventional MI and the HT, the 49 kHz sine was sampled with $f_s = 100$ kHz. The Goertzel algorithm was used to retrieve the amplitudes of the carrier and the lower and higher sideband. The samples were stored as 16-bit integers, as they are typically provided from an ADC. For a sufficient frequency resolution, 50 kSample were taken.

For an HT implementation, we computed the decimated discrete-time analytic signal of half sample rate as presented by Marple.[17] Using this algorithm, the calculation can work on the sampled data in place, and therefore, does not need any memory additional to the sampled voltages. It relies on FFT algorithms, where we used state-of-the-art efficient algorithms from the literature.[21] To the best of our knowledge, there is no more efficient way to calculate the HT on the microcontroller. However, the efficient FFT algorithms require a bigger datatype (4 byte per sample) and restrict the number of samples to a power of 2. Therefore, on the given MCU, no more than 16 kSample can be evaluated with HT.

Table 3. Comparison of memory requirements M, computation time T_c, and energy demand E with different carrier frequencies and recording durations T_r. Some combinations are not feasible because their requirements exceed the memory resources of the MCU.

	f_c, T_r		
	49 kHz 0.163 s	49 kHz 0.5 s	199 kHz 0.5 s
STFT + US	M<1 kB T_c≈0 ms E = 10 mJ	M<1 kB T_c≈0 ms E = 31 mJ	M<1 kB T_c≈0 ms E = 45 mJ
MI	M≈32 kB T_c≈100 ms E = 41 mJ	M≈100 kB T_c≈295 ms E = 126 mJ	M≈200 kB Infeasible –
HT	M≈65 kB T_c≈1.2 s E = 204 mJ	M≈200 kB Infeasible –	M≈400 kB Infeasible –

STFT + US: short-time fourier transform and undersampling; MI: modulation index; HT: Hilbert transform.

The same signal was also processed using STFT and undersampling. The signal was sampled with just 4 kHz and a chunk size of 8 was used for the computation. Because of the low sampling rate, each sample can be processed directly when it is taken and does not have to be stored. Therefore, this implementation only needs less than a kilobyte of memory, independent of the record duration, and no additional processing time is necessary after sampling.

We also measured the consumed energy during sampling and processing. The MCU was put into power-saving mode whenever possible. The results are shown in Table 3. The high memory demand makes HT and MI computation infeasible for higher sampling frequencies. In feasible situations, HT needs more computation time and energy compared to the MI. STFT with undersampling outperforms both. Because of its low requirements on memory and processing speed, the energy demand of the STFT-based algorithm can be further reduced by choosing more constrained low-power MCUs.

Discussion

Sources of AM and PM

Recent studies assumed that AM and PM/FM are caused by different physical effects.[1–3] Donskoy and

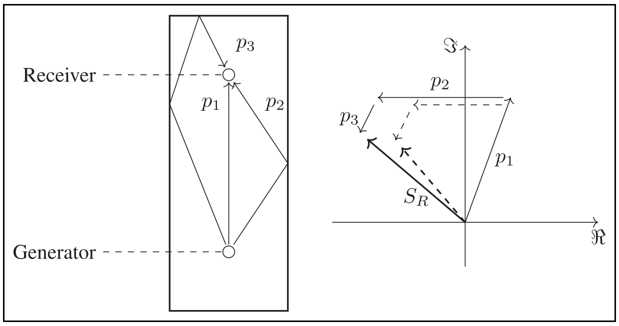

Figure 10. A simplified model of wave propagation. The dotted arrows show a situation in which the amplitude of the propagation path p_2 is decreased. This leads to a change in phase and in amplitude for the resultant wave.

Ramezani[2] suggest that FM may better indicate early fatigue damage, while Hu et al.[1] found that AM correlates better with the size of large, visible cracks. In our experiments, we cannot confirm that either AM or PM/FM is a reliable indicator on its own.

We want to point out that the modulation type measured at the receiving piezo during VAM is not necessarily the modulation type caused by a damaged area. To illustrate that, consider a simplified model of wave propagation, that assumes just one type of wave being introduced into the specimen. Although VAM is performed during a steady-state oscillation of the specimen, the steady state is produced by the superposition of many reflections of the generated signal.

Figure 10 (left) shows an example specimen with a generating ultrasonic transducer and a receiving ultrasonic transducer. Exemplarily, three paths p_1 to p_3 are sketched, over which the oscillation caused at the generator reaches the receiver. The wave has traveled a different distance on every path. Hence, the phase offset of the wave, when it reaches the receiver, may differ for every path.

If sinusoids with the same angular frequency f_c are interacting, the resultant wave S_R is again a sinusoid with the same frequency

$$S_R = \sum_k A_k e^{j(2\pi f_c t + \Phi_k)} = e^{j2\pi f_c t} \sum_k A_k e^{j\Phi_k} \tag{24}$$

The phase Φ_R and the amplitude A_R of the resultant sinusoid are given as

$$A_R = \left\| \sum_k A_k e^{j\Phi_k} \right\|, \quad \Phi_R = \arg\left(\sum_k A_k e^{j\Phi_k} \right) \tag{25}$$

This shows that the phase of the resultant wave can be modulated, even if at any individual propagating path, only AM occurs. Simultaneously, AM can occur, even though only the phase of one individual propagation path changes. Figure 10 (right) illustrates this behavior. The phase and amplitude of the wave contributed by each propagation path are plotted in the complex plane, leading to the resultant wave S_R at the receiver. We can see that if the amplitude of p_2 decreases, for example, due to crack opening, both the phase and the amplitude of the resultant wave change.

Hence, if a microcrack or fatigue damage occurs at a specific location, no matter what physical effect (crack opening, contact modulation, etc.) occurs, both amplitude and phase of the resultant wave are affected. It depends on the phase difference between the resultant wave and the wave from the individual propagation path, whether the change causes more AM, or more PM. Hence, we cannot conclude on a specific type of damage from the type of modulation alone. Recall at this point that PM and FM cannot be distinguished with any of the usually applied methods, that is, HT, IQHS, or STFT. Hence, this observation applies to studies focusing on measuring FM as well.

These observations on the simplified model also support the observations from our experiments. PM or AM alone does not yield conclusive damage indicators. It is arbitrary, how the modulation introduced in a specific propagation path affects the modulation in the resultant wave. Furthermore, for different carrier frequencies on the same specimen in the same damage state, sometimes PM dominates, and sometimes AM dominates.

General applicability of undersampling

We have demonstrated with our experiment data that in case of pumping frequencies in the range of 10 Hz, even sampling rates as low as 160 Hz can produce sufficiently accurate results. However, many studies exist that investigate NM using vastly different frequency ranges. In general, the required sampling rate depends on the pumping frequency.

For example, Ooijevaar et al.[3] use in their study a pumping frequency of 1455 Hz with a probing frequency of 50 kHz. Using the requirements on sampling frequency described in equation (13), the lowest possible sampling rate, that can still detect the modulation reliably, is roughly 12 kHz. While this sampling rate is considerably higher than the minimum sampling rates tested in this work, undersampling still enables a dramatic reduction of samples of roughly 88% compared to the Nyquist frequency of $2f_c = 100\,\text{kHz}$, that would have to be applied without undersampling.

Conclusion

This work's objective was to explore approaches to make the VAM method sufficiently efficient for use in low-power WSNs. We proposed to use the STFT instead of the HT or the IQHS algorithm to split AM and PM in a signal. The advantages of this approach in computational complexity and memory demands have been discussed extensively.

A new damage indicator based on the AM and PM measurements has been introduced and compared with the conventionally used MI. Experiments on real specimen show that this new indicator performs comparably to the MI. Furthermore, our experiments suggest that PM and AM alone cannot reliably assess damage, but considering both at the same time yields significantly improved results. A simplified model has been proposed to explain the possible source of AM and PM and why they are not reliable individually.

We discussed how undersampling could be used to drastically reduce the required amount of memory and computational resources for computing the damage indicator. Using undersampling on existing experiment data has achieved virtually identical results with only 0.2% of the conventionally required samples. The presented algorithm was implemented on a microcontroller and required only a quarter of the energy compared to the conventional MI in a test scenario, while consuming only 1% of the memory. The proposed methods allow using VAM on low-cost, low-power hardware for low pumping frequencies by drastically reducing the required memory and consequentially also the required computational power of the sensor nodes.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the agency for science, research, and gender equality of the city of Hamburg (BWFG—Behörde für Wissenschaft, Forschung und Gleichstellung).

ORCID iDs

Peter Oppermann https://orcid.org/0000-0001-6936-2307
Lennart Dorendorf https://orcid.org/0000-0003-0830-7982

References

1. Hu H, Staszewski W, Hu N, et al. Crack detection using nonlinear acoustics and piezoceramic transducers—instantaneous amplitude and frequency analysis. *Smart Mater Struct* 2010; 19(6): 065017
2. Donskoy D and Ramezani M. Separation of amplitude and frequency modulations in vibro-acoustic modulation nondestructive testing method. *Proc Mtgs Acoust* 2018; 34(1): 045002
3. Ooijevaar T, Rogge MD, Loendersloot R, et al. Vibro-acoustic modulation–based damage identification in a composite skin–stiffener structure. *Struct Health Monit* 2016; 15(4): 458–472.
4. Aymerich F and Staszewski W. Experimental study of impact-damage detection in composite laminates using a cross-modulation vibro-acoustic technique. *Struct Health Monit* 2010; 9(6): 541–553.
5. Gaglione A, Rodenas-Herraiz D, Jia Y, et al. Energy neutral operation of vibration energy-harvesting sensor networks for bridge applications. In: *Proceedings of the 2018 international conference on embedded wireless systems and networks*, Madrid, 14-16 February 2018.
6. Alavi AH, Hasni H, Jiao P, et al. Self-charging and self-monitoring smart civil infrastructure systems: current practice and future trends. In: *Proceedings of the sensors and smart structures technologies for civil, mechanical, and aerospace systems*, volume 10970. (Int Society for Optics and Photonics) Denver CO, 27 March 2019.
7. Renner C, Unterschütz S, Turau V, et al. Perpetual data collection with energy-harvesting sensor networks. *ACM Trans Sens Netw (TOSN)* 2014; 11(1): 1–45.
8. Elvin NG, Lajnef N and Elvin AA. Feasibility of structural monitoring with vibration powered sensors. *Smart Mater Struct* 2006; 15(4): 977
9. Sazonov E, Li H, Curry D, et al. Self-powered sensors for monitoring of highway bridges. *IEEE Sens J* 2009; 9(11): 1422–1429.
10. Liu P, Lim HJ, Yang S, et al. Development of a "stick-and-detect" wireless sensor node for fatigue crack detection. *Struct Health Monit* 2017; 16(2): 153–163.
11. Yang S, Jung J, Liu P, et al. Ultrasonic wireless sensor development for online fatigue crack detection and failure warning. *Struct Eng Mech* 2019; 69(4): 407–416.
12. Donskoy DM. Nonlinear acoustic methods. In: Boller C, Chang F-K and Fujino Y (eds) *Encyclopedia of structural health monitoring*. Hoboken, NJ: Wiley Online Library, 2009.
13. Lim HJ and Sohn H. Online fatigue crack prognosis using nonlinear ultrasonic modulation. *Struct Health Monit* 2019; 18(5–6): 1889–1902.
14. Klepka A, Pieczonka L, Staszewski WJ, et al. Impact damage detection in laminated composites by non-linear vibro-acoustic wave modulations. *Compo Part B Eng* 2014; 65: 99–108.
15. Lee SE, Jin S and Hong JW. Periodic nonlinear waves resulting from the contact interaction of a crack. *J Appl Phys* 2017; 122(12): 124303
16. Roder H. Amplitude, phase, and frequency modulation. *Proc Inst Radio Eng* 1931; 19(12): 2145–2176.
17. Marple L. Computing the discrete-time "analytic" signal via FFT. *IEEE Trans Sig Proc* 1999; 47(9): 2600–2603
18. Goertzel G. An algorithm for the evaluation of finite trigonometric series. *Am Math Month* 1958; 65(1): 34–35.

19. León FP and Jäkel H. *Signale und systeme*. Berlin: De Gruyter, 2015.
20. Donskoy D, Golchinfar B, Ramezani M, et al. Vibro-acoustic amplitude and frequency modulations during fatigue damage evolution. *AIP Conf Proc* 2019; 2102: 5099754.
21. Press W, Teukolsky S, Vetterling W, et al. *Numerical recipes in C. New York: Cambridge Univ Press*, 2002.