MEASUREMENT SCIENCE REVIEW, **18**, (2018), No. 4, 130-137

sciendo

ISSN 1335-8871

MEASUREMENT SCIENCE REVIEW

Journal homepage: http://www.degruyter.com/view/j/msr

Application of Conditional Averaging to Time Delay Estimation of Random Signals

Robert Hanus, Adam Kowalczyk, Anna Szlachta, Rafał Chorzępa

Rzeszów University of Technology, Faculty of Electrical and Computer Engineering, 12 Powstańców Warszawy Ave., 35-959 Rzeszów, Poland, rohan@prz.edu.pl

The article presents the possibilities of using the function of conditional average value of a delayed signal (CAV) and the function of conditional average value of a delayed signal absolute value (CAAV) to determine the time delay estimation (TDE) of random signals. For discrete CAV and CAAV estimators, the standard uncertainties of the estimation of function values at extreme points and the standard uncertainties of the TDE were given and compared with the corresponding uncertainties for the direct discrete cross-correlation function (CCF) estimator. It was found that the standard uncertainty of TDE for CAV is lower than for CCF independent of signal-to-noise ratio (SNR) for parameter values of $\alpha \geq 2$ and $M/N \geq 0.25$ (where: α - relative threshold value, M/N – quotient of number of averaging and number of samples). The standard uncertainty of TDE for CAAV will be lower than for CCF for SNR values greater than 0.35 (for $N/M = 1$).

Keywords: Time delay estimation, random signals, conditional averaging, cross-correlation.

1. INTRODUCTION

Time delay estimation (TDE) is a common problem in signal processing. This problem is important, inter alia, in radar technology, radioastronomy, medicine, localization of disturbance pathways, and non-contact measurements of two-phase flows. For the determination of time delay of stochastic signals obtained from two or more sensors, statistical methods are commonly used. This issue is widely presented in the literature, among others [1]-[10], where a number of methods based on analysis of time and frequency signals are described. The range of applications of particular methods depends on the characteristics of the measurement signals (stationarity, probability distribution, correlation), signal-to-noise ratio (SNR), and analysis parameters. Among the classical methods used for Gaussian stationary signals, the most common are the cross-correlation function (CCF) in the time domain and phase of cross-spectral density in the frequency domain [1], [8], [10]-[15]. For the above types of signals under certain conditions, differential and combined methods [4]-[6], the Hilbert Transform-based correlation method [16]-[17], and the relatively little-known methods based on conditional averaging [18]-[20] can be used.

This paper presents the applicability of the conditional average value of the delayed signal (CAV) and the conditional average value of the delayed signal absolute value (CAAV) to the TDE in relation to the CCF. For the selected signal models, the range is determined for the SNR values for which the methods under consideration have lower standard uncertainties than the CCF method with the specified analysis parameters.

The paper is organized as follows. Section 2 gives the mathematical models of measurement random signals. The basics of the TDE using the classical cross-correlation method are reviewed in section 3. Section 4 presents the application of conditional averaging methods CAV and CAAV to TDE and the analysis of measurement uncertainties for these methods in comparison with the CCF. The last section contains the summary of the presented study and final conclusions. This article is an extended and improved version of a previous conference publication [20].

2. MODELS OF SIGNALS

Measurement of the time delay of random signals during their propagation in physical systems involves active and passive experiments. In the active experiment, external random extortion is used. In the passive experiment, natural conditions in the object of measurement are used to generate measurement signals. The full measurement model should take into account inaccuracies in terms of input and output signals, resulting from difficult experimental conditions, and the impact of disturbance. Simplification of the random signal time delay measurement model may lead to a reduction in accuracy of the TDE.

DOI: 10.1515/msr-2018-0019

In time delay estimation, the models of stochastic signals $x(t)$ and $z(t)$ obtained under stationary conditions with two sensors in a passive experiment can be presented as [1]:

$$x(t) = s(t) + n_1(t) \tag{1}$$

$$z(t) = c\left[s(t-\tau_0)\right] + n_2(t) = y(t) + n_2(t) \tag{2}$$

where: $s(t)$ and $z(t)$ – observed input and output stationary random signals; c - constant factor (usually c = 1); τ_0 – transmission time delay; $n_1(t)$, $n_2(t)$ – stationary signals, uncorrelated with the signal $s(t)$ and with each other, which may in general contain constant, periodic and random components. This paper assumes that the signals $n_1(t)$ and $n_2(t)$ are stationary Gaussian white noises N(0,σ_{n1}), N(0,σ_{n2}), uncorrelated with the signal s(t) and mutually with each other.

With the above assumptions for signal models (1) and (2), the following relationships are true:

$$\sigma_x^2 = R_{xx}(0) = \sigma_s^2 + \sigma_{n1}^2 \tag{3}$$

$$\sigma_z^2 = R_{zz}(0) = c^2\sigma_s^2 + \sigma_{n2}^2 \tag{4}$$

$$R_{sz}(\tau) = cR_{ss}(\tau - \tau_0) + R_{n1n2}(\tau - \tau_0) \tag{5}$$

Where: σ_s, σ_x, σ_z, σ_{n1}, and σ_{n2} - standard deviations of signals $s(t)$, $x(t)$, $z(t)$, $n_1(t)$, and $n_2(t)$, respectively; $R_{xx}(\tau)$, $R_{zz}(\tau)$, and $R_{ss}(\tau)$ - autocorrelation functions of signals $x(t)$, $z(t)$, and $s(t)$, respectively; τ - time delay; $R_{sz}(\tau)$ - CCF of signals $s(t)$ and $z(t)$; $R_{n1n2}(\tau)$ - CCF of signals $n_1(t)$ and $n_2(t)$.

In the active experiment for determining the transmission time delay, the conditions of action of the test input signal are specified and controlled (no disturbances). Conversely, output noise can lead to the action of only a random component $n(t)$ with the distribution N(0,σ_n). With these assumptions, the models of stochastic signals received from the sensors are represented by the following formulae:

$$x(t) = s(t) \tag{6}$$

$$z(t) = c \cdot x(t - \tau_0) + n(t) = y(t) + n(t) \tag{7}$$

which constitute a simplification of models (1) and (2) with the assumptions $n_1(t) = 0$ and $n_2(t) = n(t)$.

Since the signals $x(t)$ and $n(t)$ are not correlated, for the delayed signal (7) we have the following equations:

$$\sigma_z = \sqrt{(c\sigma_x)^2 + \sigma_n^2} \tag{8}$$

$$R_{xz}(\tau) = cR_{ss}(\tau - \tau_0) \tag{9}$$

where $R_{xz}(\tau)$ - the cross-correlation function of signals $x(t)$ and $z(t)$.

This paper assumes that the signal $s(t)$ is low-pass white noise with bandwidth B. The autocorrelation function of this signal in equations (5) and (9) can be represented by:

$$R_{ss}(\tau) = KB\left(\frac{\sin 2\pi B\tau}{2\pi B\tau}\right) \tag{10}$$

or another modified exponential model [21].

Experimental research usually uses normal or quasi-normal probability distributions of the processed signals.

3. Correlation principle of TDE

Most of the descriptions in the literature on the cross-correlation method of determining the transmission time delay use dependencies (1) and (2) or (6) and (7). The CCF achieves the maximum value for $\tau = \tau_0$, so the transmission time delay can be defined as the main argument of this function (Fig.1.):

$$\tau_0 = \arg\{\max R_{xz}(\tau)\} = \arg\{R_{xz}(\tau_0)\} \tag{11}$$

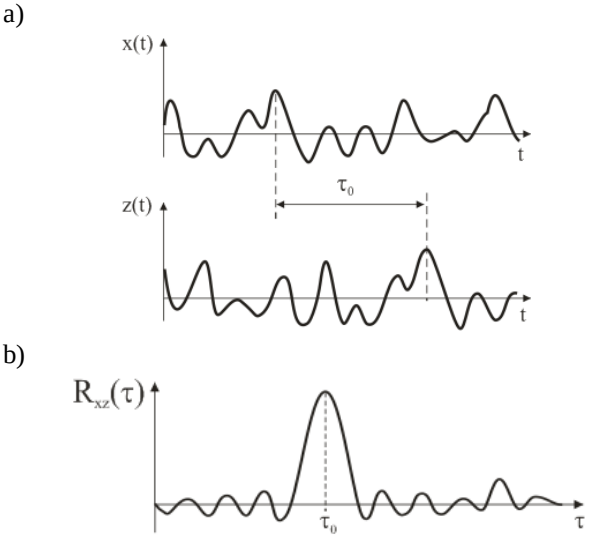


Fig.1. The concept of determining the transmission time delay τ_0 from the cross-correlation function: a) the waveforms $x(t)$ and $z(t)$; b) the CCF $R_{xz}(\tau)$.

The normalized CCF (9) is equal to:

$$\rho_{xz}(\tau) = \frac{R_{xz}(\tau)}{\sqrt{R_{xx}(0)R_{zz}(0)}} = \frac{cR_{ss}(\tau - \tau_0)}{\sigma_x\sigma_z} \tag{12}$$

and after substitution $\tau = \tau_0$:

$$\rho_{xz}(\tau_0) = \frac{cR_{ss}(0)}{\sigma_x\sigma_z} = \frac{c\sigma_s^2}{\sigma_x\sigma_z} \tag{13}$$

Assuming the same disruption at the input and output ($\sigma_{n1} = \sigma_{n2} = \sigma_n$) for signal models (1) and (2), after substituting (3) and (4) to the equation (13), we obtain [22]:

$$\rho_{xz}(\tau_0)=\left[1+\frac{1}{SNR}\left(1+\frac{1}{c^2}\right)+\frac{1}{c^2SNR^2}\right]^{-1/2} \quad (14)$$

where

$$SNR=\left(\frac{\sigma_s}{\sigma_n}\right)^2 \quad (15)$$

Similarly, for signal models (6) and (7) we have:

$$\rho_{xz}(\tau_0)=\left[1+\frac{1}{c^2SNR}\right]^{-1/2} \quad (16)$$

In the absence of disruption in both cases $\rho_{xz}(\tau_0) = 1$.

The plots of the relations (15) and (16), referred to hereafter as model I and model II, respectively, are shown in Fig.2. Particularly equations (15) and (16) and their graphs can be very useful in practice. Based on them, for obtained experimental values of normalized CCF $\rho_{xz}(\tau)$ and assumed signal models, the SNR value can be estimated at an unknown value of $n(t)$.

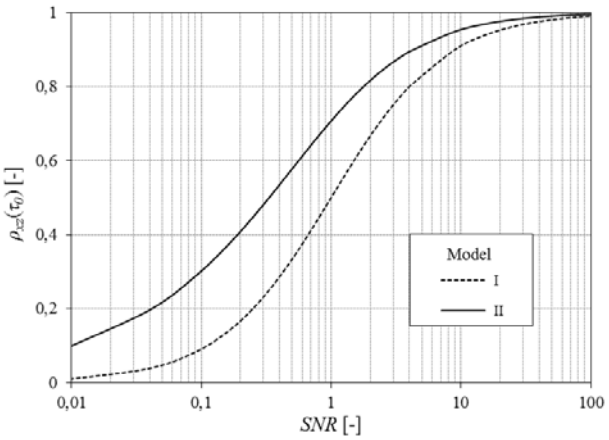


Fig.2. Plot of the relation (14) (Model I) and (16) (Model II) for $c = 1$.

For further analysis, the equation (16) will be used as the most commonly used signal model in practice. If the waveforms $x(t)$ and $z(t)$ are of length T_{total}, the variance of the CCF estimator for $\tau = \tau_0$ is given by [1]:

$$\sigma^2\left[\hat{R}_{xz}(\tau_0)\right]\approx\frac{1}{2BT_{total}}\left[R_{xx}(0)R_{zz}(0)+R_{xz}^2(\tau_0)\right] \quad (17)$$

Dependency (17) is appropriate for large values of T_{total} ($T_{total}\geq 10|\tau|$ and $BT_{total}\geq 5$).

For digital analysis methods, we can take $2BT_{total} = N_{total}$ [15] where $N_{total}=\text{int}(T_{total}/\Delta t)$, int - integer, and Δt is a properly chosen sampling period. Transforming (17) we obtain:

$$\sigma^2\left[\hat{R}_{xz}(\tau_0)\right]\approx\frac{\sigma_x^2\sigma_z^2}{N_{total}}\left[1+\rho_{xz}^{-2}(\tau_0)\right] \quad (18)$$

When estimating CCF using pairs of uncorrelated samples divided into N-cycles, on the basis of (18) taking into account (16) we obtain the relative standard uncertainty of the function value:

$$u_{rel}\left[\hat{R}_{xz}(\tau_0)\right]=\frac{\sigma\left[\hat{R}_{xz}(\tau_0)\right]}{R_{xz}(\tau_0)}=\frac{1}{\sqrt{N}}\sqrt{2+\frac{1}{c^2SNR}} \quad (19)$$

A discrete CCF estimator can be expressed as:

$$\hat{R}_{xz}(l)=\frac{1}{N-l}\sum_{n=1}^{N-l}x(n)z(n+l) \quad (20)$$

where: $l = \tau/\Delta t$, $n = t/\Delta t$.

For a large number of samples, N is used in the denominator of dependency (20) instead of $N - l$.

The standard uncertainty of transmission time delay τ_0, determined from CCF, is given by [20]:

$$u[\tau_0]_{CCF}=\frac{1}{\pi B}\sqrt[4]{\frac{3}{4}}\left\{\left(u_{rel}\left[\hat{R}_{xz}(\tau_0)\right]\right)^{1/2}\right\}^{1/2} \quad (21)$$

For signal models (6) and (7), using the equation (19) in (21) we obtain:

$$u[\tau_0]_{CCF}=\frac{1}{\pi B}\sqrt[4]{\frac{3}{4}}\left\{\frac{1}{\sqrt{N}}\left[2+\frac{1}{c^2SNR}\right]^{1/2}\right\}^{1/2} \quad (22)$$

4. Application of conditional averaging to TDE of the random signals

The transmission time delay can be determined with the use of conditional expected values: delayed signal $z(t)$ and delayed signal absolute value $|z(t)|$. Sections 4.1 and 4.2 illustrate the concept of particular methods and their standard uncertainties with respect to the cross-correlation method for signal models (6) and (7).

4.1. Conditional average value of the delayed signal

The conditional expected value of the delayed signal $z(t)$ (7) for the condition $x(t) = x_p$ is defined as [20]:

$$\begin{aligned}A_z(\tau)=A_z\big|_{x(t)=x_p}(\tau)=E\left\{z(t+\tau)\big|_{x(t)=x_p}\right\}=\\=\int_{-\infty}^{\infty}z(t+\tau)\,p\left(z(t+\tau)\big|_{x(t)=x_p}\right)dz\end{aligned} \quad (23)$$

where $p\left(z(t+\tau)\big|_{x(t)=x_p}\right)$ represents the conditional probability density of the signal $z(t)$ for the condition $x = x_p$; $E(\cdot)$ is the expected value; x_p is the set threshold value.

In the case of independence of signal $x(t)$ and disturbance $n(t)$, the expected value of the delayed signal $z(t)$ is also independent of $n(t)$:

$$\begin{aligned} &E\left[z(t+\tau)\big|_{x(t)=x_p} \right] = E\left[z(\tau)\big|_{x_p} \right] = \\ &= E\left[y(\tau)\big|_{x_p} \right] + E\left[n(\tau)\big|_{x_p} \right] = E\left[y(\tau)\big|_{x_p} \right] \end{aligned} \tag{24}$$

If the signals $x(t)$ and $y(t)$ are normal together, with average values of zero, then the conditional probability density of the signal $y(t + \tau)$ at the condition $x(t) = x_p$, is expressed as follows [19]:

$$\begin{aligned} &p\left[y(\tau)\big|_{x_p} \right] = \frac{1}{\sigma_y \sqrt{2\pi}\sqrt{1-\rho_{xy}^2(\tau)}} \cdot \\ &\cdot \exp\left\{ -\frac{2}{2\left[1-\rho_{xy}^2(\tau)\right]} \left[\frac{\rho_{xy}(\tau)}{\sigma_x} x_p - \frac{y(\tau)}{\sigma_y} \right]^2 \right\} \end{aligned} \tag{25}$$

with the conditional expected value:

$$E\left[y(t+\tau)\big|_{x(t)=x_p} \right] = E\left[y(\tau)\big|_{x_p} \right] = \frac{\sigma_y}{\sigma_x} \rho_{xy}(\tau) \cdot x_p \tag{26}$$

and conditional variance:

$$\sigma^2\left[y(t+\tau)\big|_{x(t)=x_p} \right] = \sigma_y^2\left[1-\rho_{xy}^2(\tau)\right] \tag{27}$$

The relationship between $A_z(\tau)$ and the normalized CCF $\rho_{xz}(\tau)$ is given by the formula:

$$A_z(\tau) = \frac{\sigma_z}{\sigma_x} \rho_{xz}(\tau) \cdot x_p \tag{28}$$

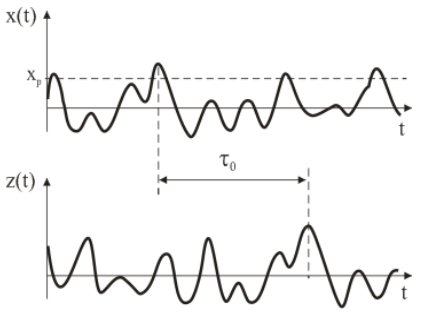
A good estimate of the conditional expected value is the arithmetical conditional average value of the signal. Its designation in practice consists in detecting mutually uncorrelated moments of transition of the signal $x(t)$ through the level x_p, run at these moments registering fragments of the delayed signal $z(t)$ and averaging those fragments in the set:

$$\hat{A}_z(l) = \overline{z(l)}\Big|_{x(n)=x_p} = \frac{1}{M} \sum_{n=1}^{M} z(n+l)\Big|_{x(n)=x_p} \tag{29}$$

where M is the number of averaged uncorrelated segments of the delayed signal $z(t)$.

The position of the main maximum of CAV function defines the transmission time delay as in the case of CCF (Fig.3.).

a)



b)

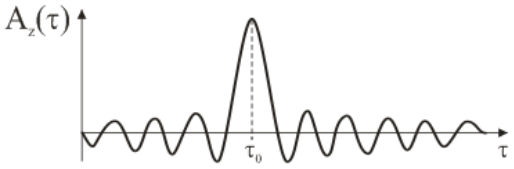


Fig.3. The concept of TDE using CAV function: a) the signals $x(t)$ and $z(t)$, b) the CAV function $A_z(\tau)$.

The CAV variance for M averaging is determined by the formula:

$$\sigma^2\left[\hat{A}_z(\tau)\right] = \frac{\sigma_z^2\left[1-\rho_{xz}^2(\tau)\right]}{M} \tag{30}$$

4.1.1. Comparison of standard uncertainties of estimation of CAV and CCF function values in the neighborhood of extreme points

The relative standard uncertainty of the CAV function estimation for $\tau = \tau_0$ taking into account (28), (30), (16), and (8) can be presented by [20]:

$$u_{rel}\left[\hat{A}_z(\tau_0)\right] = \frac{\sigma\left[\hat{A}_z(\tau_0)\right]}{A_z(\tau_0)} = \frac{1}{\sqrt{M}} \frac{\sigma_n}{c\, x_p} = \frac{1}{c\alpha\sqrt{M \cdot SNR}} \tag{31}$$

where $\alpha = (x_p/\sigma_x)$ - relative threshold value.

After dividing equation (31) by (19), the quotient of relative uncertainties of CAV and CCF is obtained as:

$$\frac{u_{rel}\left[\hat{A}_z(\tau_0)\right]}{u_{rel}\left[\hat{R}_{xz}(\tau_0)\right]} = \frac{1}{\alpha}\left[\frac{M}{N}\left(2c^2 SNR + 1\right) \right]^{-1/2} \tag{32}$$

Since CAV and CCF are determined on the basis of uncorrelated signal samples, the M/N quotient may be 1 or less than unity. The plot of the relation $u_{rel}[\hat{A}_z(\tau_0)] / u_{rel}[\hat{R}_{xz}(\tau_0)] = f(SNR)$ for $M/N = 1$, $c = 1$ and several selected values of α is shown in Fig.4. In this case, the relative standard uncertainty $u_{rel}[\hat{A}_z(\tau_0)]$ will be less than $u_{rel}[\hat{R}_{xz}(\tau_0)]$ for the value $\alpha \geq 1$.

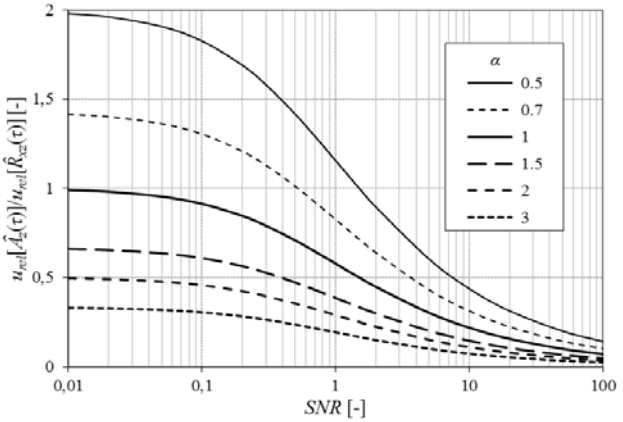


Fig.4. Plot of the relation $u_{rel}\left[\hat{A}_z(\tau_0)\right]/u_{rel}\left[\hat{R}_{xz}(\tau_0)\right] = f(SNR)$ for $M/N = 1$, $c = 1$ and selected values of α.

In the work [19] it has been shown that the optimum value of α is about 2. Fig.5. shows a plot of the dependency $u_{rel}\left[\hat{A}_z(\tau_0)\right]/u_{rel}\left[\hat{R}_{xz}(\tau_0)\right] = f(SNR)$ for $\alpha = 2$, $c = 1$ and several selected M/N quotient values.

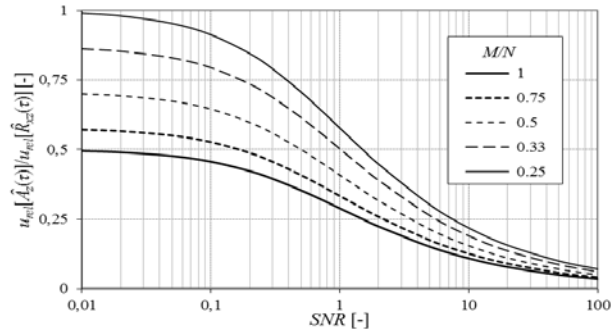


Fig.5. Plot of the relation $u_{rel}\left[\hat{A}_z(\tau_0)\right]/u_{rel}\left[\hat{R}_{xz}(\tau_0)\right] = f(SNR)$ for $\alpha = 2$, $c = 1$ and selected values of M/N ratio.

4.1.2. Comparison of the standard uncertainties of TDE for CAV and CCF

For the CAV function, the standard uncertainty of the TDE can be represented by the equation [20]:

$$u[\tau_0]_{CAV} = \frac{1}{\pi B}\sqrt[4]{\frac{3}{4}}\sqrt{u_{rel}\left[\hat{A}_z(\tau_0)\right]} \tag{33}$$

After taking into account (30) we get:

$$u[\tau_0]_{CAV} = \frac{1}{\pi B}\sqrt[4]{\frac{3}{4}}\frac{1}{\sqrt{c\alpha}}(M \cdot SNR)^{-1/4} \tag{34}$$

By comparing (34) and (22) we finally obtain:

$$\frac{u[\tau_0]_{CAV}}{u[\tau_0]_{CCF}} = \frac{1}{\sqrt{\alpha}}\left[\frac{M}{N}\left(2c^2 SNR + 1\right)\right]^{-1/4} \tag{35}$$

The plot of the relation $u[\tau_0]_{CAV}/u[\tau_0]_{CCF} = f(SNR)$ for $M/N = 1$, $c = 1$ and a few selected values of α is shown in Fig.6.

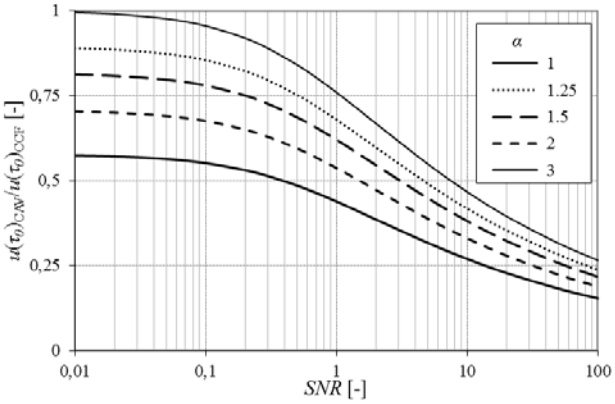


Fig.6. Plot of the relation $u[\tau_0]_{CAV}/u[\tau_0]_{CCF} = f(SNR)$ for $M/N = 1$, $c = 1$ and a few selected values of α.

Fig.7. shows the plot of the relation (35) for $\alpha = 2$, $c = 1$ and a few selected M/N quotient values.

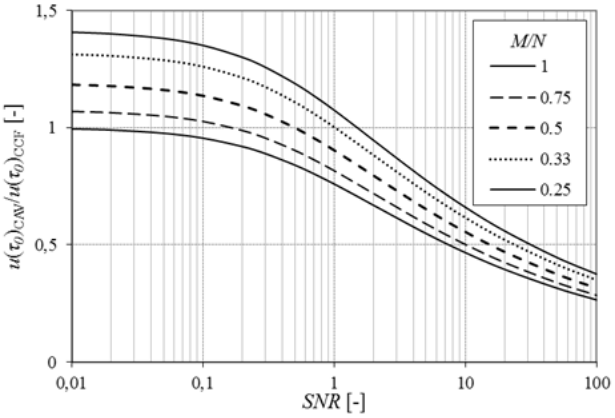


Fig.7. Plot of the relation $u[\tau_0]_{CAV}/u[\tau_0]_{CCF} = f(SNR)$ for $\alpha = 2$, $c = 1$ and selected values of M/N.

Based on Fig.7., it can be stated that the standard uncertainty of TDE using CAV is lower than for CCF independent of SNR values for $\alpha \geq 2$ and $M/N \geq 0.25$.

4.2. Conditional average value of the absolute value of delayed signal

The conditional expected value of the delayed signal absolute value $|z(t)|$ for the condition $x(t) = 0$ is defined by the equation [21]:

$$\begin{aligned} A_{|z|}(\tau) = A_{|z|}\big|_{x(t)=0}(\tau) &= E\left\{\left|z(t+\tau)\right|_{x(t)=0}\right\} = \\ &= \int_0^\infty |z(t+\tau)|\, p\left(\left|z(t+\tau)\right|_{x(t)=0}\right) dz \end{aligned} \tag{36}$$

where $p\left(\left|z(t+\tau)\right\|_{x(t)=0}\right)$ represents the conditional probability density of the $z(t)$ signal absolute value for the condition $x(t) = 0$.

For a normal probability density function $p\left(\left|z(t+\tau)\right\|_{x(t)=0}\right)$ on the basis of (36), the CAAV $A_{|z|}(\tau)$ is obtained in the form [23]:

$$A_{|z|}(\tau) = \sigma_z \sqrt{\frac{2}{\pi}\left[1-\rho_{xz}^2(\tau)\right]} \tag{37}$$

Transmission time delay is determined by the position of the main minimum of the CAAV function:

$$\tau_0 = \arg\{\min A_{|z|}(\tau)\} = \arg\{A_{|z|}(\tau_0)\} \tag{38}$$

After normalization, the equation (37) takes the form:

$$a_{|z|}(\tau) = \frac{A_{|z|}(\tau)}{A_{|z|}(\tau)_{\max}} = \sqrt{1-\rho_{xz}^2(\tau)} \tag{39}$$

Fig.8. shows a comparison of the normalized CAAV $a_{|z|}(\tau)$ and CCF $\rho_{xz}(\tau)$ functions for SNR = ∞ ($n(t) = 0$), and SNR = 5.

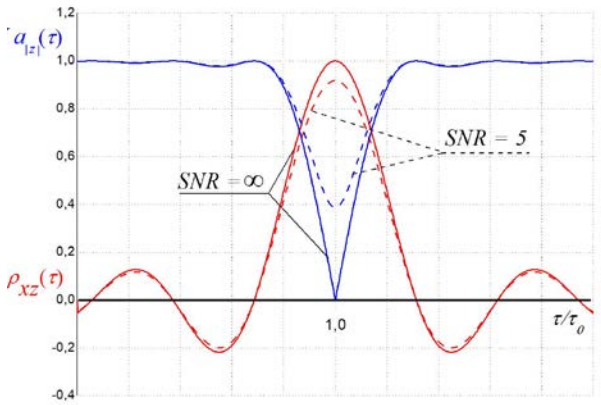


Fig.8. Examples of normalized CAAV $a_{|z|}(\tau)$ and CCF $\rho_{xz}(\tau)$ functions.

The function $a_{|z|}(\tau)$ is characterized by a sharper extremum than the function $\rho_{xz}(\tau)$. A decrease of the SNR value results in a decrease of the main CCF maximum by formula (16) and, respectively, growth of the main CAAV minimum according to:

$$a_{|z|}(\tau_0) = \sqrt{1-\left(1+\frac{1}{c^2 SNR}\right)^{-1}} \tag{40}$$

A good estimate of the expected conditional value (36) is the arithmetical conditional average value of the delayed signal absolute value. Determining the CAAV estimator involves detecting mutually uncorrelated transition times of the original signal $x(t)$ across zero, triggering at these moments registration of fragments of the delayed signal absolute value $|z(t)|$ and averaging those fragments in the set according to the formula:

$$\hat{A}_{|z|}(l) = \overline{\left|z(l)\right|}\Big|_{x(n)=0} = \frac{1}{M}\sum_{n=1}^{M}\left|z(n+l)\right|\Big|_{x(n)=0} \tag{41}$$

The variance of the CAAV estimator for M averages is determined by the equation [18]:

$$\sigma^2\left[\hat{A}_{|z|}(\tau_0)\right] \approx \frac{1}{M}\sigma_z^2\left(1-\frac{2}{\pi}\right)\left[1-\rho_{xz}^2(\tau_0)\right] \tag{42}$$

4.2.1. Comparison of the standard uncertainties of estimation of CAAV and CCF values in the neighborhood of extreme points

Based on equation (42) after taking into account (16), the relative standard uncertainty of the CAAV value estimation is obtained as:

$$u_{rel}\left[\hat{A}_{|z|}(\tau_0)\right] = \frac{\sigma\left[\hat{A}_{|z|}(\tau_0)\right]}{A_{|z|}(\tau)_{\max}} \approx \left[\frac{1}{M}\left(\frac{\pi}{2}-1\right)\frac{1}{1+c^2 SNR}\right]^{1/2} \tag{43}$$

By comparing (43) and (19) we find [23]:

$$\frac{u_{rel}\left[\hat{A}_{|z|}(\tau_0)\right]}{u_{rel}\left[\hat{R}_{xz}(\tau_0)\right]} \approx \left[\frac{N}{M}\frac{\left(\frac{\pi}{2}-1\right)\left(\frac{1}{1+c^2 SNR}\right)}{\left(2+\frac{1}{c^2 SNR}\right)}\right]^{1/2} \tag{44}$$

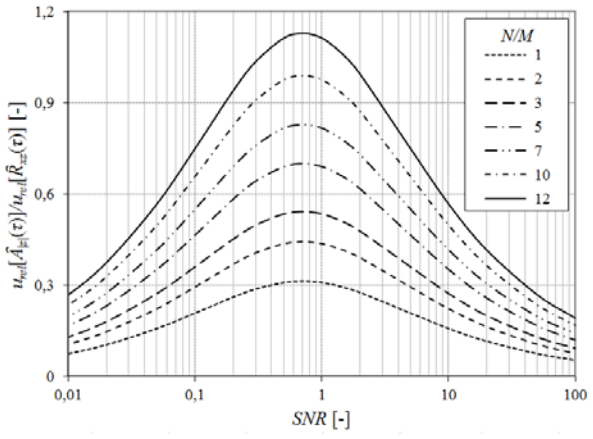


Fig.9. Plots of the relation $u_{rel}\left[\hat{A}_{|z|}(\tau_0)\right] / u_{rel}\left[\hat{R}_{xz}(\tau_0)\right] = f(SNR)$ for a few values of N/M ratio [23].

The plot of the relation (44) for $c = 1$ and several N/M quotient values is shown in Fig.9. As can be seen in the SNR range under consideration, the relative standard uncertainty of CAAV is always less than the corresponding

CCF uncertainty for *N/M* values less than or equal to 10. In practice, the value of the *N/M* quotient depends on the interval of the correlation of the measurement signals, which determines the choice of uncorrelated samples.

4.2.2. Comparison of the standard uncertainties of TDE for CAAV and CCF

The standard uncertainty of the τ_0 transmission time delay estimation from the CAAV function is given by [22]:

$$u[\hat{\tau}_0]_{CAAV} = \frac{1-\rho_{xz}^2(\tau_0)}{\rho_{xz}(\tau_0)}\left[\frac{1}{M}\left(\frac{\pi}{2}-1\right)\right]^{1/2} \tag{45}$$

After taking into account equations (45), (16), and (19), we obtain the relation $u[\hat{\tau}_0]_{CAAV} / u[\hat{\tau}_0]_{CCF} = f(SNR)$:

$$\frac{u[\hat{\tau}_0]_{CAAV}}{u[\hat{\tau}_0]_{CCF}} = \left\{\frac{N\left(\frac{\pi}{2}-1\right)}{Mc^2 SNR(2c^2 SNR+1)}\right\}^{1/2} \tag{46}$$

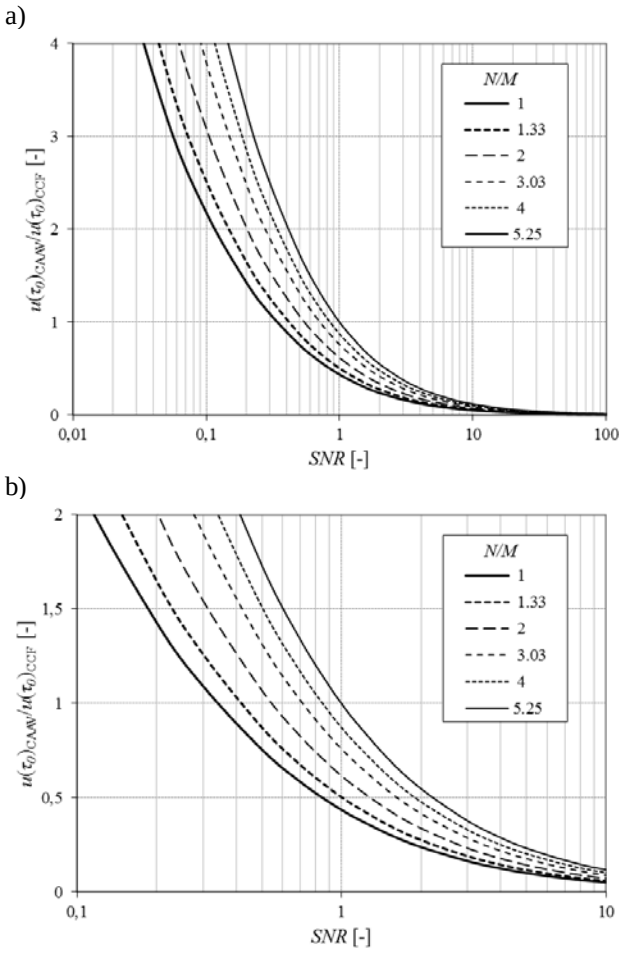


Fig.10. Plots of the relation $u[\hat{\tau}_0]_{CAAV} / u[\hat{\tau}_0]_{CCF} = f(SNR)$ for $c = 1$ and a few *N/M* quotient values; a) SNR range from 0.01 to 100, b) SNR range from 0.1 to 10.

The plot of the relation (46) for $c = 1$ and several *N/M* quotient values is shown in Fig.10. Fig.10.b) shows an enlarged area for which the quotient $u[\hat{\tau}_0]_{CAAV} / u[\hat{\tau}_0]_{CCF}$ assumes values close to 1. Based on the graphs shown in Fig.10., it can be concluded that the standard uncertainty of TDE using CAAV will be less than for CCF for SNR values greater than 0.35 (for $N/M = 1$). An increase in the *N/M* quotient results in an increase in the SNR limit value, for which the standard uncertainty of the CAAV time delay estimation is less than for CCF. For SNR = 1 equalization of the values $u[\hat{\tau}_0]_{CAAV}$ and $u[\hat{\tau}_0]_{CCF}$ occurs for $N/M = 5.25$.

5. SUMMARY

The article presents the application of two methods of conditional signal averaging to determine the transmission time delay of stochastic signals. These methods are based on analysis of the CAV and CAAV functions. In the first part of the paper we give the models of measurement signals used in the TDE of random signals and, for comparative purposes, the principle and main metrological characteristics of the cross-correlation method.

The main part of the paper presents the concept of application of the conditional averaging to TDE. For discrete CAV and CAAV estimators, the standard uncertainties of the estimation of function values in the neighborhood of the extremes and the standard uncertainties of the TDE were analyzed and compared with the corresponding uncertainties for the direct discrete CCF estimator. For the most commonly used TDE models of measurement signals, the range of SNR values is defined, for which the methods under consideration have lower standard uncertainties of estimation at specified analysis parameters. It was found that the standard uncertainty of TDE using CAV is lower than for CCF independent of SNR values for parameter values of $\alpha \geq 2$ and $M/N \geq 0.25$. The standard uncertainty of TDE applying CAAV will be lower than for CCF for SNR values greater than 0.35 ($N/M = 1$). An increase in *N/M* results in an increase in the *SNR*, for which the standard uncertainty of the CAAV time delay estimation is lower than for CCF.

This paper presents the basics and the results of theoretical analysis of CAV and CAAV methods. The authors carried out a simulation and experimental study of the properties of conditional averaging methods in TDE of stochastic signals [23]-[25]. Successful attempts have also been made to use the conditional signal averaging methods in radioisotope measurements of two-phase flows [26], [27].

REFERENCES

[1] Bendat, J.S., Piersol, A.G. (2010). *Random Data: Analysis and Measurement Procedures* (Fourth Edition). Wiley.

[2] Assous, S., Linnett, L. (2012). High resolution time delay estimation using sliding discrete Fourier transform. *Digital Signal Processing*, 22, 820–827.

[3] Waschburger, R., Kawakami, R., Galvao, H. (2013). Time delay estimation in discrete-time state-space models. *Signal Processing*, 93, 904–912.

[4] Jacovitti, G., Scarano, G. (1993). Discrete time technique for time delay estimation. *IEEE Transactions on Signal Processing*, 41 (2), 525-533.

[5] Chen, J., Benesty, J., Huang, Y. (2006). Time delay estimation in room acoustic environments: An overview. *EURASIP Journal on Advances in Signal Processing*, 2006, 026503.

[6] Hanus, R., Zych, M., Petryka, L., Świsulski, D. (2014). Time delay estimation in two-phase flow investigation using the γ-ray attenuation technique. *Mathematical Problems in Engineering*, 2014, 475735.

[7] Chen, J., Huang, Y., Benesty, J. (2004). Time delay estimation. In *Audio Signal Processing for Next-Generation Multimedia Communication Systems*. Springer, 197-227.

[8] Blok, E. (2002). Classification and evaluation of discrete subsample time delay estimation algorithms. In *14th International Conference on Microwaves, Radar and Wireless Communications (MIKON-2002)*, May 20-22, 2002. IEEE, 764-767.

[9] So, H.C. (2001). On time delay estimation using an FIR filter. *Signal Processing*, 81, 1777-1782.

[10] Zych, M., Hanus, R., Vlasak, P., Jaszczur, M., Petryka, L. (2017). Radiometric methods in the measurement of particle-laden flows. *Powder Technology*, 318, 491-500.

[11] Mosorov, V. (2008). Flow pattern tracing for mass flow rate measurement in pneumatic conveying using twin plane electrical capacitance tomography. *Particle & Particle Systems Characterization*, 25 (3), 259-265.

[12] Zhang, L., Wu, X. (2006). On the application of cross correlation function to subsample discrete time delay estimation. *Digital Signal Processing*, 16, 682–694.

[13] Beck, M.S., Pląskowski, A. (1987). *Cross-Correlation Flowmeters*. CRC Press.

[14] Mosorov, V. (2006). Phase spectrum method for time delay estimation using twin-plane electrical capacitance tomography. *Electronics Letters*, 42 (11), 630-632.

[15] Hanus, R., Zych, M., Petryka, L., Mosorov, V., Hanus, P. (2015). Application of the phase method in radioisotope measurements of the liquid - solid particles flow in the vertical pipeline. *EPJ Web of Conferences*, 92, 02020.

[16] Hanus, R. (2015). Application of the Hilbert Transform to measurements of liquid-gas flow using gamma ray densitometry. *International Journal of Multiphase Flow*, 72, 210–217.

[17] Shors, S.M., Sahakian, A.V., Sih, H.J., Swiryn, S. (1996). A method for determining high-resolution activation time delays in unipolar cardiac mapping. *IEEE Transactions on Biomedical Engineering*, 43 (12), 1192-1196.

[18] Hanus, R. (2001). Accuracy comparison of some statistic methods of time delay measurements. *Systems Analysis Modelling Simulation*, 40 (2), 239-244.

[19] Kowalczyk, A., Szlachta, A. (2010). The application of conditional averaging of signals to obtain the transportation delay. *Przegląd Elektrotechniczny*, 86 (1), 225-228. (in Polish)

[20] Kowalczyk, A., Hanus, R., Szlachta, A. (2011). Time delay estimation of stochastic signals using conditional averaging. In *8th International Conference on Measurement (Measurement 2011)*, April 27-30, 2011. Bratislava, Slovakia: Institute of Measurement Science SAS, 32-37.

[21] Kowalczyk, A., Szlachta, A., Hanus, R., Chorzępa, R. (2017). Estimation of conditional expected value for exponentially autocorrelated data. *Metrology and Measurement Systems*, 24 (1), 69–78.

[22] Hanus, R. (2010). Standard uncertainty comparison of time delay estimation using cross-correlation function and the function of conditional average value of the absolute value of delayed signal. *Przegląd Elektrotechniczny*, 86 (6), 232-235. (in Polish)

[23] Kowalczyk, A., Hanus, R., Szlachta, A. (2011). Investigation of the statistical method of time delay estimation based on conditional averaging of delayed signal. *Metrology and Measurement Systems*, 18 (2), 335-342.

[24] Szlachta, A., Hanus, R., Kowalczyk, A. (2011). Virtual instrument for the estimation of the time delay using conditional averaging of random signals. In *International Workshop on ADC Modelling, Testing and Data Converter Analysis and Design and IEEE 2011 ADC Forum*, June 30 - July 1, 2011. IMEKO, 81-86.

[25] Kowalczyk, A., Hanus, R., Szlachta, A. (2016). Time delay measurement method using conditional averaging of the delayed signal module. *Przegląd Elektrotechniczny*, 92 (9), 279-282.

[26] Hanus, R., Szlachta, A., Kowalczyk, A., Petryka, L., Zych, M. (2012). Radioisotope measurement of two-phase flow in pipeline using conditional averaging of signal. In *16th IEEE Mediterranean Electrotechnical Conference (MELECON 2012)*, March 25-28, 2012. IEEE, 144-147.

[27] Hanus, R., Zych, M., Kowalczyk, A., Petryka, L. (2015). Velocity measurements of the liquid-gas flow using gamma absorption and modified conditional averaging. *EPJ Web of Conferences*, 92, 02021.

Received February 02, 2018.
Accepted July 16, 2018.

MEASUREMENT SCIENCE REVIEW, **18**, (2018), No. 4, 138-146

sciendo

ISSN 1335-8871

MEASUREMENT SCIENCE REVIEW

Journal homepage: http://www.degruyter.com/view/j/msr

A Fast-Multi-Channel Sub-Millikelvin Precision Resistance Thermometer Readout Based on the Round-Robin Structure

Jiong Ding, Suijun Yang, Shuliang Ye

Institute of Industry and Trade Measurement Technology, China Jiliang University, Xueyuan Street, No.258, 310018, Hangzhou, China, dingjiong@cjlu.edu.cn, corresponding author: Shuliang Ye, itmt_paper@126.com

The fast response multipoint high-precision temperature measurement is often necessary in many dynamical measurement fields and industrial applications. However, limited by the existing electric circuit architecture, either the AC or DC bridges have the shortcoming that the rates or precisions degenerate markedly in the multi-channel scanning mode. To overcome this disadvantage, a round-robin structural low-cost ratiometric resistance thermometer readout based on several commercial 32-bit sigma-delta analogue-to-digital converters (Σ-Δ ADCs) was presented in this article. The experimental results show that the precision of this readout corresponds to 0.1 mK at 1 Hz when sampling four channel resistors simultaneously, while the precision and rate are not degenerating with the channel number increasing. In addition, the uncertainty of the readout is investigated in this article. It shows that the presented readout can achieve an uncertainty as low as 2.1 mK at 1 Hz (K = 2).

Keywords: Temperature metrology, resistance thermometry, round-robin structure, multi-channel scanning.

1. INTRODUCTION

The fast response synchronous multipoint high-precision temperature monitoring is often necessary in many metrology fields and industrial applications. For example, in the mass comparator system, the influences from the variety of the environmental temperature should be dynamically compensated in order to satisfy the low uncertainty mass dissemination requirement [1]. In thin film platinum resistance sensor fabrication, the fast-multi-channel resistance thermometer readout is required for online calibration. For the investigation of the super-cooling phenomena of electrolyte solutions, the multi-point temperatures should be recorded precisely and rapidly [2]. Moreover, in many temperature field reconstruction applications based on several thin-film resistance sensors, the high precision and fast measurement speed multi-channel thermometer readout is always needed [3]. According to the exciting current, the thermometer bridges are divided into two categories, the AC and DC bridges. The typical AC bridges are F18 and F900 from WIKA. The typical DC bridges are MI 6015T, FLUKE 1595, and MicroK 70 [4]. The AC bridges and MI 6015T are based on the transformers. They are considered as the most accurate resistance thermometer readouts. Unfortunately, their high cost, large size and low speed discourage their use in fast multi-channel high precision temperature monitoring. The DC bridges, except the MI 6015T, are mainly based on the analogue-to-digital converter (ADC), named ratiometric resistance thermometer readout. The ratiometric resistance thermometer readout seems a possible solution for these requirements due to its low cost and high integration. However, whether the AC bridge or DC bridge, the ADC based bridge or the transformer-based bridge, its precision degenerates observably or its measurement speed reduces obviously when several channels are available. According to the above situation, developing a low-cost multi-channel resistance thermometer, which can overcome the precision degeneration or speed reduction when working in the multi-channel scanning mode, is significant in fast multipoint high-precision temperature monitoring.

With the great technology progress in noise reduction of electronic components, an increasing number of ADC based ratiometric thermometer readouts have been developed in recent years. These writings have pointed out the difficulty of obtaining precision at the sub-mK level, especially when it works in the multi-channel scanning mode. Schweiger designed a fast multi-channel precision thermometer readout with system noise less than 3 mK at 1 Hz [2]. He just copied the single channel structure eight times to extend the channel number. Smorgon designed a single channel low-cost ratiometric front end for industrial platinum resistance thermometer (PRT) applications. The equivalent temperature precision of this readout system was about 2 mK [5]. Ambrosetti developed a versatile and high-resolution readout system for resistance temperature detector (RTD). The system resolution (RMS) corresponds to 0.38 mK at 1 Hz [6].

DOI: 10.1515/msr-2017-0020

With the ingenious design and high-quality fabrication, some well-known benchtop primary ratiometric resistance thermometer readouts, such as the Model 1595A from Fluke-Hart Scientific incorporated or MicroK 70 from ISOTECH incorporated, have the uncertainties of 0.02 mK after 1-minute moving average, approximately. They also can be easily extended to a multi-channel system with the automatically controlled switchbox. However, the extended channels always share the same ADC. When they work in the multi-channel scanning mode, more time is required to finish one cyclic measurement or the sample time for each channel should be decreased to overcome the time cost increase. This leads to the precision degenerating markedly. Actually, to accomplish several channels, measurement in one second is not available in almost all of the primary resistance thermometer readouts, because in the primary standard temperature laboratory, the fast measurement is usually not needed. In order to guarantee the accuracy, the shortest sample period for each channel is one second.

To obtain the equivalent temperature precision better than 0.1 mK at 1 Hz in multi-channel scanning mode, a commercial 32-bit Σ-Δ ADC based round-robin structural ratiometric resistance thermometer readout is presented in this study. Through selecting a series of low temperature coefficient resistors corresponding to a PRT at different temperature values, the performance of the precision, the linearity and the uncertainty of the presented readout were investigated by long time testing.

2. DESCRIPTION OF THE MULTI-CHANNEL RESISTANCE THERMOMETER READOUT

A. The architecture of the classic ratiometric resistance thermometer readout

For the highest accuracy measurement, the four-wire PRT and four-wire standard reference resistor are essential in the ADC based ratiometric resistance thermometry. Traditionally, the current source reversing technique is employed for parasitic thermal electromotive forces (EMFs) and other systematic drift elimination [7]. The simplified schematic diagram of the single channel resistance thermometer readout is shown in Fig.1.a). A well-controlled constant direct current passes through the reference (REF) resistor and the PRT in series and develops two different voltage signals. Then, taking two pairs of readings with currents in opposite senses, the following are obtained:

$$ForwardV_{PRT} = \alpha[I_F \times R_{PRT} + EMF_1 + Drift_1] \quad (1)$$

$$ForwardV_{REF} = \alpha[I_F \times R_{REF} + EMF_2 + Drift_2] \quad (2)$$

$$BackwardV_{PRT} = \alpha[-I_B \times R_{PRT} + EMF_1 + Drift_1] \quad (3)$$

$$BackwardV_{REF} = \alpha[-I_B \times R_{REF} + EMF_2 + Drift_2] \quad (4)$$

where α is the gain of the amplifier. I_F and I_B are the values of forward and backward current. R_{PRT} and R_{REF} are the value of the unknown PRT and REF resistor, respectively. EMF_1 is the parasitic thermal electromotive force between PRT and the connector. EMF_2 is the parasitic thermal electromotive force between reference resistor and the connector. $Drift_1$ is the systematic drift of the electronic circuit (including the amplifier and the ADC) when measuring the voltage of the PRT. $Drift_2$ is the systematic drift of the electronic circuit when measuring the voltage of the REF. The average of the PRT and REF voltage values is given as:

$$V_{PRT} = ForwardV_{PRT} - BackwardV_{PRT} = \alpha[I_F + I_B] \times R_{PRT} \quad (5)$$

$$V_{REF} = ForwardV_{REF} - BackwardV_{REF} = \alpha[I_F + I_B] \times R_{REF} \quad (6)$$

Through computing the ratio of the measurement values of these average voltages by the microcontroller, the ratio of the REF resistor and PRT is obtained as the following equation:

$$M = \frac{R_{PRT}}{R_{REF}} = \frac{V_{PRT}}{V_{REF}} \quad (7)$$

The above method eliminates the sources of systematic errors like EMF, gain of amplifier, and circuit drift in resistance ratio computation. The accuracy of the temperature measurement mainly depends on the accuracy of the reference resistor [8]. However, when it extends to the multi-channel scanning mode, as shown in Fig.1.b), each channel shares the ADC through the mechanical relays. This noise of measurement system becomes larger, because the data conversion time of the ADC for each channel is reduced. Meanwhile, the common mode error generated from the CMRR (common-mode rejection ratio) of the amplifier has not been considered in the above equations. This is also a significant error in the ADC based ratiometric resistance thermometry [8].

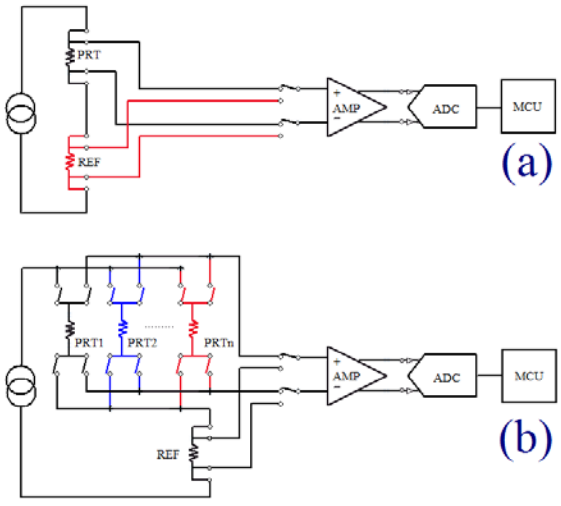


Fig.1. Topologies of ADC based ratiometric resistance thermometer readout [9]. The microcontroller controls the switchers, manages the ADC, computes the resistor's ratio, displays and sends the results. a) Classic single channel topology; b) Classic multi-channel extension topology. The combination of amplifier and ADC samples of the resistors in time-sharing mode.

B. The architecture and measurement procedure of the presented multi-channel readout

ADC is the core component in the ratiometric resistance thermometer readout. However, the high performance and well calibrated reference resistor is the most expensive one. If we copy the single channel architecture several times to avoid the precision degeneration in multi-channel scanning mode, the price will be very high. To overcome the precision degeneration in an economical method, an architecture based on the round-robin structure, as shown in Fig.2., is studied in this article. In this architecture, a constant direct current passes through one reference (REF) resistor and n PRTs in series and develops n+1 different voltage signals. Each voltage signal is amplified and sampled by the different amplifier and ADC. For ease of understanding, a two-channel round-robin architecture is selected as an example to describe the measurement procedure.

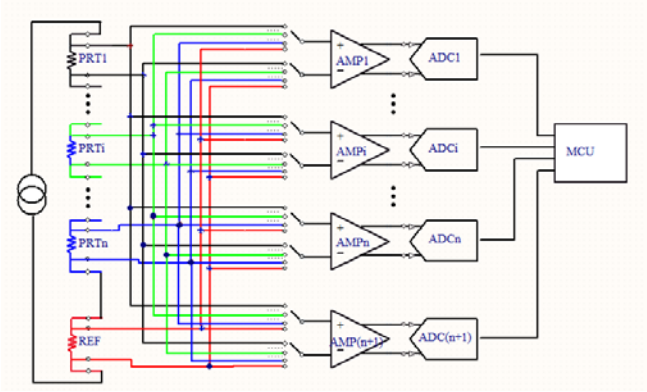


Fig.2. The architecture of the multi-channel readout based on the round-robin structure. There is one reference resistor and n PRTs, (n+1) amplifiers and ADCs. The black lines mean the signal from PRT1. The green lines mean the signal from PRTi. The blue lines mean the signal from PRTn. The red lines mean the signal from REF. Each amplifier and ADC combination samples every PRT and REF signal orderly according to the round-robin mechanism described in the followed paragraphs.

Step A1: the direction of the sense current is configured as forward, from the PRTs to the REF resistor. The ADC_1 samples the voltage from PRT_1. The ADC_2 samples the voltage from PRT_2. The ADC_3 samples the voltage from the REF resistor. The expressions of the input voltages of the ADCs are as follows:

$$\begin{bmatrix} FADC_{11} \\ FADC_{22} \\ FADC_{3r} \end{bmatrix} = \begin{bmatrix} (EMF_1 + Drift_1)\times\alpha_1 \\ (EMF_2 + Drift_2)\times\alpha_2 \\ (EMF_r + Drift_3)\times\alpha_3 \end{bmatrix} + \begin{bmatrix} \alpha_1 \times R_{PRT1} \times I_F \\ \alpha_2 \times R_{PRT2} \times I_F \\ \alpha_3 \times R_{REF} \times I_F \end{bmatrix} + \begin{bmatrix} \beta_1 \times (\frac{R_{PRT1}}{2} + R_{PRT2} + R_{REF}) \times I_F \\ \beta_2 \times (\frac{R_{PRT2}}{2} + R_{REF}) \times I_F \\ \beta_3 \times (R_{REF}) \times I_F \end{bmatrix} \quad (8)$$

where $FADC_{11}$ is the voltage from PRT_1 sampled by ADC_1. $FADC_{22}$ is the voltage from PRT_2 sampled by ADC_2. $FADC_{3r}$ is the voltage from REF sampled by ADC_3. α_1, α_2, and α_3 are the differential gains of the AMP_1, AMP_2, and AMP_3, respectively. β_1, β_2, and β_3 are the common gains of the AMP_1, AMP_2, and AMP_3, respectively. I_F is the value of forward current. EMF_1 and EMF_2 are the parasitic thermal electromotive forces between PRT_1, PRT_2 and their connectors, respectively. EMF_r is the parasitic thermal electromotive force between REF resistor and connector. $Drift_1$ is the systematic voltage drift of AMP_1 and ADC_1. $Drift_2$ is the systematic voltage drift of AMP_2 and ADC_2. $Drift_3$ is the systematic voltage drift of AMP_3 and ADC_3.

Step A2: the direction of the sense current remaining unchanged and modifying the switchers' connections between resistors and ADCs, the ADC_1 samples the voltage from PRT_2. The ADC_2 samples the voltage from the REF. The ADC_3 samples the voltage from the PRT_1. The expressions of the input voltages of the ADCs are as follows:

$$\begin{bmatrix} FADC_{12} \\ FADC_{2r} \\ FADC_{31} \end{bmatrix} = \begin{bmatrix} (EMF_2 + Drift_1)\times\alpha_1 \\ (EMF_r + Drift_2)\times\alpha_2 \\ (EMF_1 + Drift_3)\times\alpha_3 \end{bmatrix} + \begin{bmatrix} \alpha_1 \times R_{PRT2} \times I_F \\ \alpha_2 \times R_{REF} \times I_F \\ \alpha_3 \times R_{PRT1} \times I_F \end{bmatrix} + \begin{bmatrix} \beta_1 \times (\frac{R_{PRT2}}{2} + R_{REF}) \times I_F \\ \beta_2 \times (\frac{R_{REF}}{2}) \times I_F \\ \beta_3 \times (\frac{R_{PRT1}}{2} + R_{PRT2} + R_{REF}) \times I_F \end{bmatrix} \quad (9)$$

where $FADC_{12}$ is the voltage from PRT_2 sampled by ADC_1. $FADC_{2r}$ is the voltage from REF sampled by ADC_2. $FADC_{31}$ is the voltage from PRT_1 sampled by ADC_3.

Step A3: through changing the switchers' connections, the ADC_1 samples the voltage from REF. The ADC_2 samples the voltage from PRT_1. The ADC_3 samples the voltage from PRT_2. The expressions of the input voltages of the ADCs are as follows:

$$\begin{bmatrix} FADC_{1r} \\ FADC_{21} \\ FADC_{32} \end{bmatrix} = \begin{bmatrix} (EMF_r + Drift_1)\times\alpha_1 \\ (EMF_1 + Drift_2)\times\alpha_2 \\ (EMF_2 + Drift_3)\times\alpha_3 \end{bmatrix} + \begin{bmatrix} \alpha_1 \times R_{REF} \times I_F \\ \alpha_2 \times R_{PRT1} \times I_F \\ \alpha_3 \times R_{PRT2} \times I_F \end{bmatrix} + \begin{bmatrix} \beta_1 \times (\frac{R_{REF}}{2}) \times I_F \\ \beta_2 \times (\frac{R_{PRT1}}{2} + R_{PRT2} + R_{REF}) \times I_F \\ \beta_3 \times (\frac{R_{PRT2}}{2} + R_{REF}) \times I_F \end{bmatrix} \quad (10)$$

where $FADC_{1r}$ is the voltage from REF sampled by ADC_1. $FADC_{21}$ is the voltage from PRT_1 sampled by ADC_2. $FADC_{32}$ is the voltage from PRT_2 sampled by ADC_3.

Step B1: changing the direction of the sense current to the backward direction, from the REF resistor to the PRTs and modifying the switchers' connections between PRTs or REF and ADCs. The ADC_1 samples the voltage from PRT_1. The ADC_2 samples the voltage from PRT_2. The ADC_3 samples the voltage from REF resistor. The expressions of the input voltages of the ADCs are as follows:

$$\begin{bmatrix} BADC_{11} \\ BADC_{22} \\ BADC_{3r} \end{bmatrix} = \begin{bmatrix} (EMF_1 + Drift_1)\times\alpha_1 \\ (EMF_2 + Drift_2)\times\alpha_2 \\ (EMF_r + Drift_3)\times\alpha_3 \end{bmatrix} - \begin{bmatrix} \alpha_1 \times R_{PRT1} \times I_B \\ \alpha_2 \times R_{PRT2} \times I_B \\ \alpha_3 \times R_{REF} \times I_B \end{bmatrix} + \begin{bmatrix} \beta_1 \times (\frac{R_{PRT1}}{2}) \times I_B \\ \beta_2 \times (R_{PRT1} + \frac{R_{PRT2}}{2}) \times I_B \\ \beta_3 \times (R_{PRT1} + R_{PRT2} + \frac{R_{REF}}{2}) \times I_B \end{bmatrix} \quad (11)$$

where I_B is the value of backward current. $BADC_{11}$ is the voltage from PRT_1 sampled by ADC_1. $BADC_{22}$ is the voltage from PRT_2 sampled by ADC_2. $BADC_{3r}$ is the voltage from REF sampled by ADC_3.

Step B2: Through changing the switchers' connections, the ADC_1 samples the voltage from PRT_2. The ADC_2 samples the voltage from REF. ADC_3 samples the voltage from REF. The expressions of the input voltages of the ADCs are as follows:

$$\begin{bmatrix} BADC_{12} \\ BADC_{2r} \\ BADC_{31} \end{bmatrix} = \begin{bmatrix} (EMF_2 + Drift_1)\times\alpha_1 \\ (EMF_r + Drift_2)\times\alpha_2 \\ (EMF_1 + Drift_3)\times\alpha_3 \end{bmatrix} - \begin{bmatrix} \alpha_1 \times R_{PRT2} \times I_B \\ \alpha_2 \times R_{REF} \times I_B \\ \alpha_3 \times R_{PRT1} \times I_B \end{bmatrix} + \begin{bmatrix} \beta_1 \times (R_{PRT1} + \frac{R_{PRT2}}{2}) \times I_B \\ \beta_2 \times (R_{PRT1} + R_{PRT2} + \frac{R_{REF}}{2}) \times I_B \\ \beta_3 \times (\frac{R_{PRT1}}{2}) \times I_B \end{bmatrix} \quad (12)$$

where $BADC_{12}$ is the voltage from PRT_2 sampled by ADC_1. $BADC_{2r}$ is the voltage from REF sampled by ADC_2. $BADC_{31}$ is the voltage from PRT_1 sampled by ADC_3.

Step B3: the direction of the sense current remaining the same and modifying the switchers' connections, the ADC_1 samples the voltage from REF. The ADC_2 samples the voltage from PRT_1. The ADC_3 samples the voltage from PRT_2. The expressions of the input voltages of the ADCs are as follows:

$$\begin{bmatrix} BADC_{1r} \\ BADC_{21} \\ BADC_{32} \end{bmatrix} = \begin{bmatrix} (EMF_r + Drift_1)\times\alpha_1 \\ (EMF_1 + Drift_2)\times\alpha_2 \\ (EMF_2 + Drift_3)\times\alpha_3 \end{bmatrix} - \begin{bmatrix} \alpha_1 \times R_{REF} \times I_B \\ \alpha_2 \times R_{PRT1} \times I_B \\ \alpha_3 \times R_{PRT2} \times I_B \end{bmatrix} + \begin{bmatrix} \beta_1 \times (R_{PRT1} + R_{PRT2} + \frac{R_{REF}}{2}) \times I_B \\ \beta_2 \times (\frac{R_{PRT1}}{2}) \times I_B \\ \beta_3 \times (R_{PRT1} + \frac{R_{PRT2}}{2}) \times I_B \end{bmatrix} \quad (13)$$

where $BADC_{1r}$ is the voltage from REF sampled by ADC_1. $BADC_{21}$ is the voltage from PRT_1 sampled by ADC_2. $BADC_{32}$ is the voltage from PRT_2 sampled by ADC_3.

After the switchers' connections reconfigure three times with currents in opposite senses, each ADC obtains 2×3 different voltages from every PRT and REF. The sum of the PRTs and REF voltage values in opposite senses from ADC_1 is:

$$\begin{aligned} V_{11} &= FADC_{11} - BADC_{11} = \alpha_1 \times R_{PRT1} \times [I_F + I_B] \\ &+ \beta_1 \times [(\frac{R_{PRT1}}{2} + R_{PRT2} + R_{REF}) \times I_F - \frac{R_{PRT1}}{2} \times I_B] \end{aligned} \quad (14)$$

$$\begin{aligned} V_{12} &= FADC_{12} - BADC_{12} = \alpha_1 \times R_{PRT2} \times [I_F + I_B] \\ &+ \beta_1 \times [(\frac{R_{PRT2}}{2} + R_{REF}) \times I_F - (R_{PRT1} + \frac{R_{PRT2}}{2}) \times I_B] \end{aligned} \quad (15)$$

$$\begin{aligned} V_{1r} &= FADC_{1r} - BADC_{1r} = \alpha_1 \times R_{REF} \times [I_F + I_B] \\ &+ \beta_1 \times [\frac{R_{REF}}{2} \times I_F - (R_{PRT1} + R_{PRT2} + \frac{R_{REF}}{2}) \times I_B] \end{aligned} \quad (16)$$

With this procedure, the EMFs and systematic drifts are eliminated. The ratio of α to β is the CMRR of the amplifier. It can be calibrated prior to the measuring procedure. Due to the currents in opposite directions generated from the same constant current source, the forward current I_F and backward current I_B are approximately equal. Through computing the ratio of the measurement values of these average voltages, the ratio of the PRTs and REF resistor is obtained from ADC_1 as the following equations:

$$M_{11} = \frac{V_{11}}{V_{1r}} = \frac{R_{PRT1} + \frac{1}{2 \times CMRR}(R_{PRT2} + R_{REF})}{R_{REF} - \frac{1}{2 \times CMRR}(R_{PRT1} + R_{PRT2})} \quad (17)$$

$$M_{12} = \frac{V_{12}}{V_{1r}} = \frac{R_{PRT2} + \frac{1}{2 \times CMRR}(-R_{PRT1} + R_{REF})}{R_{REF} - \frac{1}{2 \times CMRR}(R_{PRT1} + R_{PRT2})} \quad (18)$$

In the above equations, the R_{REF} and $CMRR$ are the known quantities. Through the measurements, the values of M_{11} and M_{12} are obtained. The values of PRT_1 and PRT_2 can be obtained by solving these equations.

$$R_{PRT1} = \frac{4(CMRR)^2 M_{11} - 2CMRR(1+M_{12}) + M_{11} - M_{12} + 1}{4(CMRR)^2 + 2CMRR(M_{11} + M_{12}) + M_{11} - M_{12} + 1} \times R_{REF} \quad (19)$$

$$R_{PRT2} = \frac{4(CMRR)^2 M_{12} - 2CMRR(1+M_{11}) - M_{11} + M_{12} + 1}{4(CMRR)^2 + 2CMRR(M_{11} + M_{12}) + M_{11} - M_{12} + 1} \times R_{REF} \quad (20)$$

The ADC_2 and ADC_3 also sampled the voltages of PRT_1, PRT_2, and REF, respectively. According to the equations (8)-(13), the values of PRT_1 and PRT_2 can also be solved through the voltages from ADC_2 and ADC_3. Then averaging the values of PRT_1 and PRT_2 from the three ADCs, the averaged result is obtained:

$$\overline{PRT_i} = \frac{1}{2+1} \sum_{j=1}^{2+1} PRT_{ij} \quad (21)$$

where the $PRTij$ is the value of PRTi computed from ADCj. When the n PRTs want to be measured based on the round-robin structure, we need (n+1) ADCs. And the resistance value calculation method can be easily obtained through extending the above equations.

Compared with the single channel architecture, the major characteristic of the round-robin structure is that there are (n+1) ADCs or Σ-Δ engines that intersect sampling concurrently. The noise of the final result of PRTi is lower than the level from single ADC due to the parallel averaging, while the measurement speed has not declined compared with the single channel architecture. The other improvement of this measurement procedure is the common mode error correction. After CMRR correction and combining with the ADC linear calibration, the high linear performance of the resistance thermometer readout will be obtained.

C. The CMRR calibration based on the RBC

The principle of using the resistance bridge calibrator (RBC) to estimate the linearity of the thermometer readout is the least-squares fit [10]. According to the principle, the RBC can also be used for CMRR estimation. In this study, the RBC is a set of four resistors that can be connected in different configurations to generate a total of 35 distinct resistances [11]. Putting the RBC as the PRT_1, a 100 Ω high precision resistor as the PRT_2 in the above two-channel system, the values of the RBC can be obtained by the equation (19). According to the maximum likelihood estimation equation is as follows:

$$S^2 = \frac{1}{N-\rho}\sum_{i=1}^{N}(P_{i,meas} - P_{i,calc})^2 \qquad (22)$$

the variance of the differences between the measured and calculated values can be calculated. Where the N = 35 is the number of measured ratios. $P_{i,meas}$ are the measured ratios calculated by equation (19). $P_{i,cals}$ are the ratios calculated from the fitted values of RBCs. ρ is the negative number of fitted parameters. Here, it is 6 (four RBC resistances, CMRR, PRT2). With the least-squares fit by the function *fminsearch* in Matlab, the CMRR of the AMP_1 can be calibrated. Meanwhile, with the same procedure, the CMRR of the AMP_2 and AMP_3 can also be calibrated.

D. The details of key components in the multi-channel readout

The ADC, reference resistor, amplifier, and switcher are the key components of the readout system for noise and drift reduction. ADC is the kernel in ratiometric resistance thermometer readouts. With the development of digital integrated circuit design and chip manufacturing processes, the resolution of the Σ-Δ ADC achieved a significant progress in the recent years. In 2015, several commercial 32-bit Σ-Δ ADCs were available in the market. In this article, one of the available 32-bit Σ-Δ ADCs named AD7177-2 from Analog Device Incorporated is employed [12]. The output data rates of the AD7177-2 range from 5 samples per second (SPS) to 10,000 SPS. Its root mean square noise is lower than 0.1 μV at 16.67 SPS. This means the effective resolution of the AD7177-2 can achieve 0.01 ppm, approximately. Meanwhile, the AD7177-2 integrates several digital filters for both the 50 Hz and the 60 Hz power line interference rejection. For each AD7177-2, like for the most of commercial ADCs, there are several input channels, but only one Σ-Δ engine. In order to realize the round-robin structure, it needs (n+1) chip ADCs.

The stability and reputability of readout depends largely on the performance of the reference resistor. The reference resistor should have low temperature coefficient and time drift. In this design, a 100 Ω ultra-high precision resistor named VHP203 from Vishay Precision Group is employed as the reference resistor [13]. With the bulk metal foil technology, hermetic sealing and oil filling, the temperature coefficient of the resistor is better than ±0.2 ppm/K. Its shelf life stability is better than 2 ppm for at least 6 years. Its tolerance is up to ±10 ppm. To obtain the higher temperature measurement accuracy, the reference resistor is calibrated by a direct current comparator bridge (MI 6015T) before soldering. The outcome of this calibration is 100.0002154 Ω. The uncertainty is 8.7 μΩ (k = 1). Also, the designed multi-channel readout can use a high-performance external standard resistor as the reference for more accuracy requirement, such as Tinley 5685A.

Instrumentation amplifier is also the key component in the ratiometric resistance readout. Due to the tolerance of the gain resistor, the gain of each channel is different to each other. However, the gain difference and drift can be eliminated through the above mathematical analysis. The most important parameters for selecting an amplifier in the ratiometric resistance readout are the noise, distortion, and common-mode rejection ratio (CMRR). The AD8422 is a high precision, low noise, rail-to-rail instrumentation amplifier from Analog Device Incorporated [14]. Its peak-to-peak output noise is 0.15 μV. The nonlinearity is better than 0.5 ppm. The minimum CMRR is 110 dB at Gain = 10.

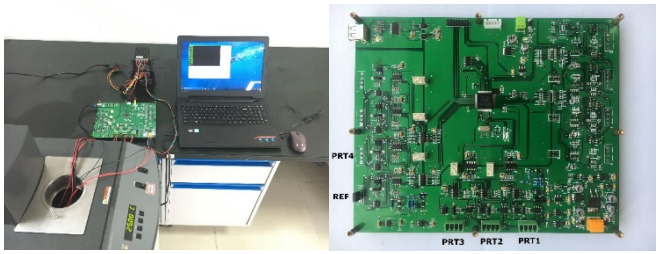
Another important component in the readout is the switcher. There are two main kinds of switchers, one is mechanical relay, and the other one is CMOS multiplexer. In this design, the CMOS multiplexer ADG888 from Analog Device Incorporated is selected as the switcher between amplifier and resistor, because of its ultralow resistance distortion [15]. The reason of using the CMOS multiplexer is that it has faster response then the relay. It is suitable for rapid channel switch. The other advantage of using the CMOS multiplexer is reducing the size and power consumption of the readout. This significantly decreases the parasitic thermal EMFs on the switch.

3. RESULTS

A. Noise performance estimation

To assess the noise performance of the round-robin structural resistance thermometer readout, a series of different value four-wire resistors kept in the oil bath (Fluke 7341 at 295.15 K) with a peak-to-peak stability better than 0.05 K are used in this investigation. The multi-channel readout is configured as single channel mode, double channel mode, triple channel mode, and quadruple channel mode. The whole experimental setup is shown in Fig.3. The values of the tested resistors are 50 Ω, 100 Ω, 150 Ω, and 350 Ω, respectively. In the single channel mode, each value resistor was measured independently for noise performance

estimation. In the double channel mode, the combination of 50 Ω and 100 Ω, the combination of 100 Ω and 150 Ω, the combination of 150 Ω and 350 Ω, and the combination of 50 Ω and 350 Ω were measured for noise performance estimation. In the triple channel mode, the combination of 50 Ω, 100 Ω, 150 Ω, the combination of 50 Ω, 100 Ω and 350 Ω, the combination of 50 Ω 150 Ω and 350 Ω, and the combination of 100 Ω, 150 Ω and 350 Ω were measured. In the quadruple channel mode, the combination of 50 Ω, 100 Ω, 150 Ω and 350 Ω was measured. Setting the exciting current as 1 mA (the uncertainty of the current is 15.1 nA (k = 1)) and the data out rate as 1 Hz without moving average, the standard deviation is calculated in Fig.4. for different value resistors and channel modes.



a) b)

Fig.3. The photos of experimental systems. a) The tested system, including resistors, circuit, computer, power supply and oil bath. b) The photo of the designed readout.

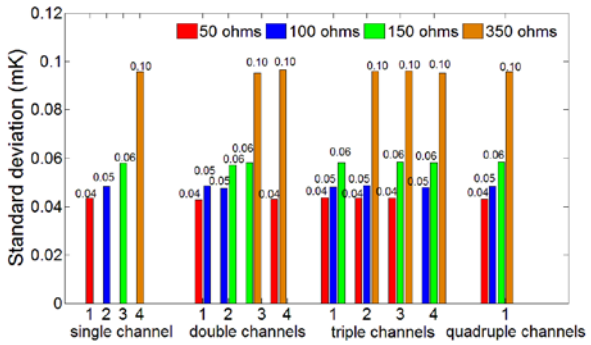


Fig.4. The noise performance of the round-robin structural readout in different channel configuration modes. There are four kinds of different value resistors in this test. It corresponds to the temperature range from 144 K to 900 K.

Almost all the primary resistance thermometers are not available in the fast scanning mode. In the article, a like-for-like comparison of the noise performance between the round-robin structural readout and a commercial portable readout, named Fluke 1529, is carried out. The Fluke 1529 is a four-channel integrating ADC based thermometry readout. Its schematic is similar to the topology shown in Fig.1.a). It can work in single channel mode, double channel mode and quadruple channel mode. Actually, Fluke 1529 cannot measure four channels in a second, because when it works in the multi-channel scan mode, the shortest sample period the user can set is 0.5 seconds. Using the Fluke 1529 to measure the above resistors, setting the exciting current as 1 mA, the standard deviation is calculated in Fig.5. In order to keep the time for one cycle measurement equal in each mode, the sample period in the single channel mode is set as 2 second; it is set as 1 second in the double channel mode; it is set as 0.5 seconds in the quadruple channel mode.

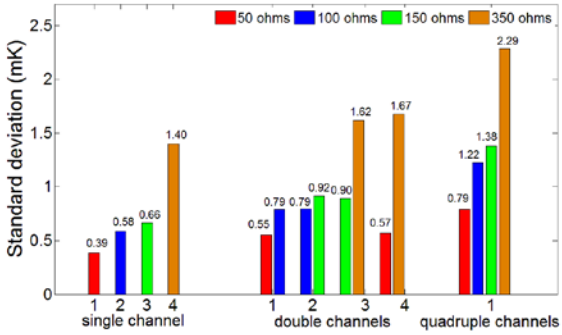


Fig.5. The noise performance of the Fluke 1529 in different channel configuration modes. The value of the internal reference resistor is 100 Ω, approximately. In the single channel mode, the integrating time of the ADC is longer than in the multi-channel scan mode through modifying the instrument configuration.

The above results show that standard deviation of the measured data is increasing with the resistor's value. In the presented readout, the precision was not degenerating with the channel number increasing. The precision and the number of channels is independent of one another. Considering the test resistor as the 100 Ω SPRT, the equivalent temperature precision of the readout achieves 0.1 mK at 1 Hz in either number of channel modes, while in the Fluke 1529, the precision degenerates with the channel number increasing. Because the Fluke 1529 uses the classic multi-channel extension topology as shown in Fig.1.b), these channels share the ADC conversion time in the multi-channel scanning mode.

Table 1. The precision comparison between designed multi-channel readout and the other classic resistance thermometers.

Bridge	**Value [Ω]**	**Current [mA]**	**Average time [second]**	**Standard deviation [mK]**
Our readout	100	1.0	32	0.020
ASL F900	100	1.0	/	0.005
FLUKE 1595A	100	1.0	30	0.010
Isotech MicroK 70	25	1.0	/	0.009
MI 6020T	100	1.0	/	0.005

In resistance thermometry, the output ratio between standard platinum resistance thermometer (SPRT) and reference resistor is always moving average for ten or sixty seconds, because the dynamic response time of the large diameter SPRT is very slow. To compare the precision with some well-known resistance bridges and ratiometric resistance thermometers, the relationship between the

precision and moving average number is estimated in the quadruple channel mode. As shown previously, the noise depends largely on the window size of moving average *n*. When *n* is set as 32, the noise performance of the designed readout is approaching several primary resistance thermometer readouts, as shown in Table 1. The data of noise performance of the well-known resistance thermometers come from their official specifications.

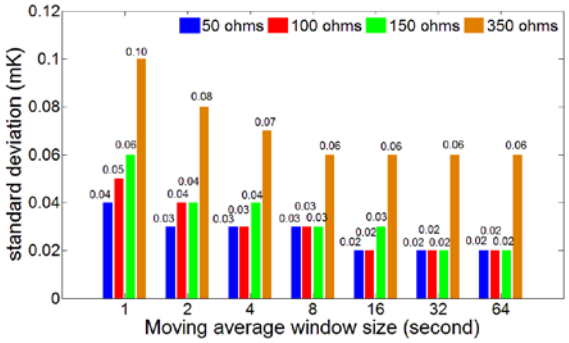


Fig.6. The standard deviation of the round-robin structural readout under different moving average window sizes.

B. *Self-heating correction and linearity evaluation*

In this design, the self-heating effects are corrected by the zero-power extrapolation method. It is a mature method in the resistance thermometry field [16]. In this study, the values of the self-heating current are 1 mA and $\sqrt{2}/2$ mA. From the self-heating correction point of view, the round-robin structure has the advantage compared to the classic multi-channel extension topology in the multi-channel mode, because in the round-robin structure, all the resistors are in series. No matter which resistor is sampled, the current heats the resistor stably. While in the classic multi-channel extension topology as shown in Fig.1.b), when one resistor is sampled, the current will not pass through the other resistors. This makes the temperature of the resistor or PRT unstable. The measurement noise becomes large.

After the self-heating and CMRR correction, the linearity of the designed multi-channel readout is evaluated using a resistance bridge calibrator (RBC). The manual RBC includes a set of four ultra-high precision resistors from Vishay Precision Group that can be configured in series or parallel combinations to produce resistances ranging from 43 Ω to 346 Ω (The nominal values of the four resistors are 77.1862 Ω, 100.0000 Ω, 129.8168 Ω, and 216.8144 Ω). The linearity evaluation principle proposed by Dr. Rod White from Measurement Standards Laboratory of New Zealand is adopted in this investigation [16]. The RBC is connected to channel PRT_1 of the designed readout. The channels of PRT_2, PRT_3, and PRT_4 are connected to three precision resistors. The ratio of resistance between the calibrator and the reference resistor was measured in the quadruple channel mode. The linearity evaluation results are shown in Fig.7. The nonlinear error is smaller than 0.5 ppm in the whole range. In the ADC based resistance thermometry readout, this error mainly comes from the ADC and amplifier. However, the total nonlinear error measured in this investigation is smaller than the nonlinear error value of the AD7177-2 and the gain nonlinearity of AD8422 in their datasheets. This is because for getting one ratio between reference resistor and PRT, the voltages sampled from the five ADCs are averaged.

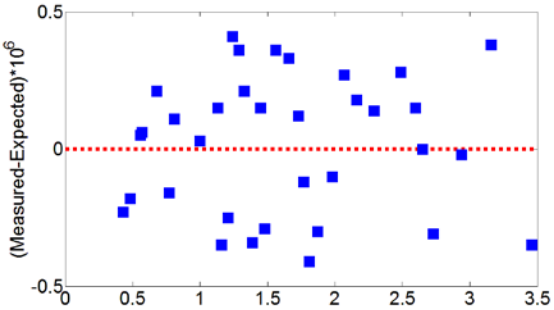


Fig.7. Residuals of nonlinearity evaluation of the designed multi-channel resistance thermometer readout.

C. *Uncertainty investigation*

The uncertainty contributions of the thermometer readout are the linearity, measurement noise, and reference resistor stability [5], [17]. The uncertainty of the linearity is calculated as the largest deviation with respect to the fit line. As shown in Fig.7., the largest deviation is 0.49 ppm. The uncertainty of the measurement noise comes from the standard deviation calculated in Fig.4. The uncertainty of the reference resistor stability is calculated from the datasheet of VHP203. We assume the environment temperature range of the designed thermometer readout working is from 295.15 K to 300.15 K. The uncertainty of the reference resistor stability is 1 ppm according to the following uncertainty combination equation. The results are shown in Table 2.

$$u_O = \sqrt{\left(\left|\frac{\partial PRT}{\partial Ratio}\right| \times u_{Ratio}\right)^2 + \left(\left|\frac{\partial PRT}{\partial REF}\right| \times u_{REF}\right)^2 + \left(u_{linearity}\right)^2} \tag{23}$$

Table 2. The uncertainty investigation of the multi-channel readout.

Value [Ω]	u_{Ratio} [ppm] (k=1)	u_{REF} [ppm] (k=1)	$u_{linearity}$ [ppm] (k=1)	u_O [μΩ] (k=1)	Equivalent uncertainty [mK] (k=2)
50	0.15	1.00	0.49	56.84	0.30
100	0.19	1.00	0.49	112.97	0.59
150	0.23	1.00	0.49	153.52	0.80
350	0.39	1.00	0.49	391.71	2.03

The results show that the equivalent uncertainty (K = 2, at 1 Hz) of the multi-channel readout is better than 2.1 mK. Compared with some well-known primary resistance bridges, the measurement uncertainty of the multi-channel readout keeps a certain distance [18], [19]. The reason is that the temperature of the reference resistor has not been controlled

and the nonlinear of the ADCs in the designed multi-channel readout has not been well calibrated yet.

D. Temperature coefficient evaluation

Putting the designed readout in the drying oven at different temperatures, the temperature coefficient and noise performance of the readout working in the quadruple channel mode is evaluated. Setting the data out rate as 1 Hz, the noise performance at different temperatures is shown in Fig.8. The results show that the noise performance is increasing with the environmental temperature imperceptibly, due to the Johnson noise becoming larger with the temperature.

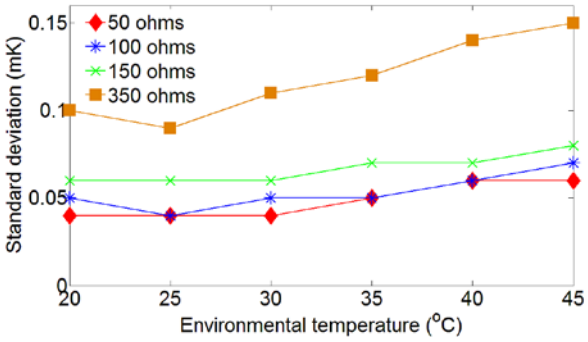


Fig.8. Standard deviation under different environmental temperatures.

Through comparison of the measured averaged ratios in 5 minutes under different environmental temperatures, the temperature coefficient of the readout is evaluated. Fig.9. shows the measuring drift under different environmental temperatures. The linear fitting result shows that the temperature coefficient of the readout is lower than 0.2 mK/K (at ratio range from 0.5 to 3.5), when the environmental temperature changes from 20 °C to 45 °C. According to the ratio measurement principle, the main contribution of the temperature coefficient of the readout comes from the reference resistor. The temperature coefficient of the VHP 203 precision resistor is lower than 0.2 ppm/K. This means the readout can adapt to some hostile industrial application environments.

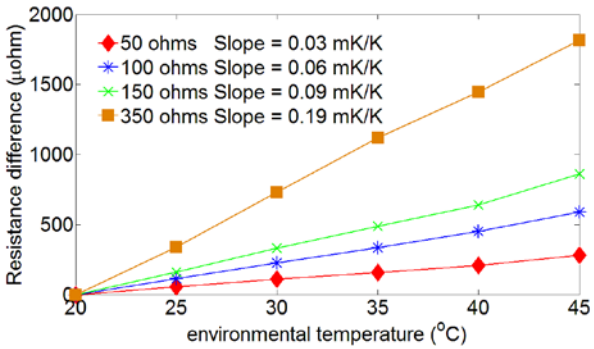


Fig.9. The temperature coefficient of the readout when measuring different value resistors.

4. Discussion / Conclusions

In this article, a fast-multi-channel sub-millikelvin precision resistance thermometer readout is presented based on the round-robin structure. When it works in four channel scanning mode, the precision corresponds to 0.1 mK at 1 Hz. The overall uncertainty of the designed readout is better than 2.1 mK at 1 Hz. The main advantage of the round-robin structure is that it can avoid the precision or rate degeneration in the multi-channel scanning mode. In the round-robin architecture, for any step, each ADC samples the different PRT or reference resistor, respectively. After a whole measurement procedure, each ADC obtains the ratio between every PRT and reference resistor. Through averaging the ratio from all the ADCs, the precision of temperature measurement improves greatly.

The other advantage is the low cost of the designed multi-channel readout. Firstly, there is only one expensive reference resistor. Compared with the channel number extension through sample coping, the round-robin architecture saves the cost of multiple reference resistors. Secondly, in the designed readout, a kind of the newest commercial 32-bit Σ-Δ ADC is employed as the engine of the system. Typically, the integrating ADC is employed in digital multimeter and ratiometric resistance thermometer readout for a long time. However, its precision is largely limited by the clock frequency and jitter. The precision of the commercial integrating ADC is usually not higher than 24 bits. Walker uses several integrating ADCs parallelly for single channel to noise reduction in FLUKE 1595A [20]. This is contrary to the cost reduction. Bramley and Pickering developed the MicroK 70 established on a new type Σ-Δ ADC licensed by the National Physical Laboratory (NPL) [8]. This ADC is an application specific integrated circuit (ASIC). The manufacturing cost is not low. However, in the designed readout, we use the universal commercial ADCs. The price is just several dollars. By using this 32-bit Σ-Δ ADC, the precision of this readout is 0.1 mK at 1 Hz. It is close to the precision of some well-known resistance bridges.

Meanwhile, the errors generated by the CMRR of the amplifiers have been considered in this article. This error exists in all the ADC based resistance thermometer readouts. In order to reduce this error, Bramely presented a substitution topology in MicroK 70. In this topology, the PRT and the reference resistor are connected in parallel [8]. However, this topology cannot eliminate this error completely. When the difference between the PRT and reference resistor is considerable, this error cannot be ignored. In the round-robin structure, the error generated by the CMRR is larger, because all the resistors are in series. It should be considered cautiously. In this article, the CMRR of the amplifiers are calibrated first. Then, through solving the equations (19) and (20), this error is eliminated. Furthermore, distinguishing and calibrating the error from the CMRR and nonlinearity of ADC, respectively, should be adopted in ADC based

resistance thermometer readouts calibration. After CMRR and ADC calibration, the better linearity should be obtained.

ACKNOWLEDGMENT

This work was supported by Zhejiang Provincial Natural Science Foundation of China (Grant No. LQ15F030003, LQ17F010011), Zhejiang Key Discipline of Instrument Science and Technology (Grant No. JL150501).

REFERENCES

[1] Frohlich, T., Fehling, T., Heydenbluth, D. (2009). Mass dissemination using a robot system. *tm – Technisches Messen*, 76 (7-8), 382-387.

[2] Schweiger, H.G., Multerer, M., Gores, H.J. (2007). Fast multichannel precision thermometer. *IEEE Transactions on Instrumentation and Measurement*, 56 (5), 2002-2009.

[3] Eke, R., Kavasoglu, A.S., Kavasoglu, N. (2012). Design and implementation of a low-cost multi-channel temperature measurement system for photovoltaic modules. *Measurement*, 45, 1499-1509.

[4] Pearce, J.V., Gray, J., Veltcheva, R.I. (2016). Characterisation of a selection of AC and DC resistance bridges for platinum resistance thermometry. *International Journal of Thermophysics*, 37, 109.

[5] Smorgon, D., Fernicola, V.C., Coslovi, I. (2011). Low-cost ratiometric front-end for industrial PRT applications. *International Journal of Thermophysics*, 32, 2317-2324.

[6] Ambrosetti, R., Matteoli, E., Ricci, D. (2012). Note: A versatile, stable, high-resolution readout system for RTD and thermistor sensors. *Review of Scientific Instruments*, 83, 096101.

[7] Taylor, H.R., Navarro, H.A. (1983). A microcomputer-based instrument for applications in platinum resistance thermometry. *Journal of Physics E: Scientific Instruments*, 16, 1100-1104.

[8] Bramley, P., Pickering, J.R. (2006). Better accuracy in temperature calibration and measurement through a new type analog-to-digital converter. *Cal Lab Magazine*, 10/11, 21-26.

[9] Walker, R., Willgress, N. (2003). Achieving 0.25 mK uncertainty with an integrated-circuit resistance thermometer readout. In *XVII IMEKO World Congress*, Dubrovnik, Croatia, 1594-1597.

[10] White, D.R., Williams, J.M., Ramsey I.E. (1997). A simple resistance network for the calibration of resistance bridges. *IEEE Transactions on Instrumentation and Measurement*, 42 (5), 1068-1074.

[11] White, D.R., Clarkson, M.T., Saunders, P., Yoon, H. (2008). A general technique for calibrating indicating instruments. *Metrologia*, 45, 199-210.

[12] Analog Devices, Inc. (2015-2016). *32-bit, 10 kSPS, Sigma-Delta ADC with 100 µs settling and true rail-to-rail buffers*. Data Sheet AD7177-2. http://www.analog.com/media/en/technical-documentation/data-sheets/AD7177-2.pdf.

[13] Vishay Precision Group. (2014). *Hermetically sealed miniature ultra high precision Z-foil technology resistors with TCR of 0.05 ppm/°C, tolerance of ± 0.001 % and load life stability of ± 0.005 %, unaffected by humidity*. VHP203 (Z-Foil). http://www.vishaypg.com/docs/63146/vhp203.pdf.

[14] Analog Devices, Inc. (2013-2015). *High performance, low power, rail-to-rail precision instrumentation amplifier*. Data Sheet AD8422. http://www.analog.com/media/en/technical-documentation/data-sheets/AD8422.pdf.

[15] Analog Devices, Inc. (2005-2017). *0.4Ω CMOS, dual DPDT switch in WLCSP/LFCSP/TSSOP*. Data Sheet ADG888. http://www.analog.com/media/en/technical-documentation/data-sheets/ADG888.pdf.

[16] Nicholas, J.V., White, D.R. (2001). *Traceable Temperatures: An Introduction to Temperature Measurement and Calibration*. John Wiley & Sons.

[17] Walker, R. (2010). Recent advances in resistance thermometry readouts. In *NCSL International Workshop and Symposium.*

[18] Palencar, R., Sopkuliak, P., Palencar, J., Duris, S., Suroviak, E., Halaj, M. (2017). Application of Monte Carlo method for evaluation of uncertainties of ITS-90 by standard platinum resistance thermometer. *Measurement Science Review*, 17 (3), 108-116.

[19] Duris, S., Palencar, R., Ranostaj, J. (2008). Contribution of the SPRT calibration to uncertainty of temperature T-90 measured by the calibrated SPRT. *Measurement Science Review*, 8 (1), 5-10.

[20] Walker, R. (2011). Automatic linearity calibration in a resistance thermometry bridge. *International Journal of Thermophysics*, 32 (1-2), 215-223.

Received January 18, 2018.
Accepted July 13, 2018.

MEASUREMENT SCIENCE REVIEW, **18**, (2018), No. 4, 147–157

sciendo

ISSN 1335-8871

MEASUREMENT SCIENCE REVIEW

Journal homepage: http://www.degruyter.com/view/j/msr

The Algebraic Structure of Quantity Calculus

Álvaro P. Raposo[1]

[1]*Department of Applied Mathematics, Universidad Politécnica de Madrid, Av. Juan de Herrera, 6, 28040, Madrid, Spain, alvaro.p.raposo@upm.es*

The algebraic structure underlying the quantity calculus is defined axiomatically as an algebraic fiber bundle, that is, a base structure which is a free Abelian group together with fibers which are one dimensional vector spaces, all of them bound by algebraic restrictions. Subspaces, tensor product, and quotient spaces are considered, as well as homomorphisms to end with a classification theorem of these structures. The new structure provides an axiomatic foundation of quantity calculus which is centered on the concept of dimension, rather than on the concept of unit, which is regarded as secondary, and uses only integer exponents of the dimensions.

Keywords: Quantity calculus, algebraic structure, group of dimensions.

1. Introduction

The foundations of quantity calculus have been the subject of research for more than a century, and recent papers as those by Kitano [1], Krystek [2], Atkey et al. [3] or Domotor [4] show that its formalization is not completely settled yet. Moreover, in the last years many authors are calling for a discussion of fundamental concepts regarding quantity calculus [5, 6, 7, 8, 9]. In fact, the first two papers mentioned represent two different visions on the subject. The paper by Kitano assumes a structure of quantity calculus centered in the definition of units, upon which the rest of the structure is built. In contrast, the paper by Krystek recovers the call of some authors in past decades to locate the concept of dimension in the center of the scene. The latter view is what the widespread use of dimensional analysis (see, e.g., [10, 11]) suggests that the structure of quantity calculus should be, but the fact is that there is not a well defined algebraic structure supporting it. The goal of this paper is to provide such a structure. On another basis, the paper by Domotor deserves a separate comment, for it dives into ontological discussion of the relationship between quantities and their values and provides an algebraic tool to describe it by means of torsors, which is further commented below.

By quantity calculus it is understood the algebra of the operations performed between physical quantities, which are three: product of quantities, product of a number times a quantity and addition of quantities of the same kind. It is customary to write a quantity q as the product of a number times a unit which, in the present notation of the International Vocabulary of Metrology (VIM) [12], is denoted as

$$q = \{q\}\,[q], \tag{1}$$

where $[q]$ stands for the unit and $\{q\}$ is the number of times q comprises that unit. Notice that, as the VIM explicitly warns in its Note 5 under the definition of quantity, this concept is that of a scalar quantity, and the way to consider vector or tensor quantities is through their components which accommodate to this definition. The operations are then performed as

$$q_1 q_2 = (\{q_1\}\{q_2\})\,([q_1][q_2]), \tag{2a}$$

$$\alpha q = (\alpha\{q\})\,[q], \tag{2b}$$

$$q_1 + q_2 = (\{q_1\} + \{q_2\})\,[q_1], \tag{2c}$$

where α is a number, q, q_1, and q_2 are quantities and, in the last equation, $[q_1] = [q_2]$ is assumed. Although this way of performing the operations is standard, the algebraic structure to which these three operations give rise is still under study, and it is the particular form of equation (2c) what makes it quite different from other usual structures such as rings, vector spaces or algebras.

The history of quantity calculus, in the words of Domotor [4], is the story of the search for the intrinsic algebraic structures underlying dimensional analysis, measurement units and measurement uncertainty of quantity values. This history is unsually long, and the reader is referred to the excellent review by de Boer [13] for a detailed account. However, a few landmarks are worth mentioning in order to situate the contribution of the present paper. Its origin goes back to Fourier in 1822 [14], when he introduced the idea of dimension of a quantity and the concept of homogeneity of dimensions in a valid physical equation, thus initiating the view centered on dimensions. Contemporarily, Gauss, in 1832 [15], proposed the first so called absolute system of

DOI: 10.1515/msr-2017-0021

units, relying on the units for length, mass and time, which eventually became the cgs system. Another step is due to Maxwell in 1873 [16], when he stated that the expression of a quantity has two factors or components, as in equation (1); in fact, the present notation in that equation comes from that of Maxwell. In 1914 Buckingham published his celebrated Pi Theorem [17], although there was not a solid framework on which to base its proof. For the remaining duration of the 20th century there is a quest for the formalism underlying the quantity calculus. According to de Boer, Wallot was the first to claim the concept of quantity as central, and not just as the product of a number and a unit. Then Landolt [18], in 1945, made the first attemp to give an axiomatic foundation, with the important step of recognizing the group properties of the operations between quantities. In the 1950s Fleischmann [19], also according to de Boer, took another step forward by distinguishing the quantities from the kinds of quantities, the latter exhibiting the structure of a group. In the following years several authors contributed to set up a description of the algebraic structure but, curiously enough, two divergent paths were followed. On the one hand authors such as Fleischmann, Quade [20] or de Boer himself called for a description centered on the quantities and the concept of dimension. On the other hand, authors such as Drobot [21], Whitney [22], and Carlson [23] developed an algebraic structure which resembled that of quantity calculus, but was centered on the concept of unit over which the rest of the set of quantities was built. The recent paper by Kitano [1] can be seen as the zenith of this viewpoint, for he introduces a theory of comparison of the different systems of units which provides new insight into the subject. In a rough summary, this structure starts with a system of units $\{u_1, \ldots, u_n\}$ and writes any quantity q in a unique way as

$$q = \alpha\, u_1^{r_1} \cdots u_n^{r_n}, \tag{3}$$

where α is a real number and $r_1, \ldots, r_n$ are rational numbers. Therefore, the algebraic structure depicted by this theory is $\mathbb{R} \times \mathbb{Q}^n$, where the factor $\mathbb{R}$ hosts the numerical value of q relative to this system of units, while $\mathbb{Q}^n$ hosts the rational exponents of the units and exhibits a linear space structure. However, the rules (not the algebraic properties) to perform the three operations need to be set in the axioms, which is not quite acceptable in an axiomatic description of an algebraic structure. As a consequence of the central role of the unit systems, we can find quantities which, dependending on the unit system of choice, can be compared or cannot, can be added or cannot.

Notwithstanding its merits, the accepted description of the algebraic structure of quantity calculus has two drawbacks that need to be addressed. Firstly, as said before, it relies on the concept of unit as central. This is not satisfactory since units are the result of an arbitrary agreement and, thus, can be easily changed, while the concept of dimension is more resilient. (The arbitrariness, of course, is only from the algebraic viewpoint; in fact, the choice of base units in a system of units, say the SI, is far from arbitrary, for they need to be as constant and accurate as possible and reproducible). The history of the units in electromagnetism at the end of the 19th century and beginning of the 20th shows it clearly: the mess of different systems of units, all of them trying to describe electromagnetic quantities in a purely mechanical context, was definitely clarified with the proposal of Giorgi [24] to introduce a separate unit for one of the electromagnetic quantities, that is, introducing a new and independent dimension.

Secondly, the use of fractional exponents in the units. It is a remarkable fact that, albeit the possibilities allowed by Bridgman's theorem on dimensional analysis [25], dimensionful physical quantities enter into physically valid equations only through the three operations mentioned at the beginning. Therefore, only integer exponents of quantities, and thus of units, should be expected. Although the use of fractional exponents is widespread, we claim that it is not necessary. Several examples come to mind, as the case of square roots of quantities (the use of Pythagoras' theorem, the period of a pendulum, the standard deviation of a random quantity…) but in all cases the square root acts on a quantity which is already a square. An algebraic structure for quantity calculus, which allows fractional exponents, is oversized.

It is noticeable that Quade [20], following Fleischmann, developed a description of the algebraic structure entirely with integer exponents of the units instead of fractional ones, and also tried to center it on the concept of dimension. Unfortunately, this approach, which pointed in the correct direction, was superseded by the one described above. The cause of moving the focus from dimensions to units is, probably, the absence of a known algebraic structure which fits exactly the operations between quantities. But, fortunately, the recent paper by Krystek [2] resumes the quest for this structure.

Drawing from the ideas of Quade and Krystek, the goal of this paper is to introduce a new algebraic structure, based on a simple set of axioms, which accounts exactly for the properties of quantity calculus and overcomes the two aforementioned problems. The structure is centered in the group of dimensions and the quantities are placed upon it gathered in fibers over the dimensions. Hence, the structure can be described as an algebraic fiber bundle. The axioms allow to define systems of units and, as a consequence, but not as a definition, to write quantities as in equation (1). Then the properties of the maps $\{\cdot\}$ and $[\cdot]$ are studied and the validity of equations (2) is established (not postulated, as usual). These are the contents of section 2. Section 3 shows how to construct new spaces of quantities from old ones as subspaces or by means of tensor products or quotients. The quotient of a space of quantities is seen to be the tool to reduce dimensions as it is usually done, for instance, to get natural units. In section 4 the tool for the comparison is defined and studied: the homomorphism of spaces of quantities, which allows us to characterize and classify the spaces. This classification is compared with that of Kitano in a final section of conclusions and also a comment on Domotor's torsor theory is considered.

2. GROUP OF DIMENSIONS AND SPACE OF QUANTITIES

Our object of study is a set Q of quantities and the operations defined within it, which is referred to as a space of quantities. The elements of Q will be denoted by lowercase latin letters, particularly q, r, s. These quantities, as it is detailed below, can be multiplied and added among them and also multiplied by scalar numbers from a field F, whose elements will be denoted by lowercase Greek letters, particularly α, β.

Example 2.1. *For further reference, we assign symbols to the following spaces of quantities:* Q_{geom}*, the space of quantities of geometry, that is, all the quantities needed to deal with lengths, areas, volumes, angles, etc.* Q_{time}*, the space of quantities to measure time.* Q_{kin}*, the space of quantities of kinematics.* Q_{mech}*, the space of quantities of mechanics.* Q_{phys}*, the space of quantities of physics.*

2.1. Group of dimensions

As noticed in the introduction, main role in the structure is played by the dimension of a quantity. Under this viewpoint, the dimension is an intrinsic property of a quantity, in contrast to its numerical value, which depends on the unit chosen, or the unit itself, which can be changed arbitrarily. Therefore, each quantity must have a firm link with its dimension in the present scheme better than a link with a unit or its numerical value with respect to that unit, despite the latter being what equation (1) suggests. To that end, let us first define properly the set of dimensions. The properties which characterize this set have been well described in the paper by Krystek [2], and they are just summarized here: dimensions can be multiplied and show the structure of an Abelian group with two further properties which characterize this group. First, no element is torsion, for there is no dimensionful quantity which multiplied by itself finitely many times becomes a quantity of dimension one. Second, it is finitely generated. Therefore, we adopt the following definition.

Definition 2.2. *A group of dimensions is a finitely generated free Abelian group.*

In this paper such a group is generally denoted by $\mathscr{D}$ and its elements by uppercase letters in roman sans-serif type such as $\mathsf{A}, \mathsf{B}\ldots$ (as stated in the VIM). The identity element of the group, denoted $1_{\mathscr{D}}$, is the dimension of the so called dimensionless quantities (quantities of dimension one are preferred).

Two properties of finitely generated free Abelian groups (or, equivalenty, free $\mathbb{Z}$-modules) are of interest to us [26]. In the first place, there exists the concept of basis: an independent (finite) set of generators. If $\{\mathsf{A}_1, \ldots, \mathsf{A}_k\}$ is such a basis for a group $\mathscr{D}$ then any element B has a unique expression in terms of the form $\mathsf{B} = \mathsf{A}_1^{n_1} \cdots \mathsf{A}_k^{n_k}$, where the exponents $n_1, \ldots, n_k$ are integer numbers. The number k of generators of any basis is called the rank of the group, and is a characteristic property of it. In the second place, such a group is isomorphic with the direct product of k infinite cyclic groups: $\mathscr{D} \cong \langle \mathsf{A}_1 \rangle \times \cdots \times \langle \mathsf{A}_k \rangle$, where $\langle \mathsf{A}_i \rangle = \{\mathsf{A}_i^n : n \in \mathbb{Z}\}$.

Example 2.3. *The groups of dimensions of the systems of quantities given in example 2.1 are, respectively, the following:* $\mathscr{D}_{\text{geom}} = \langle \mathsf{L} \rangle$*, the free Abelian group generated by* L*, which denotes length.* $\mathscr{D}_{\text{time}} = \langle \mathsf{T} \rangle$*, generated by* T *(time).* $\mathscr{D}_{\text{kin}} = \langle \mathsf{L}, \mathsf{T} \rangle$*, generated by* L *and* T*.* $\mathscr{D}_{\text{mech}} = \langle \mathsf{L}, \mathsf{T}, \mathsf{M} \rangle$*, generated by* L*,* T *and* M *(mass).* $\mathscr{D}_{\text{phys}} = \langle \mathsf{L}, \mathsf{T}, \mathsf{M}, \mathsf{I}, \Theta \rangle$*, generated by* L*,* T*,* M*,* I *(electric current) and* Θ *(temperature).*

2.2. Space of quantities

The link between a quantity and its dimension is made by means of a projection map dim: $Q \to \mathscr{D}$. This map is a surjection.

Example 2.4. *If* h *is Planck's constant then* $\dim(h) = \mathsf{L}^2\mathsf{T}^{-1}\mathsf{M}$*; and* θ*, the angle at a vertex of a triangle, yields* $\dim(\theta) = 1_{\mathscr{D}}$*; each one in the appropriate setting.*

In order to reflect that the dimension of a product of quantities is the product of the dimensions of the quantities the projection map must be a homomorphism with respect to the product of quantities.

In this way we follow the general concept studied by Atkey et al. [3] of a fiber bundle approach to quantities and dimensions. All the quantities with the same dimension, say A, form a set called a fiber, for it can be written as the inverse image of that dimension: $\dim^{-1}(\mathsf{A})$. As the VIM explicitly states, quantities of the same kind belong to the same fiber, while the opposite is not necessarily true. However, the algebraic structure cannot distinguish this detail. In each fiber quantities can be added and multiplied by scalars in a field F, resulting in quantities of the same dimension. These operations give the fiber the structure of a vector space over the field F. Moreover, since the comparison of each quantity in the fiber with a reference in the fiber, the unit, yields a single number, as in equation (1), that vector space is one dimensional (the latter in the sense of vector space dimension over F). There are some quantities which are intrinsically positive (mass, absolute temperature) and, thus, a full linear space seems to be oversized for them. However, differences of these quantities must also be considered in the framework. Since the algebraic structure cannot distinguish if a quantity is an absolute temperature or a difference of temperatures, the full linear space structure has to be allowed.

The field F is usually assumed to be that of the real numbers but so far there is no algebraic reason to restrict the definition to it. We are now ready to give an axiomatic definition of a space of quantities which takes into account all the aforementioned elements.

Definition 2.5. *A space of quantities with group of dimensions* $\mathscr{D}$ *over the field* F *is a set* Q*, together with a surjective map* dim: $Q \to \mathscr{D}$ *such that:*

(i) for each $\mathsf{A} \in \mathscr{D}$*, the fiber* $\dim^{-1}(\mathsf{A})$ *has the structure of a one dimensional vector space over* F*,*

(ii) there is a product defined in Q which makes it into an Abelian monoid and the map dim *is a monoid homomorphism, that is, for $q, r \in Q$,*

$$\dim(qr) = \dim(q)\dim(r),$$

and

(iii) the product distributes over the addition in each fiber, that is, for $q, r_1, r_2 \in Q$ with $\dim(r_1) = \dim(r_2)$,

$$q(r_1 + r_2) = qr_1 + qr_2,$$

and the product associates with the product by scalars in the sense of

$$\alpha(qr) = (\alpha q)r,$$

where $q, r \in Q$ and $\alpha \in F$.

The rank of Q is the rank of its group of dimensions.

This structure can be thought of as an algebraic fiber bundle, where the base structure is the group $\mathscr{D}$ and where over each element of it we place a fiber which is a one dimensional vector space. If A is a dimension and we denote $Q_{\mathsf{A}} = \dim^{-1}(\mathsf{A})$ the fiber over it, then the space is made of the disjoint union of its fibers $Q = \cup_{\mathsf{A}\in\mathscr{D}} Q_{\mathsf{A}}$. Therefore, this structure is a particular instance of the structure presented in [3] as a pair dimension group – fibers: $\langle \mathscr{D}, \{Q_{\mathsf{A}}\}\rangle$.

The fibers are not independent, for they have algebraic bounds given by the condition of the projection map being a monoid homomorphism. All fibers are isomorphic as vector spaces, and isomorphic to the field F, but there is one fiber of particular interest: the fiber $\dim^{-1}(1_{\mathscr{D}})$, the set of quantities of dimension one. The identity element in Q is denoted 1_Q and, since dim is a homomorphism, necessarily $\dim(1_Q) = 1_{\mathscr{D}}$, so 1_Q is a quantity of dimension one, as expected. Therefore, the fiber $\dim^{-1}(1_{\mathscr{D}})$ is not only an F-vector space, but an F-algebra of dimension 1, that is, naturally isomorphic with the field F by means of the isomorphism assigning the number 1 in F with 1_Q. For this reason, this fiber can be identified with F when needed.

2.3. System of units

We now turn to the task of defining system of units. It has been noticed that a system of units is nothing but a choice of a nonzero quantity of each dimension, that is, a basis in each fiber. Remember that a system of units is called coherent if the product of the units of any two quantities q and r gives the unit in the system for the quantity qr. The tool for a precise definition is the concept of section, which is a map that chooses one, and only one, element in Q from each fiber.

Definition 2.6. *A section of the space of quantities Q is a map $\sigma : \mathscr{D} \to Q$ such that $\dim \circ \sigma = \mathrm{id}_{\mathscr{D}}$. A section is called coherent if the map is a group homomorphism. The zero section, denoted σ_0, is the section which selects the zero element of each fiber. A nonzero section is a section none of which images is a zero element.*

Then we have the following definition.

Definition 2.7. *A system of units in a space of quantities is a nonzero section of it. The system is called coherent if the section is coherent.*

Before proceeding further, a word on the zeros of Q is necessary. Since each fiber has a zero element there are many zeros in the space Q, all of which constitute $\sigma_0(\mathscr{D})$, the image of the zero section. In this construction each zero has a dimension, so $0\,\mathrm{m\,s^{-1}}$ is a different quantity than $0\,\mathrm{kg}$.

Therefore, rather than speaking of the zero element, in this structure we have to speak of a zero element to refer ourselves to any of these elements in the image of the zero section. Despite of that, when no confusion is possible we write $q = 0$ to symbolize that the quantity q is a zero, without stating explicitly its dimension. Nevertheless, these zeros behave as is expected from an ordinary zero: the product of a quantity with a zero is a zero, as can be easily verified. However, it must be noticed that there is nothing in the definition of a space of quantities to prevent the existence of zero divisors, i.e. nonzero quantities q and r such that their product qr is a zero. As an extreme example consider a space of quantities with a product defined as $qr = 0$ for any dimensionful quantities q and r; it satisfies all the axioms of definition 2.5. Zero divisors, if any, are by no means isolated for, if q is a zero divisor, then αq, with $\alpha \in F$ is also a zero divisor, so the entire fiber of q is made of zero divisors. Also if s is another quantity such that sq is not zero, then sq is another zero divisor. Of course zero divisors do not show up in spaces of quantities of actual measurements, therefore in what follows we only consider spaces of quantities free of zero divisors. Some advantages we gain from that are collected in the next proposition.

Proposition 2.8. *In a space of quantities the following properties are equivalent:*

i. There are no zero divisors.

ii. The set of nonzero quantities is a group with respect to the product.

iii. There exists a coherent system of units.

PROOF. We show the first property to be equivalent to each of the other two. In order to show that the set of nonzero quantities is a group we need to show it is closed under the product and every element has an inverse. The absence of zero divisors is just the former condition, so let us show the inverse of a nonzero quantity q. Let $\tilde{q}$ be a nonzero quantity in the inverse fiber of q, meaning $\dim(\tilde{q}) = \dim(q)^{-1}$. Therefore, $q\tilde{q}$ is nonzero and dimensionless and, thus, there is a nonzero scalar α such that $q\tilde{q} = \alpha 1_Q$. The quantity $s = \alpha^{-1}\tilde{q}$ satisfies $qs = 1_Q$. On the contrary, if there are zero divisors, the set of nonzero quantities is not closed under the product, so it is not a group.

Now assume again Q is free of zero divisors and define a section $\sigma : \mathscr{D} \to Q$ by assigning a nonzero element in the corresponding fiber to each element in a basis of $\mathscr{D}$ and the rest

of elements by asking σ to be a homomorphism. The absence of zero divisors ensures that σ is a nonzero, in addition to coherent, section. On the contrary, assume now that q and r are nonzero elements of Q such that qr is a zero, and let σ be a coherent section of Q. Assume $\sigma(\dim(q))$ and $\sigma(\dim(r))$ are nonzero. Then they are of the form $\sigma(\dim(q)) = \alpha q$ and $\sigma(\dim(r)) = \beta r$ for some nonzero α and β in F. We have $\sigma(\dim(qr)) = (\alpha q)(\beta r) = \alpha\beta\, qr$ which is a zero. Hence, a coherent section is not nonzero, so there exits no coherent system of units. □

The following is an explicit construction of a space of quantities given a group of dimensions $\mathscr{D}$. Moreover, in section 4 we justify that, to some extent, this is the only example of a space of quantities free of zero divisors.

Example 2.9. *For a field F and a finitely generated free Abelian group $\mathscr{D}$, the set $F \times \mathscr{D}$ together with the map* dim*: $F \times \mathscr{D} \to \mathscr{D}$ which projects onto the second component, and the operations*

$$\begin{aligned}(\alpha,\mathsf{A}) + (\beta,\mathsf{A}) &= (\alpha+\beta,\mathsf{A}),\\ \beta(\alpha,\mathsf{A}) &= (\beta\alpha,\mathsf{A}),\\ (\alpha,\mathsf{A})(\beta,\mathsf{B}) &= (\alpha\beta,\mathsf{AB}),\end{aligned}$$

for $\alpha,\beta \in F$ and $\mathsf{A},\mathsf{B} \in \mathscr{D}$, becomes a space of quantities free of zero divisors. A coherent system of units is given from a group homomorphism $\chi : \mathscr{D} \to F^*$, where F^* denotes the multiplicative group of the field, by $\sigma : \mathscr{D} \to F \times \mathscr{D} : \mathsf{A} \mapsto (\chi(\mathsf{A}),\mathsf{A})$.

The first goal of a formalization of quantity calculus is to justify within the formalism the actual way in which operations between quantities are performed, that is, with the aid of a system of units and operating with the numerical values and with the units separately. The following paragraphs do this. Let us start by writing down the expression of any quantity in Maxwell's form. Let q be a quantity in a space Q free of zero divisors, and let σ be a system of units in Q. The dimension of q is $\dim(q)$ and the unit in its fiber is $\sigma(\dim(q))$. Now, since the latter is not a zero, by proposition 2.8 it has an inverse so we can define the map $\nu : Q \to F$, in which we make use of the identification of the field F with the fiber $\dim^{-1}(1_{\mathscr{D}})$ of quantities of dimension one, by

$$\nu(q) = q\,\sigma(\dim(q))^{-1}. \tag{4}$$

The quantities q and $\sigma(\dim(q))$ have the same dimension, so the product in equation (4) gives a quantity of dimension one which, after identification with an element of F, can be regarded as a number: the numerical value of q with respect to the unit $\sigma(\dim(q))$. Then we have

$$q = \nu(q)\,\sigma(\dim(q)), \tag{5}$$

where we identify $\nu(q)$ with $\{q\}$ and $\sigma(\dim(q))$ with $[q]$ as given in equation (1). In other words, the symbols $\{\cdot\}$ and $[\cdot]$ are nothing but the maps $\{\cdot\} = \nu$ and $[\cdot] = \sigma \circ \dim$. Let us study their algebraic properties.

Proposition 2.10. *In a space of quantities free of zero divisors, the map $[\cdot] : Q \to Q$ verifies*

(i) for q_1, q_2 quantities in the same fiber, and α, β in F

$$[\alpha q_1 + \beta q_2] = [q_1] = [q_2],$$

(ii) it is a homomorphism with respect to the product of quantities if and only if σ is a coherent section.

PROOF. Both items stem directly from the splitting of $[\cdot]$ as the composition $\sigma \circ \dim$. For the first one, since q_1 and q_2 are in the same fiber, then so is $\alpha q_1 + \beta q_2$, so it is clear that $\dim(\alpha q_1 + \beta q_2) = \dim(q_1) = \dim(q_2)$ and, therefore, the same applies to the map $[\cdot]$. For the second item, if σ is a group homomorphism, then $[\cdot]$ is the composition of two homomorphisms with respect to the product, so it is also a homomorphism. For the other way around, if $[\cdot]$ is a homomorphism, so is σ because the map dim is surjective. □

Proposition 2.11. *In a space of quantities free of zero divisors, the map $\{\cdot\} = \nu$ defined by equation (4) is*

(i) an F-linear homomorphism and

(ii) a homomorphism with respect to the product of quantities if and only if σ is a coherent section.

PROOF. For the first item consider two quantities q_1 and q_2 in the same fiber and two scalars α and β in the field F and compute $\{\alpha q_1 + \beta q_2\} = (\alpha q_1 + \beta q_2)\,\sigma(\dim(\alpha q_1 + \beta q_2))^{-1}$. Since $\dim(\alpha q_1 + \beta q_2) = \dim(q_1) = \dim(q_2)$ as in the previous proposition, the former expression can be written as $\alpha q_1 \sigma(\dim(q_1))^{-1} + \beta q_2 \sigma(\dim(q_2))^{-1}$, that is, $\alpha\{q_1\} + \beta\{q_2\}$.

In the second item the if part is trivial. For the only if part consider A and B in $\mathscr{D}$ and choose two nonzero quantities q and r such that $\mathsf{A} = \dim(q)$ and $\mathsf{B} = \dim(r)$. Then $\sigma(\mathsf{AB}) = qr\{qr\}^{-1}$, since qr is not a zero, then $\{qr\} \neq 0$. Now, because $\{\cdot\}$ is a homomorphism with respect to the product, the previous expresion gives $q\{q\}^{-1} r\{r\}^{-1} = \sigma(\mathsf{A})\,\sigma(\mathsf{B})$. □

These propositions set the condition to operate with quantities in the usual way: for quantities q_1 and q_2 in the same fiber and scalars α and β

$$\{\alpha q_1 + \beta q_2\} = \alpha\{q_1\} + \beta\{q_2\}; \quad [\alpha q_1 + \beta q_2] = [q_1] = [q_2],$$

and for any quantities q and r, only in case of a coherent system of units,

$$\{qr\} = \{q\}\{r\}; \quad [qr] = [q][r].$$

In other words, the equations (2) have been justified from the axioms.

3. NEW SPACES FROM OLD ONES

3.1. Subspace

Definition 3.1. *A subset S of a space of quantities Q is a subspace if, with the operations of Q and the restriction of the projection map, it is a space of quantities.*

Since the projection map dim restricted to S is the projection map of S, its image, $\dim(S)$, must be a subgroup of $\mathscr{D}$. Fortunately, it is a well known result of group theory that a subgroup of a free Abelian group is itself free Abelian [26, Theorem 10.17]. This is equivalent to the condition of S being closed under the product of quantities. In particular, $1_{\mathscr{D}}$ is in this subgroup. The subset S is also closed under addition of quantities of the same fiber and product by scalars so, if s is a nonzero quantity in S, then αs, for any scalar α, is also in S. In other words, the complete fiber containing s is in S, as should be, and for the fibers in S must be one dimensional vector spaces. This observation rules out a fiber in S containing only the zero element. As a consequence, the fiber of dimensionless quantities is contained in S. Thus, we have characterized the subspaces of a space of quantities as follows.

Proposition 3.2. *Let Q be a space of quantities with projection map* dim: $Q \to \mathscr{D}$. *A subset $S \subset Q$ is a subspace if and only if it is of the form $S = \dim^{-1}(\mathscr{E})$, where $\mathscr{E}$ is a subgroup of the group of dimensions $\mathscr{D}$.*

Trivially $Q = \dim^{-1}(\mathscr{D})$ is a subspace, arising from the improper subgroup of $\mathscr{D}$, and so is the fiber $\dim^{-1}(1_{\mathscr{D}})$, the subspace arising from the trivial subgroup of $\mathscr{D}$. Some nontrivial examples follow.

Example 3.3. *The space Q_{geom} can be seen is a subspace of Q_{kin}, for $Q_{\text{geom}} = \dim^{-1}(\langle \mathsf{L} \rangle)$, and $\langle \mathsf{L} \rangle$ is a subgroup of $\mathscr{D}_{\text{kin}}$. Analogously, Q_{kin} as a subspace of Q_{mech}, which in turn is a subspace of Q_{phys}.*

But also $\dim^{-1}(\langle \mathsf{L}^2 \rangle)$ is a subspace of Q_{geom} with group of dimensions $\langle \mathsf{L}^2 \rangle$.

3.2. Tensor product

From the examples one intuitively expects to be able to build the space Q_{kin} of kinematics quantities from Q_{geom} and Q_{time}, the spaces of quantities of geometry and time, respectively. But a simple cartesian product is not enough, for running 10 meters in 1 second gives the same speed as running 100 meters in 10 seconds. The technique is somewhat similar to the tensor product of linear spaces so we adopt the notation. Let Q and R be spaces of quantities free of zero divisors, over the field F and with groups of dimensions $\mathscr{D}_Q$ and $\mathscr{D}_R$, respectively, and projection maps $\dim_Q$ and $\dim_R$.

In the set $Q \times R$ define the element (q_1, r_1) to be related to (q_2, r_2) if there is $\alpha \in F$ such that $q_1 = \alpha q_2$ and $r_2 = \alpha r_1$ or such that $q_2 = \alpha q_1$ and $r_1 = \alpha r_2$. It is straightforward to check it is an equivalence relation. The quotient set is denoted $Q \otimes R$ and the equivalence class of the element (q, r) is denoted $q \otimes r$. Notice $\alpha q \otimes r = q \otimes \alpha r$, in particular $q \otimes 0 = 0 \otimes 0 = 0 \otimes r$.

We now define a structure of space of quantities in $Q \otimes R$. Its group of dimensions is the direct product of $\mathscr{D}_Q$ and $\mathscr{D}_R$, which is a free Abelian group with rank the sum of the ranks of each one. The projection map is defined by $\dim(q \otimes r) = (\dim_Q(q), \dim_R(r))$, which is well defined because all the elements in the class $q \otimes r$ have the same image under dim. Define a product in $Q \otimes R$ by $(q_1 \otimes r_1)(q_2 \otimes r_2) = (q_1 q_2) \otimes (r_1 r_2)$, which is independent of the representatives chosen, is commutative and associative and has an identity element: the class $1_Q \otimes 1_R$.

Define the product of the scalar $\gamma \in F$ times $q \otimes r$ by $\gamma(q \otimes r) = (\gamma q) \otimes r = q \otimes (\gamma r)$. Finally, define the addition of two elements in the same fiber $q_1 \otimes r_1$ and $q_2 \otimes r_2$ in the following manner. From the absence of zero divisors and proposition 2.8 there are nonzero $q \in Q$ and $r \in R$ such that $q_i = \alpha_i q$ and $r_i = \beta_i r$ for some $\alpha_i, \beta_i \in F$, $i \in \{1, 2\}$; define $q_1 \otimes r_1 + q_2 \otimes r_2 = (\alpha_1 \beta_1 + \alpha_2 \beta_2)(q \otimes r)$. The addition and product by scalars in the set of elements of a fiber satisfy the properties of a vector space over F and, moreover, this vector space is of dimension one, for, if q and r are nonzero elements and α and β are arbitrary scalars, then $(\alpha q) \otimes (\beta r) = (\alpha \beta)(q \otimes r)$. The zero element in each fiber is $0 \otimes 0$.

Finally, it is also straightforward to see the projection map behaves well under the product: $\dim((q_1 \otimes r_1)(q_2 \otimes r_2)) = \dim(q_1 \otimes r_1)\dim(q_2 \otimes r_2)$. Then we have the following result.

Proposition 3.4. *The set $Q \otimes R$, together with the operations defined above, is a space of quantities over the field F with group of dimensions $\mathscr{D}_Q \times \mathscr{D}_R$ and rank* $\operatorname{rank}(Q) + \operatorname{rank}(R)$.

The spaces Q and R can be identified, respectively, with $Q \otimes \dim_R^{-1}(1_{\mathscr{D}_R})$ and $\dim_Q^{-1}(1_{\mathscr{D}_Q}) \otimes R$, which are subspaces of $Q \otimes R$.

Example 3.5. *As announced before, we have $Q_{\text{kin}} \cong Q_{\text{geom}} \otimes Q_{\text{time}}$. Also $Q_{\text{mech}} \cong Q_{\text{kin}} \otimes Q_{\text{mass}}$.*

3.3. Quotient space

The quotient space is a construction intended to reduce the rank of a space of quantities by identifying certain quantities of different dimensions. The quotient cannot be taken with respect to a subspace, but another kind of subset of Q, namely, that given by a subsection.

Definition 3.6. *Let Q be a space of quantities free of zero divisors. A subsection of Q is the restriction of a nonzero coherent section $\sigma : \mathscr{D} \to Q$ to a subgroup $\mathscr{E}$ of $\mathscr{D}$.*

Its image $\Sigma = \sigma(\mathscr{E})$, which is also called subsection for brevity, is the intersection of the subspace $\dim^{-1}(\mathscr{E})$ and $\sigma(\mathscr{D})$, the image of the section. Notice that $1_Q \in \Sigma$ because the section σ is coherent. With the aid of Σ we can define an equivalence relation in Q. The quantity q_2 is equivalent modulo Σ to the quantity q_1 if $q_2 = q_1 s$ for some quantity s in Σ. It is reflexive for, as noticed before, $1_Q \in \Sigma$. It is symmetric, because s is not a zero and, by the absence of zero divisors, it is invertible in Q and $q_1 = q_2 s^{-1}$, where s^{-1} is in Σ because

the section is coherent. Finally, if $q_2 = q_1 s$ and $q_3 = q_2 s'$ for s and s' in Σ then $q_3 = q_1 ss'$, so q_3 is equivalent modulo Σ to q_1 since ss' is in Σ by the coherence of σ.

The quotient set of this equivalence relation is denoted Q/Σ and its elements, the equivalence classes, are of the form $q\Sigma$, which denotes the set of the elements qs with s running in Σ. We now provide the quotient set with suitable operations to convert it into a space of quantities. First we describe its group of dimensions. Since equivalent elements q_1 and $q_2 = q_1 s$ are identified in the quotient set, their dimensions must be identified as well. The obvious candidate for the group of dimensions is, thus, the quotient group $\mathscr{D}/\mathscr{E}$. In such a case it is only natural to define the projection map, $\widehat{\dim}\colon Q/\Sigma \to \mathscr{D}/\mathscr{E}$ by making the following diagram commutative.

$$\begin{array}{ccc} Q & \xrightarrow{\ \rho\ } & Q/\Sigma \\ {\scriptstyle \dim}\downarrow & & \downarrow{\scriptstyle \widehat{\dim}} \\ \mathscr{D} & \xrightarrow{\ \hat{\rho}\ } & \mathscr{D}/\mathscr{E} \end{array}$$

where the maps ρ and $\hat{\rho}$ are the natural projections of each set into its respective quotient set. In other words, $\hat{\rho}\circ\dim = \widehat{\dim}\circ\rho$. Unfortunately, the quotient of a free Abelian group is not necessarily free Abelian and, hence, $\mathscr{D}/\mathscr{E}$ does not necessarily qualify as a group of dimensions. Therefore, though the algebraic structure is well defined, the subsection must be carefully chosen so as the subgroup $\mathscr{E}$ makes the quotient $\mathscr{D}/\mathscr{E}$ a free Abelian group. For instance, in the group of dimensions of kinematics quantities, $\mathscr{D}_{\text{kin}} \cong \langle \mathsf{L}\rangle \times \langle \mathsf{T}\rangle$ the quotient $\mathscr{D}/\langle \mathsf{L}\rangle \cong \langle \mathsf{T}\rangle$ is free Abelian, while the quotient $\mathscr{D}/\langle \mathsf{L}^2\rangle \cong \langle \mathsf{L}\rangle/\langle \mathsf{L}^2\rangle \times \langle \mathsf{T}\rangle$ is not. From now on we assume that $\mathscr{E}$ is chosen so as to make $\mathscr{D}/\mathscr{E}$ free Abelian.

The product in Q/Σ is defined by the rule $\big(q_1\Sigma\big)\big(q_2\Sigma\big) = (q_1q_2)\Sigma$ which is easily checked to be independent of representatives. We have to check the condition which links the product and the projection map, but $\widehat{\dim}\big(q_1\Sigma q_2\Sigma\big) = \widehat{\dim}\big((q_1q_2)\Sigma\big) = \widehat{\dim}\circ\rho(q_1q_2)$ by the definition of the product in Q/Σ and the definition of ρ. Now, by the commutativity of the diagram and because both, dim and $\hat{\rho}$, are homomorphisms, the latter expression equals $\hat{\rho}\circ\dim(q_1q_2) = \hat{\rho}\circ\dim(q_1)\hat{\rho}\circ\dim(q_2) = \widehat{\dim}(q_1\Sigma)\widehat{\dim}(q_2\Sigma)$, so we conclude that $\widehat{\dim}$ is a monoid homomorphism.

The product with a scalar α from the field F is defined by $\alpha\big(q\Sigma\big) = (\alpha q)\Sigma$, which is also independent of the choice of representative q in the class $q\Sigma$. For the addition notice that if $q_1\Sigma$ and $q_2\Sigma$ are elements in the same fiber in Q/Σ, i.e. $\widehat{\dim}(q_1\Sigma) = \widehat{\dim}(q_2\Sigma)$, its sum cannot be defined simply as $(q_1+q_2)\Sigma$, because q_1 and q_2 need not be in the same fiber in Q. We only know $\dim(q_2) = \dim(q_1)\mathsf{A}$ for some $\mathsf{A}\in\mathscr{E}$. Denote $s = \sigma(\mathsf{A})$, an element in Σ, and define $q_1' = q_1 s$, so $q_1\Sigma = q_1'\Sigma$, hence $\dim(q_1') = \dim(q_1)\dim(s) = \dim(q_2)$, so they are in the same fiber in Q. Now we can define the addition as $\big(q_1\Sigma\big) + \big(q_2\Sigma\big) = (q_1' + q_2)\Sigma$. We could have taken instead an equivalent element of q_2 in the fiber of q_1 getting the same result. In the fiber of $q\Sigma$, the zero element is the class $q_0\Sigma$, where q_0 is the zero in the fiber of q, and is formed by the zeros of the fibers of Q represented in the class $q\Sigma$.

It is straightforward to check that the conditions of definition 2.5 hold for Q/Σ, so we state the result as follows.

Proposition 3.7. *If $\mathscr{E}$ is a subgroup of $\mathscr{D}$ such that $\mathscr{D}/\mathscr{E}$ is free Abelian then the set Q/Σ, together with the operations defined above, is a space of quantities with group of dimensions $\mathscr{D}/\mathscr{E}$ and rank given by* $\operatorname{rank}(\mathscr{D}) - \operatorname{rank}(\mathscr{E})$.

The mechanism of taking quotients is the algebraic tool underlying what is common practice in physics of choosing "systems of units" such that some specified universal constants become dimensionless and take on the numerical value 1. The extreme examples of this procedure are the systems of natural units, which do not cease to be a current subject of research as seen in [27]. But it has to be remarked that the mechanism goes beyond a change of system of units; it is indeed a change of space of quantities.

Example 3.8. *A usual agreement in particle physics is to "choose units such that $c = \hbar = 1$", where c is the speed of light and $\hbar$ is the reduced Planck's constant, and all the quantities are measured in powers of units of energy. In fact, it is a reduction of the space of quantities of mechanics, of rank 3, to a space of rank 1, by the way of a suitable quotient.*

Consider the dimensions of the quantities c, $\hbar$ and a unit of energy, say the electronvolt, eV*:* $\dim(c) = \mathsf{L}\mathsf{T}^{-1}$, $\dim(\hbar) = \mathsf{L}^2\mathsf{T}^{-1}\mathsf{M}$ *and* $\dim(\text{eV}) = \mathsf{L}^2\mathsf{T}^{-2}\mathsf{M}$. *They are independent in the group $\mathscr{D}_{\text{mech}}$, as can be seen by solving the equation $\dim(c)^{m_1}\dim(\hbar)^{m_2}\dim(\text{eV})^{m_3} = 1_{\mathscr{D}}$, whose unique solution is $m_1 = m_2 = m_3 = 0$. Moreover, they can generate the group since the equation $\dim(c)^{m_1}\dim(\hbar)^{m_2}\dim(\text{eV})^{m_3} = \mathsf{L}^a\mathsf{T}^b\mathsf{M}^c$ in the group is translated as the following system of linear equations in $\mathbb{Z}$:*

$$\begin{cases} m_1 + 2m_2 + 2m_3 &= a \\ -m_1 - m_2 - 2m_3 &= b \\ m_2 + m_3 &= c \end{cases} \tag{6}$$

The matrix of coefficients is unimodular, so it is invertible and the system has a unique solution for any values a, b, c in $\mathbb{Z}$. Therefore, we can write $\mathscr{D}_{\text{mech}} = \langle \dim(c), \dim(\hbar), \dim(\text{eV})\rangle$ and consider the subgroup $\mathscr{E} = \langle \dim(c), \dim(\hbar)\rangle$. Define a coherent section $\sigma : \mathscr{D}_{\text{mech}} \to Q_{\text{mech}}$ by its action on this basis:

$$\begin{aligned} &\sigma(\mathsf{L}\mathsf{T}^{-1}) = c, \\ &\sigma(\mathsf{L}^2\mathsf{T}^{-1}\mathsf{M}) = \hbar, \\ &\sigma(\mathsf{L}^2\mathsf{T}^{-2}\mathsf{M}) = \text{eV}. \end{aligned}$$

Now this section, when restricted to the subgroup $\mathscr{E}$, defines a subsection Σ. The quotient Q_{mech}/Σ has group of dimensions $\mathscr{D}_{\text{mech}}/\mathscr{E} \cong \langle \dim(\text{eV})\rangle$ and rank 1. The dimensionless quantities include $\hbar$ and c, which are identified with $1_{Q_{\text{mech}}}$, as desired, and the rest of quantities have dimensions of a power of energy. For instance, the dimension $\mathsf{L} = \dim(c)\dim(\hbar)\dim(\text{eV})^{-1}$, when carried to the quotient group, is $\hat{\rho}(\mathsf{L}) = \dim(\text{eV})^{-1}$, so lengths are measured as inverse energies.

4. HOMOMORPHISM OF SPACES OF QUANTITIES. ISOMORPHIC SPACES

In this section the tool for comparison of spaces of quantities is defined and its properties studied. The goal is the classification of spaces of quantities, which is achieved in theorem 2.

Definition 4.1. *Let Q and R be spaces of quantities over the field F. A map $\psi : Q \to R$ is a homomorphism of spaces of quantities if*

(i) for any two quantities q_1, q_2 in Q

$$\psi(q_1 q_2) = \psi(q_1)\psi(q_2),$$

that is, it is a monoid homomorphism with respect to the product, and

(ii) if q_1 and q_2 are quantities in the same fiber of Q, then $\psi(q_1)$ and $\psi(q_2)$ are in the same fiber in R and

$$\psi(\alpha q_1 + \beta q_2) = \alpha\psi(q_1) + \beta\psi(q_2),$$

for α and β in F, so ψ is a linear map in each fiber.

The homomorphism ψ induces a group homomorphism between the base groups, $\mathscr{D}_Q$ and $\mathscr{D}_R$. If $\dim_Q$ and $\dim_R$ are the respective projection maps, define the map $\phi : \mathscr{D}_Q \to \mathscr{D}_R$ so that the following diagram commutes,

$$\begin{array}{ccc} Q & \xrightarrow{\psi} & R \\ {\scriptstyle \dim_Q}\downarrow & & \downarrow{\scriptstyle \dim_R} \\ \mathscr{D}_Q & \xrightarrow{\phi} & \mathscr{D}_R \end{array}$$

that is, $\dim_R \circ \psi = \phi \circ \dim_Q$. It is well defined because $\dim_Q$ and $\dim_R$ are surjective and ψ preserves fibers and it is straightforward to check that ϕ is a group homomorphism.

The map ϕ says which fibers of Q are mapped into each fiber of R. As an example, if q is a quantity of dimension one in Q then $\dim_R(\psi(q)) = \phi(\dim_Q(q)) = \phi(1_{\mathscr{D}_Q}) = 1_{\mathscr{D}_R}$, so the fiber of quantities of dimension one in Q is mapped to the fiber of quantities of dimension one in R.

An isomorphism $Q \to R$ is a bijective homomorphism and defines Q and R as isomorphic spaces, denoted $Q \cong R$. The natural maps such as the identity map $\mathrm{id}_Q : Q \to Q$, the inclusion map $i : Q \hookrightarrow R$, where Q is a subspace of R, and the natural projection map $\rho : Q \to Q/\Sigma$, where Σ is a subsection of Q, are all homomorphisms of spaces of quantities. The trivial homomorphism is the map which sends every element in Q to the dimensionless zero of R. By a zero homomorphism we understand a homomorphism in which all the elements in Q are mapped to zero elements in R, such as the trivial map, but there are other zero homomorphisms, as many as group homomorphisms between $\mathscr{D}_Q$ and $\mathscr{D}_R$. For given such a group homomorphism ϕ, which in turn defines which fibers in Q are mapped to which fibers in R, it is enough to define $\psi : Q \to R$ by sending each $q \in Q$ to the zero element of the fiber assigned by ϕ.

In fact it is necessary to understand the behaviour of fibers and zeros under a homomorphism. It is clear that the image of a zero is a zero. If q is a nonzero element of Q but $\psi(q)$ is a zero in R, then all the fibers of q are mapped to the same zero, for $\psi(\alpha q) = \alpha\psi(q)$ which is the same zero for any $\alpha \in F$. On the other hand, if $\psi(q)$ is not a zero, then the fiber of q is mapped isomporphically (as vector spaces) to the fiber of $\psi(q)$. In particular, if $\psi(1_Q)$ is zero, then the homomorphism is a zero homomorphism, for $\psi(q) = \psi(q1_Q) = \psi(q)\psi(1_Q)$ which is a zero for any q. This expresion also proves that if $\psi(1_Q)$ is not a zero then it is 1_R, the identity in R.

So far a homomorphism can be defined by setting which fiber of R is the image of each fiber of Q and by setting which fibers of Q are mapped to zero and which of them are mapped isomorphically to their corresponding fibers. In the case of interest of spaces free of zero divisors the result can be improved.

Proposition 4.2. *Let Q be a space of quantities free of zero divisors and $\psi : Q \to R$ a homomorphism of spaces of quantities. Then $\psi(1_Q) = 0$ implies ψ is a zero homomorphism, while $\psi(1_Q) \neq 0$ implies that each fiber is mapped isomorphically onto a fiber in R.*

PROOF. The first part has already been proved. Assume now that $\psi(1_Q)$ is not a zero and let q be a nonzero element of Q which, thus, has an inverse q^{-1}. Since $\psi(1_Q) = \psi(qq^{-1}) = \psi(q)\psi(q^{-1})$ is not a zero, we conclude that $\psi(q)$ is not a zero. Since its fiber is a one dimensional vector space, the latter says that the fiber of q is mapped isomorphically to the fiber $\psi(q)$. □

Of course, only the nonzero homomorphisms are of interest for us to be able to compare spaces of quantities, so from now on we only consider this kind of homomorphisms. The following are basic properties of homomorphisms of spaces of quantities.

Proposition 4.3. *Let $\psi : Q \to R$ be a nonzero homomorphism of spaces of quantities. Then:*

(i) the image of a subspace of Q is a subspace of R,

(ii) the preimage of a subspace of R is a subspace of Q,

(iii) the preimage of a section of R is a section of Q and

(iv) the preimage of a subsection of R is a subsection of Q.

PROOF. Let S be a subspace of Q, which is characterized, by proposition 3.2, as $S = \dim_Q^{-1}(\mathscr{E})$ for a subgroup $\mathscr{E}$ of $\mathscr{D}_Q$. The projection of its image is $\dim_R(\psi(S)) = \phi(\dim_Q(S)) = \phi(\mathscr{E})$, which is a subgroup of $\mathscr{D}_R$. Since ψ is a nonzero homomorphism, every fiber in S is mapped onto a fiber in $\psi(S)$, so we conclude that $\psi(S)$ coincides with $\dim_R^{-1}(\phi(\mathscr{E}))$, so it is a subspace of R.

Consider now S to be a subspace of R. Its inverse image $\psi^{-1}(S)$ is made of the fibers which are mapped into S. But these fibers are given by $\dim_Q^{-1}(\phi^{-1}(\dim_R(S)))$. Since $\dim_R(S)$ is a subgroup of $\mathscr{D}_R$, so is $\phi^{-1}(\dim_R(S))$ with respect to $\mathscr{D}_Q$ and, thus, $\psi^{-1}(S)$ is a subspace of Q.

Let σ be a section of R. Since all fibers of R are represented in the section, it is clear that its inverse image, $\psi^{-1}(\sigma(\mathscr{D}_R))$, contains at least an element from each fiber in Q. We now show that there is no more than one from each fiber. Assume q_1 and q_2 are elements in the same fiber in Q with $\psi(q_1)$ and $\psi(q_2)$ in the section of R. Then $\psi(q_1)$ and $\psi(q_2)$ belong to the same fiber in R, which means $\psi(q_1) = \psi(q_2)$ because they are in a section. Therefore, since fibers in Q are mapped isomorphically to fibers in R, this leads to $q_1 = q_2$.

Finally, considering a subsection as the intersection of a subspace and a section in R it is clear that the inverse image of such intersection is a intersection of a subspace and a section in Q, thus, a subsection. □

The kernel of a nonzero homomorphism $\psi : Q \to R$ is defined as $\ker\psi = \psi^{-1}(1_R)$. Since $\{1_R\}$ is a subsection of R, its inverse image defines, by proposition 4.3, a subsection in Q. The image of a homomorphism, $\operatorname{im}\psi = \psi(Q)$ is, by the same proposition, a subspace of R. The kernel and the image so defined satisfy an isomorphism theorem.

Theorem 1. *Let $\psi : Q \to R$ be a nonzero homomorphism of spaces of quantities. Then*

$$Q/\ker\psi \cong \operatorname{im}\psi$$

as spaces of quantities.

PROOF. The first step is to check that the quotient of the theorem is indeed a space of quantities. Let us denote by K the kernel of ψ, which is a subsection of Q. Thus, we only have to show that the group of dimensions of Q/K is free Abelian and, to that end, we have to identify the projection of K on $\mathscr{D}_Q$. We claim this projection to be precisely the kernel of the induced group homomorphism: $\dim_Q(K) = \ker\phi$. Let q be in K. Then $\phi \circ \dim_Q(q) = \dim_R \circ \psi(q) = 1_{\mathscr{D}_R}$, so $\dim_Q(q) \in \ker\phi$ which shows one inclusion. Now let A be in $\ker\phi$, and let q be a nonzero element in the fiber $\dim_Q^{-1}(\mathsf{A})$. Then $1_{\mathscr{D}_R} = \phi(\mathsf{A}) = \dim_R \circ \psi(q)$, so $\psi(q)$ has dimension one and can be written as $\psi(q) = \alpha 1_R$ for a nonzero α in F (since ψ is a nonzero homomorphism). Consider the element $q_1 = \alpha^{-1}q$. Then $\psi(q_1) = 1_R$, so $q_1 \in K$ and $\dim_Q(q_1) = \dim_Q(q) = \mathsf{A}$, so $\mathsf{A} \in \dim_Q(K)$, which shows the other inclusion and the claim is proved. The group of dimensions of Q/K is, thus, $\mathscr{D}_R/\ker\phi$, which is isomorphic with $\operatorname{im}\phi$ by the isomorphism theorem for groups. Since $\operatorname{im}\phi$ is a subgroup of $\mathscr{D}_R$ it is free Abelian, and so is $\mathscr{D}_R/\ker\phi$. Therefore, Q/K is a space of quantities.

The rest of the proof is standard. Define the map $\hat{\psi} : Q/K \to \operatorname{im}\psi$ by $\hat{\psi}(qK) = \psi(q)$. It is straightforward to check, first, it is well defined; second, it is a homomorphism of spaces of quantities; third, it is a bijection. □

Example 4.4. *Let us revisit example 3.8 from the viewpoint of homomorphisms. Consider the map $\psi : Q_{\text{mech}} \to Q_{\text{energy}}$ given in the following form: the image of the quantity* eV *be itself, while the image of the speed of light, c, be $1_{Q_{\text{energy}}}$ as well as the image of the reduced Planck's constant, $\hbar$. Then $\ker\psi = \{c^n : n \in \mathbb{Z}\} \cup \{\hbar^n : n \in \mathbb{Z}\}$ and $\operatorname{im}\psi = Q_{\text{energy}}$. Theorem 1 says that $Q_{\text{mech}}/\ker\psi$ is isomorphic with Q_{energy}.*

The next result is the classification theorem for spaces of quantities free of zero divisors.

Theorem 2. *Two spaces of quantities over the same field, free of zero divisors, are isomorphic if and only if they have the same rank.*

PROOF. First consider two spaces of quantities $Q \cong R$. Then there is an isomorphism $\psi : Q \to R$ which induces a group homomorphism $\phi : \mathscr{D}_Q \to \mathscr{D}_R$. We only need to show the latter to be an isomorphism for it is a well known result of the theory of free Abelian groups that two such groups are isomorphic if and only if they have the same rank [26, Theorem 10.14]. But that is obvious since the map ϕ is nothing but the rule which says which fiber in Q is mapped to what fiber in R and, since ψ is an isomorphism, this mapping of fibers is a bijection.

Second, assume Q and R are two spaces of quantites of the same rank, that is, their groups of dimensions $\mathscr{D}_Q$ and $\mathscr{D}_R$ have the same rank. Therefore, there is a group isomorphism $\phi : \mathscr{D}_Q \to \mathscr{D}_R$ and it defines a bijection of the fibers in Q with the fibers in R. If we can assign a linear isomorphism between each pair of fibers, we are done. To that end it is enough to map a nonzero element of each fiber in Q with a nonzero element of its corresponding fiber in R. Now, since both Q and R are free of zero divisors, by proposition 2.8 each of them has a coherent system of units, say σ_Q and σ_R, respectively. Define a map $\psi : Q \to R$ by giving its action on the set $\sigma_Q(\mathscr{D}_Q)$ so that the following diagram commutes,

$$\begin{array}{ccc} Q & \xrightarrow{\psi} & R \\ {\scriptstyle\sigma_Q}\uparrow & & \uparrow{\scriptstyle\sigma_R} \\ \mathscr{D}_Q & \xrightarrow{\phi} & \mathscr{D}_R \end{array}$$

and extend it linearly in each fiber. This map is easily seen to be an isomorphism of spaces of quantities, so $Q \cong R$. □

As a last example, we show that, up to isomorphism, the example 2.9 is the only space of quantities over a group of dimensions and a field free of zero divisors.

Example 4.5. *Let Q be a space of quantities over the field F with group of dimensions $\mathscr{D}$ and free of zero divisors. Let σ be a coherent system of units and ν the map defined in equation (4). Then the map $\psi : Q \to F \times \mathscr{D}$ given by $\psi(q) = (\nu(q), \dim(q))$ is an isomorphism of spaces of quantities. Its inverse is $\psi^{-1}(\alpha, \mathsf{A}) = \alpha\,\sigma(\mathsf{A})$.*

This is to say that every space of quantities Q, free of zero divisors, is isomorphic with $F \times \mathscr{D}$. But the isomorphism is not canonical, for it depends on the system of units chosen.

5. Conclusions

A brand-new algebraic structure has been defined and studied which fits exactly the algebra of quantity calculus. This structure is centered in the concept of dimension of a quantity as its main property, and considers only integer exponents in the combination of dimensions (and, thus, of units).

The axioms of definition 2.5 permit to deduce the usual expresion of a quantity as the product of a number times a unit, and to deduce the way the operations are usually performed. In addition, it has been shown how to construct new spaces from old ones by means of subspaces, tensor products or quotients. The latter construction is seen to be the correct interpretation of the usual procedure in physics, misleadingly referred to as a change of units, which sets some universal constant to 1. Finally, a characterization and classification has been obtained in terms of the rank of the space of quantities, that is, the rank of its group of dimensions. This classification must be compared with that of Kitano [1], for they share some elements in common, but are not equivalent.

Kitano studies a set of physical quantities under different 'systems of units', even though these systems of units may have different number of basic units. In our approach, this should be considered as different spaces of quantities. Kitano then defines the relation $V \precsim U$ between the systems of units V and U if every pair of quantities q and p which are equal in U are also equal in V (equivalently, if every pair of distinguishable quantities in V are also distinguishable in U). This relation is a preorder in the set of systems of units. In case $V \precsim U$ and $U \precsim V$ then U and V are said to be equivalent. If the system U is supported by the algebraic structure $\mathbb{R} \times \mathbb{Q}^n$ and V by $\mathbb{R} \times \mathbb{Q}^m$, then $V \precsim U$ implies $m \leq n$, and U and V can be equivalent only if $m = n$. But the contrary is not necessarily true, for there are systems of units with $m = n$ which are not comparable. An example of the latter is given by the electrostatic and the electromagnetic systems of units (esu and emu respectively). In the esu system the quantities $\mu_0/4\pi$ and $1/c^2$, where μ_0 is the permeability of empty space and c is the speed of light, are not distinguisable, while $4\pi\varepsilon_0$ is distinguishable from any of the former, being ε_0 the permittivity of empty space. In the emu system, however, $4\pi\varepsilon_0$ is not distinguishable from $1/c^2$, but $\mu_0/4\pi$ is; therefore, these systems are not comparable. It must be stressed that Kitano's classification is based on physical grounds, not algebraic ones, and thus is different from the one introduced in the present paper as we see in the next paragraph.

Within the context of the algebraic fiber bundle introduced in this paper, a similar preorder can be defined. We can write $Q \precsim R$ for two spaces of quantities Q and R, over the same field F and free of zero divisors, if there is a homomorphism $Q \to R$ which is an injection (i.e. a monomorphism). The injection property ensures that different quantities in Q have different images in R. This is possible only if $\operatorname{rank} Q \leq \operatorname{rank} R$, as can be easily seen, and moreover, this condition is sufficient. Under this relation all spaces of quantities are comparable, and their comparison depends only on their ranks, in contrast with Kitano's classification. Resuming the previous example, the esu and emu systems can be understood as suitable quotients of the MKSA system. The MKSA system of units is defined in the space of quantities with group of dimensions $\mathscr{D} = \langle \mathsf{L}, \mathsf{T}, \mathsf{M}, \mathsf{I} \rangle$, of rank 4. In this space, the quantity $1/4\pi\varepsilon_0$ has dimension $\mathsf{L}^3\mathsf{T}^{-4}\mathsf{M}\mathsf{I}^{-2}$, while the dimension of the quantity $\mu_0/4\pi$ is $\mathsf{L}\mathsf{T}^{-2}\mathsf{M}\mathsf{I}^{-2}$. The esu system is obtained by making the quantity $1/4\pi\varepsilon_0$ equal to 1, that is, since the group $\mathscr{D}$ can be written as $\mathscr{D} = \langle \mathsf{L}, \mathsf{T}, \mathsf{I}, \mathsf{L}^3\mathsf{T}^{-4}\mathsf{M}\mathsf{I}^{-2} \rangle$, by taking the quotient of the space of quantities which gives the group of dimensions $\mathscr{D}/\mathscr{E}$, where $\mathscr{E} = \langle \mathsf{L}^3\mathsf{T}^{-4}\mathsf{M}\mathsf{I}^{-2} \rangle$, we thus, get a new space of quantities of rank 3. The emu system, on the other hand, is obtained by making $\mu_0/4\pi$ equals 1, that is, by the quotient in the group of dimensions $\mathscr{D}/\mathscr{F}$, where now $\mathscr{F} = \langle \mathsf{L}\mathsf{T}^{-2}\mathsf{M}\mathsf{I}^{-2} \rangle$, so it is also a new space of quantities of rank 3. According to theorem 2, these two spaces of quantities are isomorphic, regarded as algebraic structures, although Kitano's classification states them, righteously from a physical point of view, as incomparable. Nevertheless, it must be stressed that, under the algebraic theory depicted in this paper, the MKSA system and the esu (or emu) system are not different unit systems, but different spaces of quantities.

Finally, a comment on Domotor's torsor theory, which explains, in a series of three papers [28, 29, 4], in a convenient way, the passage from the state space of a physical measurement device to the space of quantities. His description of torsors is well suited to the algebraic structure of quantities given by Drobot and subsequent authors up to Kitano, but the same ideas may be adapted to the structure presented in this paper so as to provide the same foundation of the aforementioned passage. Indeed, the tool of torsors explains exactly, as he shows in his paper, the structure of the fibers, each of which is a vector space of dimension one.

6. Acknowledgement

Thanks are due to the anonymous reviewer whose insightful and accurate comments have improved this paper.

References

[1] Kitano, M. (2013). Mathematical structure of unit systems. *Journal of Mathematical Physics*, 54, 052901.

[2] Krystek, M. (2015). The term 'dimension' in the international system of units. *Metrologia*, 52, 297–300.

[3] Atkey, R., Ghani, N., Forsberg, F. N., Revell, T., Staton, S. (2015). Models for polymorphism over physical dimensions. In *13th International Conference on Typed Lambda Calculi and Applications (TLCA'15)*, 999–1013.

[4] Domotor, Z. (2017). Torsor theory of physical quantities and their measurement. *Measurement Science Review*, 17 (4), 152–177.

[5] Mills, I. (2016). On the units radian and cycle for the quantity plane angle. *Metrologia*, 53, 991–997.

[6] Mari, L., Giordani, A. (2012). Quantity and quantity value. *Metrologia*, 49, 756–764.

[7] Feller, U. (2011). The International System of Units–a case for reconsideration. *Accreditation and Quality Assurance*, 16, 143–153.

[8] Foster, M. (2010). The next 50 years of the SI: A review of the opportunities for the e-Science age. *Metrologia*, 47, R41–R51.

[9] Johansson, I. (2010). Metrological thinking needs the notions of parametric quantities, units and dimensions. *Metrologia*, 47, 219–230.

[10] Barrow, J.D., Gibbons, G.W. (2017). Maximum magnetic moment to angular momentum conjecture. *Physical Review D* ,95, 064040.

[11] Kostro, L. (2010). The physical meaning of the coefficients c^n/G ($n = 1,2,\ldots 5$) and the Standard Model of the Universe. *AIP Conference Proceedings*, 1316 (1), 165–179.

[12] Joint Committee for Guides in Metrology. (2012). *International vocabulary of metrology – Basic and general concepts and associated terms (VIM)*. 3rd edition. JCGM 200:2012.

[13] de Boer, J. (1994). On the history of quantity calculus and the international system. *Metrologia*, 31, 405–429.

[14] Fourier, J.B.J. (1822). *Théorie analytique de la Chaleur*. Paris: Chez Firmin Didot, père et fils.

[15] Hallock, W., Wade, H.T. (1906). *Outlines of the Evolution of Weights and Measures and the Metric System*. The Macmillan Company.

[16] Maxwell, J. (1873). *Treatise on Electricity and Magnetism*. Oxford University Press.

[17] Buckingham, E. (1914). On physically similar systems: Illustrations of the use of dimensional equations. *Physical Review*, 4, 345–376.

[18] Landolt, M. (1945). *Größe, Maßzahl und Einheit*. Rascher Verlag Zurich. (French Translation – Paris: Dunod, 1947)

[19] Fleischmann, R. (1959/60). Einheiteninvariante Größengleichungen, Dimensionen. *Der Mathematische und Naturwissenschaftliche Unterricht*, 12, 385–443.

[20] Quade, W. (1961). Über die algebraische Struktur des Größenkalküls der Physik. *Abhandlungen der Braunschweigischen Wissenschaftlichen Gesellschaft*, 13, 24–65. (Translated as an appendix to Dimensional Analysis for Economists by F.J. de Jong, Amsterdam: North Holland Publishing Company, 1967)

[21] Drobot, S. (1953). On the foundations of dimensional analysis. *Studia Mathematica*, 14, 84–99.

[22] Whitney, H. (1968). The mathematics of physical quantities: Part II: Quantity structures and dimensional analysis. *The American Mathematical Monthly*, 75 (3), 227–256.

[23] Carlson, D. (1979). A mathematical theory of physical units, dimensions and measures. *Archive for Rational Mechanics and Analysis*, 70, 289–304.

[24] Giorgi, G. (1904). Proposals concerning electrical and physical units. In *Transactions of the International Electrical Congress: St. Louis, 1904*. Albany, N.Y.: J.B. Lyon Company, 136–141.

[25] Bridgman, P. (1922). *Dimensional Analysis*. Yale University Press.

[26] Rotman, J. (1995). *An Introduction to the Theory of Groups*, 4th edition. Springer.

[27] Lainscsek, C., Gorodnitsky, I. (2003). Planck's natural units are contained in a time series. *AIP Conference Proceedings*, 676 (1), 370–371.

[28] Domotor, Z. (2012). Algebraic frameworks for measurement in the natural sciencies. *Measurement Science Review*, 12 (6), 213–233.

[29] Domotor, Z. (2016). An algebraic approach to unital quantities and their measurement. *Measurement Science Review*, 16 (3), 103–126.

Received January 19, 2018.
Accepted July 15, 2018.

MEASUREMENT SCIENCE REVIEW, **18**, (2018), No. 4, 158–161

sciendo

ISSN 1335-8871

MEASUREMENT SCIENCE REVIEW

Journal homepage: http://www.degruyter.com/view/j/msr

A New Line-Shape Asymmetry Model for Wavelength Modulation Spectroscopy in Gaseous Flows

Václav Nevrlý[1], Vít Klečka[1], Michal Vašinek[2], Václav Válek[1], Jan Suchánek[3], Michal Dostál[1,3]
Branislav Hric[1], Petr Bitala[1] and Zdeněk Zelinger[3]

[1]*Faculty of Safety Engineering, VŠB-Technical University of Ostrava, Lumírova, 13, 700 30, Ostrava-Výškovice, Czech Republic, vaclav.nevrly@vsb.cz*
[2]*Faculty of Electrical Engineering and Computer Science , VŠB-Technical University of Ostrava, 17. listopadu, 15, 708 33, Ostrava-Poruba, Czech Republic*
[3]*J. Heyrovský Institute of Physical Chemistry, The Czech Academy of Sciences, Dolejškova, 2155/3, 182 23, Prague, Czech Republic*

This communication reports technical notes on the development and application of an automated line-shape fitting procedure for wavelength modulation spectroscopy (WMS). Near-infrared transitions of carbon dioxide (CO_2) around 1573 nm were measured in vertical cold (non-reacting) flow of CO_2 at atmospheric pressure using WMS with demodulation at second harmonic frequency. Semi-empirical model based on the set of so-called Gabor functions was developed and parameters of Lorentzian line-shape profile and its asymmetry resulting from simultaneous frequency and amplitude response of the current-modulated semiconductor laser were determined. Nonlinear least-square fitting procedure employing differential evolution algorithm was successfully utilized for performing this task. Line-shape fitting procedure enabling efficient signal de-noising and background subtraction of wavelength modulation spectra was implemented into an open-source code.

Keywords: gas sensing, spectroscopy, modulation, model, line-shape asymmetry.

1. Introduction

Tunable Diode Laser Absorption Spectroscopy (TDLAS) has long been recognized as a well-established optical diagnostic tool for gas sensing in various environments (see e.g., [1] and references therein). In order to reduce low-frequency noise inherent to direct TDLAS method, wavelength modulation spectroscopy (WMS) [2] is usually preferred for measurement of low absorption signals.

For analytical purposes it is always desirable to evaluate line-shape profile [3, 4, 5, 6], rather then simply estimate peak-to-peak parameters of WMS signal, which contains less-robust information on molecular absorption and spectral broadening. Therefore, line-shape fitting procedure needs to be addressed in WMS post-processing scheme as the level of uncertainties is a critical attribute of measurement outputs.

Efficient signal de-noising and background subtraction is a serious problem when detecting low-level absorption signals. Spectral coincidence of different absorption features with molecular fingerprints of interest is another issue to be solved for successful interpretation of WMS experiments.

Here we report an automated procedure which was developed in order to address these problems without the need for prior spectral calibration of a particular WMS setup. The solution is based on a novel analytical model of spectral line-shapes relevant to WMS signals from absorbing molecules, which may be present on an optical path of the respective laser beam.

2. Subject & Methods

In a typical WMS setup, signal from the optical detector is demodulated via analog or digital lock-in amplifier at the n-th (which is often the second, i.e. $2f$) harmonic frequency of a current-modulated semiconductor laser radiation. The simplest experimental configuration for $2f$–WMS consists of a single optical path, which can include both probed volume (with an analyte of interest) and a reference cell or gaseous flow suitable for spectral calibration purposes.

The procedure, which is schematically described in Fig. 1. was developed as a first step of our aim to perform quantitative spectroscopy and species concentration measurements in the given experimental setup. Gaseous flow of carbon dioxide (CO_2) at room temperature was utilized as a feasible reference for spectral calibration. Lorentzian spectral line profile is therefore assumed to be appropriate for the given study.

DOI: 10.1515/msr-2018-0022

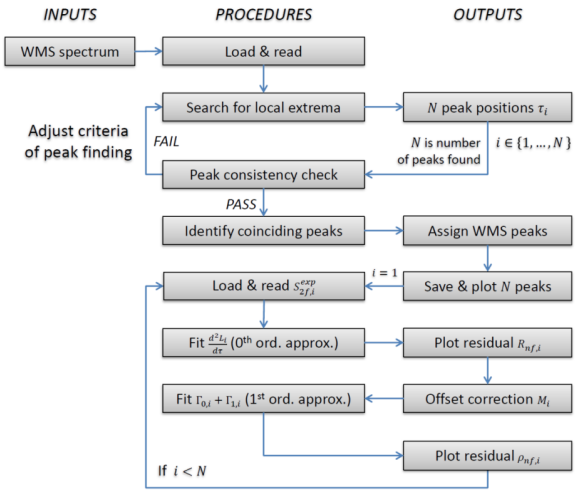


Fig. 1. Procedural steps involved in fitting of new line-shape asymmetry model.

Due to this assumption, applicability of the method reported here is rather limited to experimental conditions when collisional broadening dominates over the Doppler broadening and other physical effects on spectral line shape.

Spectral line shapes were investigated primarily because they contain crucial information for development of proper data processing methodology as well as for optimizing the overall performance of our experimental setup. Theoretical model which we used for the given purpose enabled to efficiently describe the characteristic profile of $2f$–WMS absorption signal.

Compared with more sophisticated line-shape fittings methods, no further data on experimental setup (modulation intensity, optical power change, etc.) or parameters of spectral line position and width were required as model inputs. Thus, we anticipate that the individual harmonic components of $2f$–WMS signal can be estimated prior to spectral calibration and determination of wavelength modulation response function for the given semiconductor laser.

Finally, an automated line-shape fitting procedure has been implemented in the Python programming language and its open-source release in the form of interactive script is in preparation.

2.1. Experimental setup

Measurements reported here were performed under atmospheric pressure using the body of the burner designed according to [7]. Two separate mass-flow controllers (Manufacturer: Bronkhorst High-Tech B.V.) were employed for feeding the dryed air from a compressor (Manufacturer: JUN–AIR/Gast Group Ltd.) into the central body of the burner and carbon dioxide (Manufacturer: Air Liquide Deutschland GmbH, purity 99.995%) to annular co-flow at volumetric flow rates of $Q_{v,air} = 10$ l/min and $Q_{v,CO_2} = 11$ l/min, respectively.

Tuning range of distributed feedback (DFB) laser operating around 1.573 μm (Manufacturer: Eblana Photonics Ltd.) was periodically (each 10 s) scanned by ramping up electric current at constant diode temperature maintained by laser driver/controller (Manufacturer: Thorlabs Inc.).

Sine wave ($f = 11$ kHz) was superimposed electronically on repeating saw-tooth wave to modulate the lasing wavelength. Signal from the amplified InGaAs photodetector (Manufacturer: Thorlabs Inc.) was demodulated at second harmonic ($2f$) frequency by analog lock-in amplifier (Manufacturer: Stanford Research Systems Inc.) and acquired on digital oscilloscope (Manufacturer: Teledyne LeCroy GmbH) at the sampling rate of 1 kS/s.

Finally, accumulation and averaging procedure (each 5 samples) was performed to increase the signal-to-noise ratio (SNR), thus single measurement (averaged scan) was obtained during a 50 s interval. Optical path length of the laser beam through absorbing medium (CO_2 coflow stream) was increased by one reflection on the planar mirror, thus yielding $l_p = 10$ cm.

2.2. Theoretical model

Spectrally-broadened absorption line shape in an atmospheric-pressure gaseous flow can be approximated by area-normalized Lorentzian profile given by the function:

$$L = \frac{a_0}{\pi a_2 \left(1 + \left(\frac{\tau - a_1}{a_2}\right)^2\right)}, \tag{1}$$

where a_0, a_1, and a_2 are height (absolute maximum), center and the half-width at half maximum (HWHM) of the Lorentzian function L, respectively. Spectral profile is characterized here in temporal domain (by τ ranging from $\tau = 0$ s to $\tau_S = 10$ s in our specific case) which is proportional to ramping laser current.

Second derivative of L provides zeroth-order approximation of the $2f$–WMS signal S_{2f} in case of pure frequency modulation (FM) leading to fully symmetric spectral line shape. Analytical form obtained after symbolic derivation and simplification is given by:

$$\frac{d^2L}{d\tau} = -\frac{2a_0a_2\left(a_2^2 - 3\left(a_1 - \tau\right)^2\right)}{\pi\left(a_2^2 + \left(a_1 - \tau\right)^2\right)^3}. \tag{2}$$

Fourier series expansion has been previously employed [2] to derive $2f$–WMS analytical expression of line-shape function and its asymmetry in a frequency domain.

However, unlike previous investigators [3, 4, 6], we report here an alternative use of trigonometric series postulated by Gabor [8] for representation of an arbitrary elementary signal.

Following notation given in (1), Gabor functions were reformulated into general form (for $k = 0, 1, 2, \ldots, \infty$):

$$\Gamma_{k,\cos} = \alpha_k \cos\left(2\pi k \frac{(\tau - a_1)}{2a_2} + \phi_k\right) G_w, \tag{3}$$

$$\Gamma_{k,\sin} = \beta_k \sin\left(2\pi (k + \tfrac{1}{2}) \frac{(\tau - a_1)}{2a_2} + \psi_k\right) G_w, \tag{4}$$

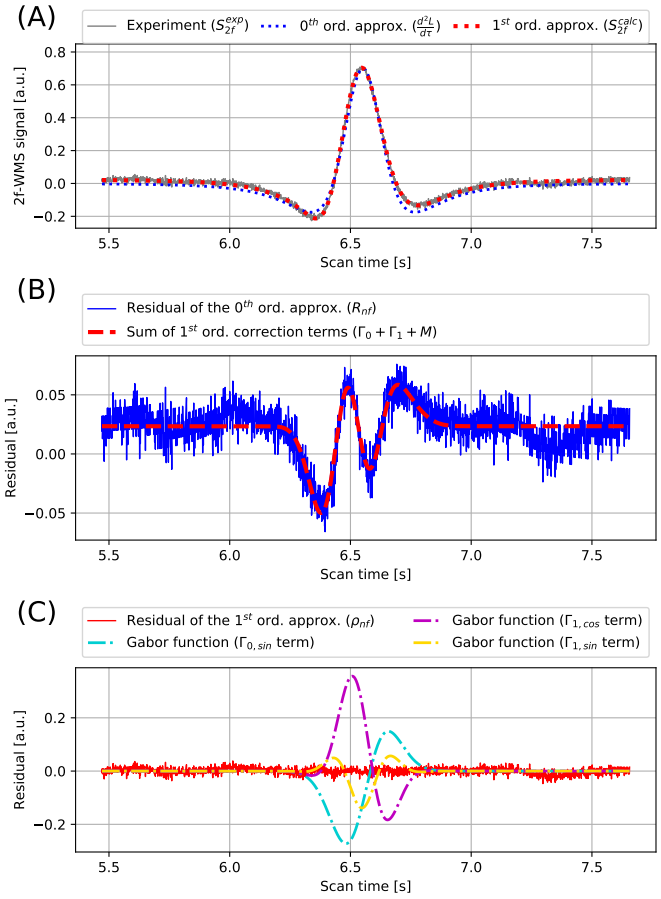


Fig. 2. Spectral profile of R(16) absorption line of carbon dioxide ($2\nu_1+2\nu_2+1\nu_3$ band) in temporal domain and results of line-shape fitting procedure. Experimental data recorded at lock-in phase shift $\Phi = -45°$ (with 30 ms integration time, 200 mV sensitivity and 100× signal expansion).

$$\sum_{k=0}^{\infty}\Gamma_k = \sum_{k=0}^{\infty}\left(\Gamma_{k,\cos}+\Gamma_{k,\sin}\right). \tag{5}$$

Here α_k, β_k, ϕ_k and ψ_k are amplitudes and phase shifts, respectively, of the corresponding term (Gabor function) and G_w is a width-adjusted (reduced) Gaussian distribution function described below in more details, see (6).

Only few terms of this expansion (for $k=0$ and $k=1$) were considered as essential for our application. To further limit the number of free parameters for least-square fitting procedure, $\Gamma_{0,cos}$ term was constrained to zero (by assuming $\phi_0 = \frac{\pi}{2}$). In spite of the fact, that $2f$–WMS is in principle zero background technique, significant offset value was observed and thus had to be included into our model following experimental trials. The offset value M was determined as an arithmetic mean of the signal in an appropriate section of the spectrum.

$$G_w = \frac{a_0}{w\sigma_i\sqrt{2\pi}}\exp\left[-\left(\frac{\tau-a_1}{2w\sigma_i}\right)^2\right], \tag{6}$$

where parameters a_0 and a_1 are height (absolute maximum) and center of the respective Lorentzian function specified according to (1). Initial value of standard deviation σ_i was determined from a_2 which is proportional to $\sigma_i\sqrt{2\ln(2)}$. Width-reduction parameter w of the G_w envelope was adjusted in order to obtain best-fit representation of the residual signal ($R_{nf} = S_{2f}^{exp} - \frac{d^2L}{d\tau}$) by minimal set of Gabor functions. Effective parameters of three Gabor functions ($\Gamma_{0,\sin}$ and both terms of Γ_1) were evaluated numerically by nonlinear least-square regression (minimization) method yielding:

$$\rho_{nf} = R_{nf} - M - \Gamma_0 - \Gamma_1 \approx 0. \tag{7}$$

First-order approximation of the $2f$–WMS signal was thus obtained as:

$$S_{2f}^{exp} \approx S_{2f}^{calc} = \frac{d^2L}{d\tau} + \Gamma_0 + \Gamma_1 + M. \tag{8}$$

3. Results

Measurements were mainly focused on the region around 6360 ± 2 cm^{-1} where we observed three spectral lines assigned as R(14), R(16) and R(18) of the $2\nu_1+2\nu_2+1\nu_3$ combination band of CO_2 [9]. Experimental data and final results of the theoretical model are depicted in Fig. 2, trace (A).

Apparently, sum of correction terms $(M+\Gamma_0+\Gamma_1)$ provides suitable regression function for fitting the R_{nf} residual with an excellent performance in the given case, see trace (B) in Fig. 2. It is worth noting that the width of Gaussian envelope relative to initially estimated HWHM (a_2) of Lorentzian function had to be reduced (to $w \approx 0.5$) in order to achieve appropriate best-fit representation.

Physical interpretation of Gabor function can then be anticipated from the trace (C) in Fig. 2. As n is odd for line-center asymmetric components of WMS signal, $\Gamma_{0,sin}$ and $\Gamma_{1,cos}$ terms are attributable to effect of $1f$ and $3f$ modulation, respectively. In analogy, $\Gamma_{1,sin}$ term resembles contributions from even n (e.g., $4f$) harmonics.

4. Conclusions/Discussion

Based on results summarized in the previous section we presume that the model reproduces some intrinsic features of WMS signal and provides an interesting alternative to non-physical methods for signal de-noising, e.g., discrete wavelet transformation (DWT) [10]. This new model can also provide initial inputs for physically sound theoretical models, enabling to estimate spectral response and phase-shift parameters (e.g., [3, 6]) from experimental line-shape profiles.

We can conclude that the entire procedure reported in this communication enables to automatically derive analytical description of line-shape asymmetry (i.e., instrumental function) corresponding to the specific experimental setup. Therefore, it has a capability to provide important inputs for simulation of complex absorption spectra when dealing with quantification of concentration or temperature fields in non-homogeneous gaseous flows (e.g., in laminar flames) based on the $2f$–WMS technique.

An open-source software implementation of an automated line-shape fitting procedure is convenient for further development or modifications in frame of various researches as well as educational activities. Extension of the model appropriate for fitting the spectral lines with Voigt profile is in progress [11].

5. Acknowledgement

The article is based upon work from COST Action (CM1404), supported by COST (European Cooperation in Science and Technology). The research was funded by the Ministry of Education, Youth and Sport of the Czech Republic (Project No. LTC17071). Financial support from the Czech Science Foundation (Project No. 17-05167S) and from VSB-Technical University of Ostrava (Faculty of Safety Engineering) via the Project No. SP2017/127 is also acknowledged.

References

[1] Hodgkinson, J., Tatam, R. P. (2013). Optical gas sensing: A review. *Measurement Science and Technology,* 24(1), 012004.

[2] Kluczynski, P., Gustafsson, J., Lindberg, A. M., Axner, O. (2001). Wavelength modulation absorption spectrometry — an extensive scrutiny of the generation of signals. *Spectrochimica Acta Part B: Atomic Spectroscopy,* 56(8), 1277–1354.

[3] Schilt, S., Thevenaz, L., Robert, P. (2003). Wavelength Modulation Spectroscopy: Combined Frequency and Intensity Laser Modulation. *Applied Optics,* 42(33), 6728.

[4] Engelbrecht, R. (2004). A compact NIR fiber-optic diode laser spectrometer for CO and CO2:. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy,* 60(14), 3291–3298.

[5] Civiš, S., Zelinger, Z., Nevrlý, V., Dorogan, A., Ferus, M., Iakovlev, V., Sirbu, A., Mereuta, A., Caliman, A., Suruceanu, G., Kapon, E. (2014). Near-infrared wafer-fused vertical-cavity surface-emitting lasers for HF detection. *Journal of Quantitative Spectroscopy and Radiative Transfer,* 147, 53–59.

[6] Zhao, G., Tan, W., Hou, J., Qiu, X., Ma, W., Li, Z., Dong, L., Zhang, L., Yin, W., Xiao, L., Axner, O., Jia, S. (2016). Calibration-free wavelength-modulation spectroscopy based on a swiftly determined wavelength-modulation frequency response function of a DFB laser. *Optics Express,* 24(2), 1723.

[7] Hartung, G., Hult, J., Kaminski, C. F. (2006). A flat flame burner for the calibration of laser thermometry techniques. *Measurement Science and Technology,* 17(9), 2485–2493.

[8] Gabor, D. (1946). Theory of communication. Part 1: The analysis of information. *Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering,* 93(26), 429–441.

[9] Henningsen, J., Simonsen, H. (2000). The (2201–0000) Band of CO2 at 6348 cm-1: Linestrengths, Broadening Parameters, and Pressure Shifts. *Journal of Molecular Spectroscopy,* 203(1), 16–27.

[10] Li, J., Parchatka, U., Fischer, H. (2012). Applications of wavelet transform to quantum cascade laser spectrometer for atmospheric trace gas measurements. *Applied Physics B,* 108(4), 951–963.

[11] Nevrlý, V., Bitala, P., Klečka, V., Vašinek, M., Zelinger, Z., Suchanek, J., Dostál, M., Válek, V. (2019). Analysis of wavelength modulation spectra for determination of OH radical concentration in an atmospheric pressure laminar premixed flames. *Journal of Physics: Conference Series*. Manuscript in preparation.

Received January 19, 2018.
Accepted July 25, 2018.

MEASUREMENT SCIENCE REVIEW, **18**, (2018), No. 4, 162-167

sciendo

ISSN 1335-8871

MEASUREMENT SCIENCE REVIEW

Journal homepage: http://www.degruyter.com/view/j/msr

Frequency and Time Fault Diagnosis Methods of Power Transformers

Miroslav Gutten[1], Daniel Korenciak[1], Matej Kucera[1], Richard Janura[2], Adam Glowacz[3], Eliasz Kantoch[4]

[1] *University of Žilina, Faculty of Electrical Engineering, Department of Measurement and Applied Electrical Engineering, Univerzitná 1, 010 26 Žilina, Slovak Republic, miroslav.gutten@fel.uniza.sk, daniel.korenciak@fel.uniza.sk, matej.kucera@fel.uniza.sk*

[2] *ČEZ, a. s., Nuclear Power Plant Temelin, 373 05 Temelin CZ, Czech Republic, richard.janura@cez.cz*

[3] *AGH University of Science and Technology, Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, Department of Automatic Control and Robotics, Al. A. Mickiewicza 30, 30-059 Kraków, Poland, adglow@agh.edu.pl*

[4] *AGH University of Science and Technology, Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, Department of Biocybernetics and Biomedical Engineering, Al. A. Mickiewicza 30, 30-059 Kraków, Poland, kantoch@agh.edu.pl*

The authors describe experimental and theoretical analyses of faults of power transformer winding. Faults were caused by mechanical effect of short-circuit currents. Measurements of transformer were carried out in high-voltage laboratory. Frequency and time diagnostic methods (method SFRA - Sweep Frequency Response Analysis, impact test) were used for the analyses. Coils of transformer windings were diagnosed by means of the SFRA method and the time impact test. The analyzed methods had a significant sensitivity to a relatively small deformation of coil. In the analysis a new technique for analyzing the effects of short-circuit currents is introduced. This technique is developed for high-voltage transformers (different types of power). The proposed analyses show that it is necessary to analyze the value of short-circuit current. Short-circuit current represents a danger for the operation of the power transformer. The proposed approach can be used for other types of transformers. Moreover, the presented techniques have a potential application for fault diagnosis of electrical equipment such as: transformers and electrical machines.

Keywords: Fault, short-circuit, transformer, diagnostics, frequency.

1. INTRODUCTION

Maintenance diagnostics of short-circuit currents of transformers should be carried out to increase the reliability in real trouble-free process. The short-circuits in operation are commonly caused by different faults such as: mechanical damage of insulation, electric insulation breakdown (caused by voltage), over heating fault, or open circuit fault [1].

The short-circuit is a serious fault for the transformer. High electric currents can damage the device easily. Electric currents cause high temperature of windings therefore their isolation is damaged. However, high electro-magnetic forces are much more dangerous. Electro-magnetic forces can damage the transformer permanently.

It is necessary to pay attention to higher reliability of power transformers, due to possible failures of an electric system and because of their cost.

Windings of the power transformer should be designed to avoid mechanical and thermal faults caused by short-circuit currents. Permanent deformations and gradual aging process of the electrical device are often caused by short-circuit current. Gradual aging process can worsen its mechanical properties. Heat shocks often cause decrease of mechanical properties of the transformer. Next it can cause unexpected damage of transformer during its operation.

Vibration based fault diagnosis of transformer winding was presented in the literature [2]. Fault diagnosis of power transformers using dissolved gas analysis was presented in the literature [3]-[7]. Electrical parameters of transformers were analyzed [8]-[10]. Signal processing and classification methods were also used for proper fault diagnosis of transformers [11]-[14] and other types of machines [15]-[18].

DOI: 10.1515/msr-2018-0023

To prevent the faulty state of transformers, the authors performed different types of diagnostic measurements. The measurements should illustrate a current state of the measured power transformer. It is essential to select a proper diagnostic method for the recognition of faulty state.

In the analysis a new technique for analyzing the effects of short-circuit currents is introduced. This technique is developed for high-voltage transformers (different types of power).

The authors focused on short-circuit of the transformer windings. The effect of magnetic field on conductor current causes the creation of forces. Mentioned forces have effect on coils of windings.

The electric currents of the transformer do not exceed rated values for normal operation. The electromagnetic forces effecting on transformer winding are low. The currents of short-circuits can reach the multiple of rated values of the transformer.

Then the forces can become dangerous for transformer coils of windings, electric and magnetic circuits, taps and the whole construction.

The short-circuit forces affecting on windings can be divided into two groups [19]:

- radial (cross),
- axial (longitude).

Radial forces are caused by electromagnetic lengthwise fields. They act in the radial direction. Radial forces are generated by leakage flux density and the current. They are paralleled with the axis of transformer winding. Radial forces extend external high-voltage windings. They also compress internal low-voltage windings. For this reason, the air interspaces are bigger.

Axial forces act in the axial direction. Similar to radial forces, axial forces are generated by leakage flux density and the current. They are dangerous from the center to border of coil. In this area of transformer the magnetic field has the high value in up and down edge of winding. Axial forces can reach dangerous values of short-circuits forces. They can also deform the outer high-voltage coil [19].

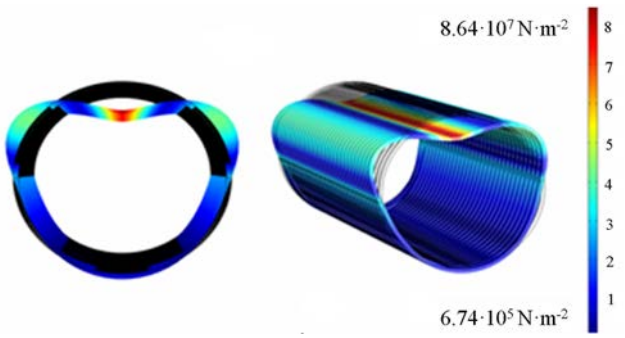


Fig.1. Simulation deformation of coil of the high-voltage winding (22 kV).

According to the literature [19] it is required to pay more attention to catching the outer high-voltage coil of winding. In case of a loose coil, the short-circuit axial forces Fd result in a vertical shift of outer coils of winding. The excessive pressure on spacers can press on the coil insulation and shift the winding. It can also cause serious damage to the transformer windings.

Fig.1. shows the simulation deformation of coil of the high-voltage winding (22 kV). It is influenced by radial forces (permanent shift of winding).

Fig.2. shows the simulation of pressure stress and permanent deformation of outer coil of winding. It was caused by short-circuit of the transformer. Axial forces are also proportional to the maximum current of the short-circuit. Even with a little imbalance between the windings, this force can be increased.

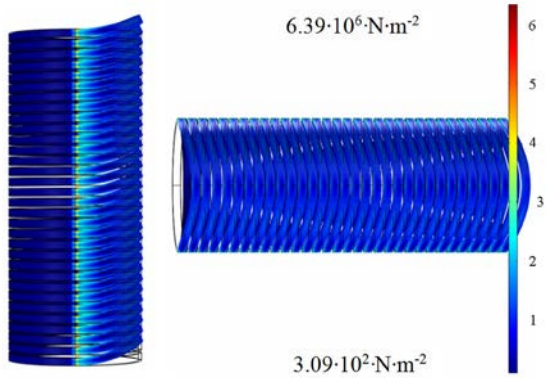


Fig.2. Permanent deformation of outer coil of winding. It was influenced by pressure stress from axial force (5 % per unit imbalance of the windings).

Based on theoretical and simulating analysis it can be noticed that the analysis of dangerous short-circuit forces is essential. It is necessary to determine the possible effects of short-circuit. It can also change the insulating state and the mechanical state of transformer.

2. EXPERIMENTAL DIAGNOSTIC METHODS OF TRANSFORMERS

The main problem of current energy transmission and distribution is that the measured data from the diagnostic measurements are not analyzed so much. There are problems to find scientific databases of such signals. Therefore, it is problem to assess the technical state of power transformers depending on their residual lifetime. A fault can occur in an unpredictable time of operation. The result of this can be the power failure for a short or a long time. It is needed to analyze the measured and computed values of parameters. It is based on knowledge of exposure to adverse effects of specific electrical equipment.

Achieving objectives by using the measurements can help to explain the adverse effects of short-circuit and propose new diagnostic techniques. Moreover, it is possible to detect an approaching fault of the electric device. Some steps may be also proposed in advance (e.g. repair of transformer) [20], [21].

The authors consider the influence of overvoltage, switching currents, short-circuit forces. Following measurements can be carried out for analyzed power transformers:

- measurements of transformer impedance depending on the frequency characteristics. It can be carried out using the SFRA method,
- time measurements of transformer coils. It can be carried out using impact test - high-voltage pulse source,
- measurements of impedance, resistivity of power transformer windings,
- analysis of the transformer oil,
- measurements of electric currents, voltages, insulation resistances using methods: RVM - Relevance Vector Machine, PDC - Polarization and Depolarization Current,
- measurements of loss factor and capacity,
- measurements of isolation parameters of the power transformer with application of dielectric spectroscopy,
- measurements of temperature of the transformer winding using FBG - Fiber Bragg Grating [22],
- selected variant of measuring methods.

Different measuring methods and diagnostic techniques can be used for fault diagnosis of transformers such as: analysis of acoustic emission, thermal images, electromagnetic radiation, analysis of temperature and moisture of transformers. The mentioned techniques can be used for detection of the effects of overvoltages and short-circuits. Effects of overvoltages and short-circuits can damage coils of the transformer. Repair of power transformer costs time and money. Mentioned techniques can be applied for disconnected power transformer [23].

Fig.3. Analysed power transformer with defect of the coil of phase A.

3. Experimental analysis of the power transformer 22 / 0.4 kV

On the basis of theoretical analyses of effects of short circuit, there were selected individual diagnostic and measurement methods for measurements on a power transformer 22 kV / 0.4 kV with defect on the high-voltage phase – coil A (Fig.3.).

Important task was to compare measured curves of single transformer winding and the sensitivity of the two diagnostic methods – high-frequency method SFRA (Sweep Frequency Response Analysis) and the method of the time analysis of high voltage impulse by impact test. Frequency and time method (SFRA and impact test) can determine the frequency or time response of characteristic quantities of affected by impedance and phase of transformer.

There is no need of changing of construction of the measured machine for both diagnostic techniques. They are also analyzed for disconnected transformer [24].

4. Measurement of the transformer using frequency method

The power transformer diagnostics set up uses the frequency scale from 20 Hz to 10 MHz at source voltage 0.2 – 20 Vpp, and output impedance 50 Ω using the Megger FRAX 150, (Fig.4.).

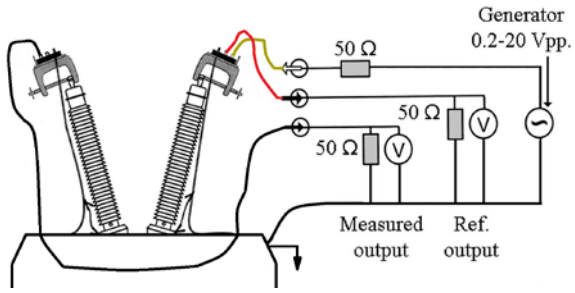


Fig.4. SFRA measurement connection using the Megger FRAX.

Mentioned analyzer can be used for magnetic core defects and detecting winding shifts of power transformers. The parameters of device are following: sampling rate 100 MS/s, absolute accuracy ± 0.5 dB for dynamic range -130 to +20 dB. The analyses were carried out for no-load and short-circuit states.

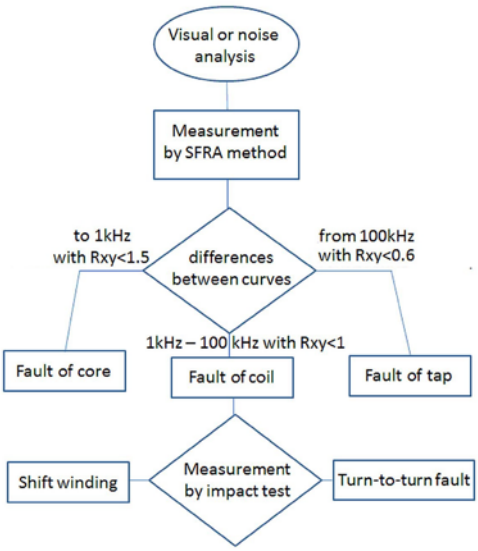


Fig.5. Block diagram of measurements of power transformer.

For low frequencies (up to 1 kHz) measured curves were good visible. For frequencies (1 kHz to 100 kHz) problems related to shift of windings turn-to-turn fault occurred [25].

For higher frequencies (>100 kHz) the faults related to short-circuit forces may appear.

The process of fault analysis on the transformer by measurement of the SFRA method with impact test is shown in Fig.5. Differences between measured frequency curves are analyzed by value of relative factor R_{xy}. This analysis is based on calculation of normalization factor covariance using equations (1), (2) and (3).

The short-circuit test detected winding state of the transformer. It showed deformation of internal winding. It also showed its shift as an activity of short-circuit. The impedance depending on frequency for short-circuit and no-load states was presented in Fig.6.

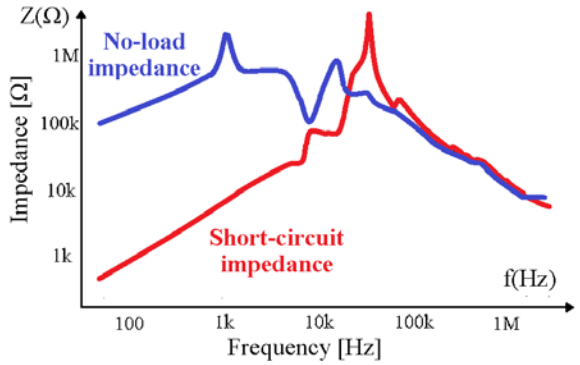


Fig.6. Impedance depending on frequency for short-circuit and no-load states.

Impedance depending on frequency of three windings of the power transformer of no-load test is depicted in Fig.7. In the Fig.7. we can see deformation of the coil A. This deformation occurred in the range of 1-10 kHz. Different values of B-C impedance mean damage of coil A for star connected transformer.

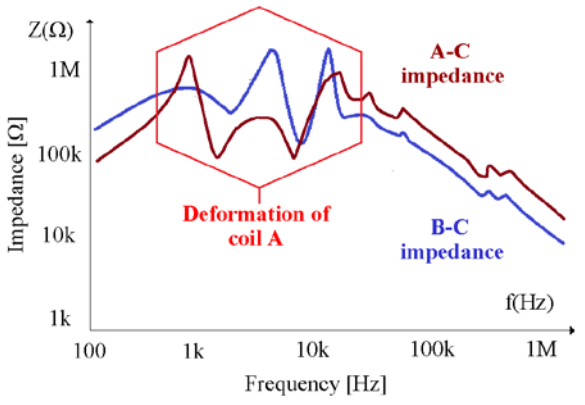


Fig.7. Impedance depending on frequency of the power transformer of no-load state.

Relative factor R_{xy} was used for analyses of differences between analyzed states.

For the analysis the authors used the equations (1), (2) and (3): Computation of the two compared standard variances is following:

$$D_x = \sum_{K=0}^{N-1}\left[X(k) - \frac{1}{N}\sum_{K=0}^{N-1}X(k)\right]^2 \qquad D_Y = \frac{1}{N}\sum_{K=0}^{N-1}\left[Y(k) - \frac{1}{N}\sum_{K=0}^{N-1}Y(k)\right]^2 \tag{1}$$

where $X(k)$ and $Y(k)$ are values of curves of analysed phase connection A-C and B-C, N is the number points.

Computation of the two compared covariances:

$$C_{XY} = \frac{1}{N}\sum_{K=0}^{N-1}\left[X(k) - \frac{1}{N}\sum_{K=0}^{N-1}X(k)\right]^2 \cdot \left[Y(k) - \frac{1}{N}\sum_{K=0}^{N-1}Y(k)\right]^2 \tag{2}$$

For computation of two analyzed states the authors used Normalization factor of covariance.

$$R_{XY} = \frac{C_{XY}}{\sqrt{D_X D_Y}} \tag{3}$$

In the frequency range from 1 kHz to 100 kHz, the value of R_{xy} was equal to 0.35 for no-load state. Magnitude depending on frequency for winding-phase A-C and B-C was depicted (Fig.8.).

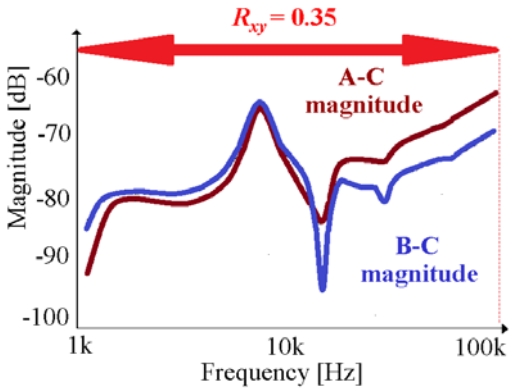


Fig.8. Magnitude depending on frequency for winding-phase A-C and B-C.

Analysis (Fig.8.) showed deformation anomaly due to short-circuit of the transformer. Damage to the power transformer is the result of damage of the phase coil A (visual damage in the Fig.3.).

5. TIME ANALYSIS OF THE TRANSFORMER

Time method using the impact test is often used for analysis of the insulation between coils. It is also used for detection of the attenuated winding sections of electrical machines (also transformers). This method allows us to verify early states of the coil faults. Short-time voltage pulses are used for the winding in order to form a voltage gradient across the coil of the winding. The coil dampens oscillations through transient sinus form at the time intervals among pulses. Each transformer winding has unique character of response. It can be analyzed by oscilloscope. Wave form is influenced by transient circuit depending on the winding inductance and inside capacity of the pulse generator. A diagram for the impact test method of three-winding power transformer is presented in Fig.9.

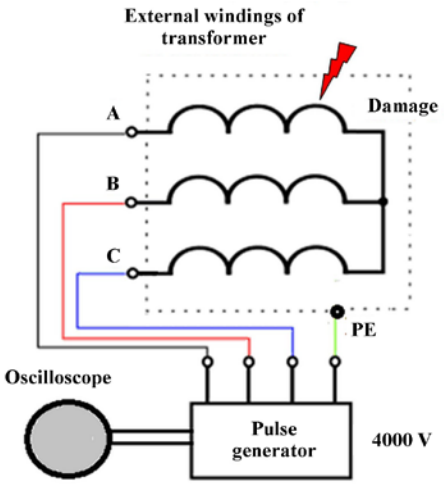


Fig.9. A diagram for the impact test method of three-winding power transformer.

Voltage of coils A and B depending on time was shown in Fig.10. It compares time curves of pulse test measurement of the transformer.

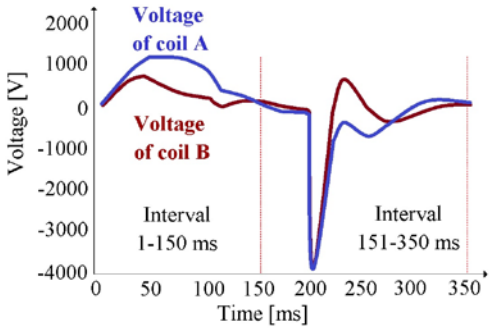


Fig.10. Analysis and comparison of transformer coils A and B using impact test.

The differences between voltage of two coils can be noticed in the range of 0-350 ms. Measurement was carried out using test of two coils. Phase of coil A is influenced by short-circuit. Phase of coil B is well visibly. The decrease of amplitude at the interval 2 (151-350 ms) is caused by change of the resistance and capacitance. It was caused by damaged insulation of coil A. State of the transformer windings can be detected if we know single curves overlap (with similar fault-free curve). Mutually shifted dependents indicate the permanent damage of one coil. Next it analyses time and amplitude differences between states. Two impulses were analyzed for two intervals. The values of amplitude coefficient related to respective ratio at time intervals were showed in Fig.11. For pairs of coils A-B and A-C at interval 0-150 ms, the maximum value of ratio is equal to 5.47. At time interval 150-350 ms, the maximum of ratio is equal to 8.11. The lower value of amplitude was for pair of fault-free coils B-C. The short-circuit as well as permanent deformation of coils A-C had higher value of amplitude.

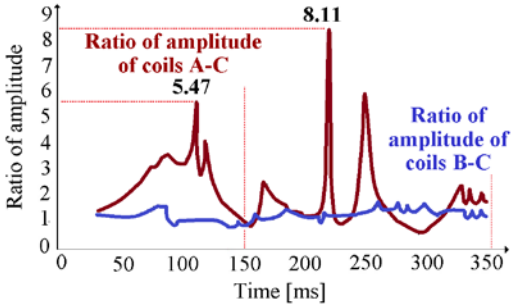


Fig.11. Ratio of amplitude of analyzed coils A-C, B-C of the transformer.

6. Conclusions

In the paper the authors showed the importance of knowledge about theoretical and experimental analysis of the effect of short-circuit forces. Measurements of transformer were carried out in high-voltage laboratory. The analyzed short-circuit forces caused non-reversible deformation of the winding of the power transformer. The proposed analyses showed that the effects of short-circuit current need proper diagnostic and measuring methods. Analyzed state of the transformer was dangerous for operation. Coils of transformer windings were diagnosed by means of the SFRA method and time impact test. The analyzed methods had a significant sensitivity for small deformation of coil. The proposed methods allowed us to analyze short-circuit of the coil of the transformer winding. The advantage of analyzed methods is that they can also detect the fault area in the transformer or its winding.

Fault diagnosis measuring methods should be developed to increase the reliability of electrical equipment. The proposed approach can be used for other types of transformers. Moreover, the presented techniques have potential application for fault diagnosis of electrical equipment such as: transformers and electrical machines.

In the future, the authors will analyze the effects of other short-circuit currents. The authors will also measure dielectric parameters of transformers. It would be also a good idea to analyze moisture of the insulating paper and the conductivity of the oil transformer.

Acknowledgment

Article was financed by the Ministry of Education of Slovak Republic, the Grant Agency VEGA under contract No. 1/0602/17.

References

[1] Hrabovcova, V., Rafajdus, P., Franko, M., Hudák, M. (2009). *Measuring and Modelling of Electrical Machines*. Žilina, Slovakia: EDIS. ISBN 978-80-8070-924-2. (in Slovak)

[2] Wang, T., He, Y.G., Luo, Q.W., Deng, F.M., Zhang, C.L. (2017). Self-powered RFID sensor tag for fault diagnosis and prognosis of transformer winding. *IEEE Sensors Journal*, 17 (19), 6418-6430.

[3] Islam, M.M., Lee, G., Hettiwatte, S.N. (2017). A nearest neighbour clustering approach for incipient fault diagnosis of power transformers. *Electrical Engineering*, 99 (3), 1109-1119.

[4] Zhang, Y.Y., Wei, H., Liao, R.J., Wang, Y.Y., Yang, L.J., Yan, C.Y. (2017). A new support vector machine model based on improved imperialist competitive algorithm for fault diagnosis of oil-immersed transformers. *Journal of Electrical Engineering & Technology*, 12 (2), 830-839.

[5] Peimankar, A., Weddell, S.J., Jalal, T., Lapthorn, A.C. (2017). Evolutionary multi-objective fault diagnosis of power transformers. *Swarm and Evolutionary Computation*, 36, 2017, 62-75.

[6] Ding, Y., Liu, Q. (2017). Data-driven fault diagnosis method for power transformers using modified Kriging model. *Mathematical Problems in Engineering*, 2017, art. ID 3068548.

[7] Wei, C.H., Long, H., Yan, L. (2017). Investigate transformer fault diagnosis performance of dissolved gas analysis with measurement error. *Electric Power Components and Systems*, 45 (8), 894-904.

[8] Yang, Q., Su, P.Y., Chen, Y. (2017). Comparison of impulse wave and sweep frequency response analysis methods for diagnosis of transformer winding faults. *Energies*, 10 (4), art. no. 431.

[9] Li, W.L., Liu, W.J., Wu, W., Zhang, X.B., Gao, Z.H., Wu, X.H. (2016). Fault diagnosis of star-connected auto-transformer based 24-pulse rectifier. *Measurement*, 91, 360-370.

[10] Ballal, M.S., Suryawanshi, H.M., Mishra, M.K., Chaudhari, B.N. (2016). Interturn faults detection of transformers by diagnosis of neutral current. *IEEE Transactions on Power Delivery*, 31 (3), 1096-1105.

[11] Dai, C.X., Liu, Z.G., Hu, K.T., Huang, K. (2016). Fault diagnosis approach of traction transformers in high-speed railway combining kernel principal component analysis with random forest. *IET Electrical Systems in Transportation*, 6 (3), 202-206.

[12] Rigatos, G., Siano, P. (2016). Power transformers' condition monitoring using neural modeling and the local statistical approach to fault diagnosis. *International Journal of Electrical Power & Energy Systems*, 80, 150-159.

[13] Illias, H.A., Chai, X.R., Abu Bakar, A. (2016). Hybrid modified evolutionary particle swarm optimisation-time varying acceleration coefficient-artificial neural network for power transformer fault diagnosis. *Measurement*, 90, 94-102.

[14] Mejia-Barron, A., Valtierra-Rodriguez, M., Granados-Lieberman, D., Olivares-Galvan, J.C., Escarela-Perez, R. (2018). The application of EMD-based methods for diagnosis of winding faults in a transformer using transient and steady state currents. *Measurement*, 117, 371-379.

[15] Wang, C., Zhang, Y.G. (2015). Fault correspondence analysis in complex electric power systems. *Advances in Electrical and Computer Engineering*, 15 (1), 11-16.

[16] Li, Z.X., Jiang, Y., Hu, C.Q., Peng, Z.X. (2017). Difference equation based empirical mode decomposition with application to separation enhancement of multi-fault vibration signals. *Journal of Difference Equations and Applications*, 23 (1-2), 457-467.

[17] Li, Z.X., Jiang, Y., Hu, C., Peng, Z. (2016). Recent progress on decoupling diagnosis of hybrid failures in gear transmission systems using vibration sensor signal: A review. *Measurement*, 90, 4-19.

[18] Glowacz, A., Glowacz, W., Glowacz, Z. (2015). Recognition of armature current of DC generator depending on rotor speed using FFT, MSAF-1 and LDA. *Eksploatacja i Niezawodnosc-Maintenance and Reliability*, 17 (1), 64-69.

[19] Brandt, M., Kascak, S. (2016). Failure identification of induction motor using SFRA method. In *2016 ELEKTRO: 11th International Conference*, 16-18 May 2016. IEEE, 269-272.

[20] Petras, J., Kurimsky, J., Balogh, J., Cimbala, R., Dzmura, J., Dolnik, B., Kolcunova, I. (2016). Thermally stimulated acoustic energy shift in transformer oil. *Acta Acoustica United with Acoustica*, 102 (1), 16-22.

[21] Brandt, M. (2016). Identification failure of 3 MVA furnace transformer. In *Diagnostic of Electrical Machines and Insulating Systems in Electrical Engineering (DEMISEE)*, 20-22 June 2016. IEEE, 6-10.

[22] Chen, W.G., Liu, J., Wang, Y.Y., Liang, L.M., Zhao, J.B., Yue, Y.F. (2008). The measuring method for internal temperature of power transformer based on FBG sensors. In *2008 International Conference on High Voltage Engineering and Application (ICHVE)*, 9-12 November 2008. IEEE, 672- 676.

[23] Werelius, P., Ohlen, M., Adeen, L., Brynjebo, E. (2007). Measurement considerations using SFRA for condition assessment of Power Transformers. In *2008 International Conference on Condition Monitoring and Diagnosis*, 21-24 April 2008. IEEE, 898-901.

[24] Chitaliya, G.H., Joshi, S.K. (2013). Finite Element Method for designing and analysis of the transformer – A retrospective. In *International Conference on Recent Trends in Power, Control and Instrumentation Engineering (PCIE 2013*). Association of Computer Electronics and Electrical Engineers, 54-58.

[25] Heathcote, M.J. (2007). *The J & P Transformer Book*, 13th Edition. Elsevier, ISBN 978-0-7506-8164-3.

Received April 17, 2018
Accepted July 10, 2018

MEASUREMENT SCIENCE REVIEW, **18**, (2018), No. 4, 168-174

sciendo

ISSN 1335-8871

MEASUREMENT SCIENCE REVIEW

Journal homepage: http://www.degruyter.com/view/j/msr

Measuring and Evaluating the Differences of Compared Images for a Correct Car Silhouette Categorization using Integral Transforms

Elena Pivarčiová[1], Daynier Rolando Delgado Sobrino[2], Yury Rafailovich Nikitin[3], Radovan Holubek[2], Roman Ružarovský[2]

[1]*Department of Manufacturing and Automation Technology, Faculty of Environmental and Manufacturing Technology, Technical University in Zvolen, Masarykova 24, 960 01 Zvolen.*
[2]*Institute of Production Technologies, Faculty of Materials Science and Technology in Trnava, Slovak University of Technology in Bratislava, Jána Bottu street, 25, 917 24, Trnava, Slovak Republic.*
[3]*Department of Mechatronic Systems, Faculty of Quality Management, Kalashnikov Izhevsk State Technical University, Studencheskaya street, 7, 426069, Izhevsk, Russia.*

The present paper focuses on the analysis of the possibilities of using integral transforms for measuring and evaluating the differences of compared images (car silhouettes) with the purpose of a correct car body categorization. Approaches such as the light intensities frequency change, the application of discrete integral transforms without the use of further supplementary information enabling automated data processing using the Fourier-Mellin transforms are used within this work. The calculation of the several metrics was verified through different combinations that implied using and not using the Hamming window and a low-pass filter. The paper introduced a method for measuring and evaluating the differences in the compared images (car silhouettes). The proposed method relies on the fact that the integral transforms have their own transformants in the case of translation, scaling and rotation, in the frequency area. Besides, the Fourier-Mellin transform was to offer image transformation that is resistant to the translation, rotation and scale.

Keywords: Automatic measurement system, Fourier-Mellin transform, image analysis, Body in White (BIW), car silhouettes.

1. INTRODUCTION

Even when it has been long known that perceived quality is one of the most important factors underlying the success of car manufacturers today [1], sometimes some of the more conventional methods used for quality control in the automotive industry are regarded as outdated and ineffective to a certain extent. Thus, given this importance of product control in the manufacturing process and the challenging nature of ensuring that the quality of each product meets the standards required by the clients in an efficient and effective manner, the combined use of such conventional (manual) and new automated methods of control have been gaining more and more popularity in recent years and significant research has been focused on the topic.

In the middle of all these needs for quality control and measurements, automatic computer vision has particularly and increasingly become an important tool in the measurement for the evaluation of quality and in the recognition of several types of effects in the automotive industry, see the relevant works appearing in [2] for a better understanding of this. The main goal of a vision measurement system lies in trying to emulate, to a given extent, some of the human vision, sense, image and scene evaluation capabilities, and this by electronic and informatics means involving tasks as recognition, digitalization and evaluation, among others. If, for example, a measured input is sufficiently accurate and does not require specific difficult processing of input data (images), the system can autonomously react to it and good results are reachable relatively easily. On the contrary, if the same system were to be implemented into a real environment, and its inputs were not images with clearly and distinctly defined objects of a constant size, rate and observation angle, the whole process would not be so easy and those images would require significant pre-processing to get the proper reaction from the system. Similarly, the accuracy of measurements also depends on a delicate process of offline calibration, and in some practical applications, the system is even supposed to be reconfigured on an almost continuous basis so as to track the target that an online calibration requires [3]. In the end, the main goal of all this reflects and enhances the need of building an automatic measurement

DOI: 10.1515/msr-2018-0024

system for quality control in the automotive industry without the requirements of constant or frequent human interference. This proposal should be also general and wide enough so that it can be used in other branches of manufacturing.

Automated visual measuring is known to be an image-processing method that has been widely applied over the years in the production line for quality control purposes [4], [5]. This focuses on mechanical parts, vehicles, casting production and even the garment industry [6], [7]. Inspection tasks, however, are time consuming and have been usually and mostly carried out through human intervention, i.e.: human inspectors. On the other hand, the inspectors' performance has been often a subject of debate in the scientific community and practice, making some authors state and prove to an extent that this is often inadequate and its accuracy is influenced by the fatigue of the intense and repetitive tasks. In addition, human inspectors need training which together with the needed skills implies investments and time to develop. Also, if compared to machines, the hours of human inspectors are relatively limited, and labor costs are a major factor posing a challenge for most manufacturers trying to compete in a densely populated and competitive marketplace [8]. On the other hand, an automatic measurement system needs almost no human presence and thus, eventually, this translates into more time available to the experts to better and more effectively work in the measurement and to analyze the composite image data increasing the speed and effectiveness of their work [9].

Some of the key advantages of computer machine vision measurement systems lie in the use of contactless non-destructive measuring [10], the use of ultraviolet and infra-red radiation for the exploring and inspection tasks, as well as in the augmented flexibility allowing for immediate changes of system configuration and the possibility of parallel measuring of more parameters at the same time. These types of systems are applied in many areas of society, for example: in traffic, medicine, manipulators and robotics [11], [12], assembly cells, production of synchronous DC motors [13], welding [14] and of course, in quality inspection.

In this study, as it is in principle similarly stated in [15], an automatic measurement system is understood and seen as an optical system with image detection and verification which is to be used for the detection and proper categorization (classification and sorting) of car silhouettes. One of the most problematic issues for the automatic measuring system relates to the comparison of the recorded images with a reference template, i.e.: car silhouette template comparison [16]. Besides, the system also needs to possess the ability to detect whether the given casting is faultless or defective. This last task is relatively easy for the human eye and mind, because humans have experience which has been in many cases cumulated since birth. However, this same problem is difficult to process for a computer, getting computers to fully understand things and the effects of the surrounding world has been a myth for long and still is a present challenge for science.

During the car production/assembly phase, it is highly necessary to monitor and control the production process to detect potential errors of different types and nature and, among other things, to also sort the car bodies by specific models and distribute them according to their type and the place where such parts are required, see Fig.1. One of the key parts of the process is the one related to the creation of the Body in White (BIW) assemblies. In the process of their obtaining, all models of cars are welded on a single line. This welding line almost always contains samples of welded car bodies of all models which are to be compared with the ones being produced every single time. Here an automatic visual measuring system plays a decisive role in the quality of the car bodies, and robots are mainly the ones responsible for "seeing, comparing, and controlling that these match what the ideal car of a given model should look like, in other words: they are to compare the model of a given product with its equivalent product being manufactured at a given time.

Fig.1. 3D inline measuring technology from ZEISS [17].

Another possible application for this type of automated system is found in the distribution of the BIW to the respective assembly halls/lines, in other words, in the sorting of these according to individual car body types.

Taking into consideration all the above-mentioned elements, in this paper we propose a method that compares the reference image with the tested image using the amplitude and phase spectra of the Fourier-Mellin transform, also their invariability to translation, rotation and scale. Given that the Fourier-Mellin transform reduces the mutual rotation of the images along the horizontal axis and the shift in scaling along the vertical axis, then any two images, one of which may offset, rotated and/or with a change in scale, have identical Fourier-Mellin transform amplitudes. For comparison purposes and/or the determination of the level of similarity, we used in the paper several metrics, for example the phase correlation and percentage comparison analysis.

2. SUBJECT & METHODS

A phase-only correlation technique/method that is based on Fourier-Mellin transform was used in the automatic measurement system for the purposes of image analysis and comparison.

A. The basics of the Fourier-Mellin transform

The Fourier-Mellin transform allows comparing images which are offset, rotated and have changed scale. This method takes advantage of the fact that the shift differences are annulated because the amplitude spectrum of the image and its displaced copy are identical, only their phase spectrum is different. Subsequently the log-polar transform causes the rotation and scale to appear outwardly as a shift. On the other hand, phase correlation is based on the fact that two similar images in a cross-spectrum create a continuous sharp extreme in the place of registration and the noise is distributed randomly in discontinuous extremes.

This is the reference image $a(x, y)$ and tested image $b(x, y)$, that happens to be the copy of the image $a(x, y)$ and is rotated, translated and has changed scale.

In the case of the Fourier transform of images (1), it applies:

$$\begin{aligned} A(u,v)&=F\{a(x,y)\}=|A(u,v)|e^{-j\theta(u,v)} \\ B(u,v)&=F\{b(x,y)\}=|B(u,v)|e^{-j\theta(u,v)} \end{aligned} \quad (1)$$

where $|A(u,v)|$, $|B(u,v)|$ are the spectral amplitude, $\theta(u,v)$ is the phase, $F\{.\}$ is the Fourier transform [16].

On the other hand, the Fourier transform of log-polar amplitudes is equivalent to the Fourier-Mellin transform [18]. Next, we detail the following cases:

a) The log-polar transform of amplitudes $|A(u,v)|$, $|B(u,v)|$ from the Cartesian coordinate system to the log-polar coordinate system.

The Fourier transform is displayed next in a log-polar plane as indicated by the transformation of coordinates, Fig.2.

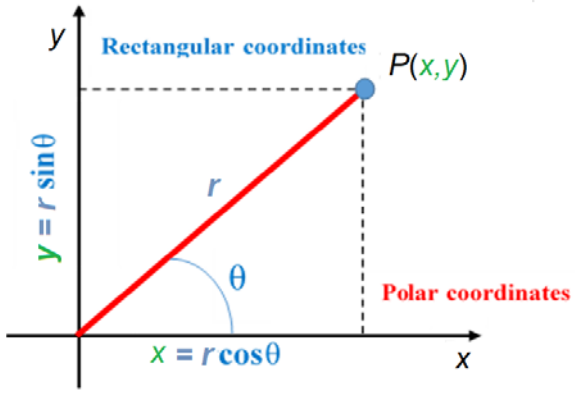


Fig.2. Transform of rectangular to polar coordinates.

In order to convert from Cartesian to log-polar coordinates we do as stated in (2):

$$r=\sqrt{x^2+y^2} \quad (2)$$

The beginning of the log-polar coordinate system (m_0, n_0) should be in the middle of the video matrix, and this in order to ensure the maximum number of pixels possible.

The maximum sampling radius for the conversion will be defined as in (3):

$$\begin{aligned} \rho_{max} &= \min(m_0, n_0) \ldots inscribed\ circle \\ \rho_{max} &= \min(m_0, n_0) \ldots inscribed\ circle \end{aligned} \quad (3)$$

The relationship between the polar coordinates (ρ, θ), used for the samples of the test image, and the log-polar image (e^r,θ) is explained by (4):

$$(\rho, \theta)=(e^r, \theta) \quad (4)$$

For mapping the pixels of the test image (x_i, y_i) to the pixels of the output image (r_m, θ_n) (5) was used, see [19], where $(\rho_m, \theta_n) = (e^{r_m}, \theta_n)$

$$\begin{aligned} x_i &= round(\rho_m . \cos\theta_n + m_0) \\ y_j &= round(\rho_m . \sin\theta_n + n_0) \end{aligned} \quad (5)$$

b) Fourier transform of log-polar amplitudes (6).

$$\begin{aligned} \boldsymbol{A}_{ip}(\boldsymbol{v}, \boldsymbol{\omega}) &= F\{|A_{ip}(e^r, \theta|\} \\ \boldsymbol{B}_{ip}(\boldsymbol{v}, \boldsymbol{\omega}) &= F\{|B_{ip}(e^r, \theta|\} \end{aligned} \quad (6)$$

In this case the log-polar transform of the amplitude spectrum causes the rotation and scale to be shown as a translation, these can be better understood from the compound figure that appears below, where the amplitude and the phase spectrum of the Fourier transform of the real image of the car part appears in *a*), and its Fourier spectrum in log-polar coordinates is shown in *d*).

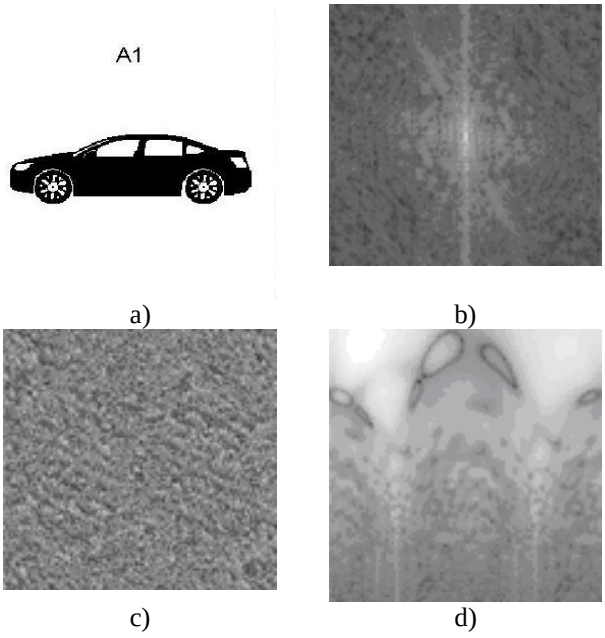


Fig.3. The amplitude and the phase spectrum of the Fourier transform: a) Original image, b) Fourier spectrum amplitude in the Cartesian coordinates, c) Fourier spectrum phase in the Cartesian coordinates, d) Fourier spectrum in log-polar coordinates.

B. Comparison of the observed images of measured objects and their mutual interaction

In general terms the use of the Fourier-Mellin transform reduces the mutual rotation of images to translation along the horizontal axis, and also the change of scale to vertical axis translation [20]. Both of these cases/spectra were compared. These authors analyzed two images, one of which was moved and/or rotated and/or had the scale changed, and they proved that both ended up having identical amplitude of Fourier-Mellin transform. Based on this study, reference designs were compared in our case with test images in order to determine the degree of similarity and correlation between them for the purposes of our automatic measurement system.

Besides, several well-known and other more specific metrics were also used for the calculation of a comparative score helping to evaluate the results of the comparison of the obtained images. This score/metrics quantifies the similarities between the test image and the reference image. The calculation of the metrics has been verified in different combinations, and this with or without the application of the Hamming window, and similarly with or without the application of the low-pass filter. The representative metrics are used to determine the similarity between reference and test images. These metrics are detailed next.

POC, Phase-Only Correlation

This comparative metric is calculated as the maximum phase correlation according to (7), where *F* stands for Fourier transform of images *a*, *b*, F^{-1} is the inverse Fourier transform, and *F** is the complex conjugate image [20].

$$\mathrm{C} = F^{-1}\left(\frac{F_a F_b{}^*}{|F_\mathrm{a}||F_\mathrm{b}|}\right) \tag{7}$$

The phase correlation gives a distinct maximum if two images are similar. On the contrary, a more insignificant maximum will be created if two images are not similar. The size of the maxima is used as a measure of similarity between two images.

MPOC, Modified Phase-Only Correlation

Since the energy of the signal is lower in the high-frequency domain, phase components are not reliable in high-frequency ones [16]. The effect of unreliable phase components in high frequencies can be limited by using filters or modifying the POC function by using the spectral weighting function.

The function of spectral weighting *W*(*u*,*v*) has been used to improve the detection by removing minor ingredients with high frequency, which have a low reliability, where u, v are 2D coordinates, β is a parameter, which checks the width of the function, and α is used only for normalizing purposes (8) [21].

$$\mathrm{W}(u,v) = \left(\frac{u^2+v^2}{\alpha}\right) e^{-\frac{u^2+v^2}{2\beta^2}} \tag{8}$$

Such a modified image phase correlation function of ***a*** and ***b*** appearing in (9) is given by [27].

$$\tilde{q}_{a,b}(x,y) = F^{-1}\left\{W(u,v)\frac{F_a(u,v)F_b^*(u,v)}{|F_a(u,v)||F_b(u,v)|}\right\} \tag{9}$$

The extreme value of a function $\tilde{q}_{a,b}(x,y)$ is invariable at the change of shift and brightness. In our case, this value was used to measure the similarity of images: if two images are similar, their function MPOC gives a crisp extreme, if they are different, then the extreme decreases considerably.

PD, Percent Discrimination

This metric has relative amount of similarity between reference and test image as seen in (10) [22].

$$\mathrm{PD} = \frac{2[C_{ab}]_{max}}{[C_{aa}]_{max}+[C_{bb}]_{max}} x100\% \tag{10}$$

where: $[C_{aa}]_{max}$,$[C_{ab}]_{max}$, and $[C_{bb}]_{max}$ are maximal phase correlations if they are compared to a reference image *a*(*x*, *y*) with itself, tested image *b*(*x*, *y*) with itself, and the reference image *a*(*x*, *y*) with the image *b*(*x*, *y*).

These metrics were used to compare the log-polar amplitudes of Fourier-Mellin transform of images.

C. A functional testing of the proposed method

Designs of car body silhouettes available over the internet have been used as an example for the functional testing of the proposed method, see Fig.4. For the image comparison purposes, 3 basic car silhouette samples were considered, and 10 images were used for each case what yielded a total of 30. Besides, in order to obtain better and more reliable results, a series of extra comparison tests were conducted where each image was compared with the others and itself, what in the end generated a total of 90 tested images. It is important to underline that 30 of these were comparison of correct (right) pairs and the remaining 60 of incorrect ones.

Fig.4. The design of car bodies, Stock Vector Illustration. Silhouette cars on a white background.

The before-mentioned metrics Maximum phase correlation (POC, Phase Only Correlation), Modified phase correlation (MPOC, Modified Phase Only Correlation), Percentage comparison (PD, Percent Discrimination)) were used to determine the similarity of the reference and test images –

calculation of the comparison score. The calculated score was compared with the threshold value, which determines the degree of correlation necessary for the comparison to be taken as a match.

The different threshold values *t* regulating the decision-making process were set based on several performed tests:

- images generating a result higher than *t* are evaluated as identical or highly similar, i.e.: it can be stated it is the same car body,
- images generating results lower than or equal to *t* are regarded to be inconsistent, i.e.: it cannot be stated it is the same car body.

The possible decisions resulting from the system are of the type: match, mismatch, and no consistent result at all. It is however possible to find varying degrees of strong matches and mismatches.

The number of correctly rejected and correctly accepted images depends on the threshold value defined, see Fig.5. The value is adjustable depending on the requirements of the process, what may make the system more or less accurate in some cases. As the test results showed, the thresholds values *t* of the relevant criteria was set as follows:

- threshold values I.
- threshold values II.
- selected threshold value t
- it's not the same car body
- minimum of the correct pairs
- maximum of the wrong pairs
- maximum of the threshold values I. and II.
- score ≤ threshold value

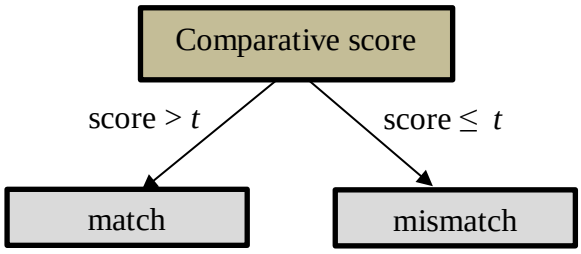


Fig.5. Comparative score issuing a match and/or mismatch of images.

3. Results

In our particular case, the images were adjusted to 256 shades of grey and to the dimensions of 256 × 256 pixels for obtaining faster results with the Fourier transform. The test images with the parameters for the transformation were created from the basic/primary images. The images were translated in the horizontal and/or vertical direction, rotated around the different angles in both directions and also had the scales changed. The resulting images were compared with other car bodies.

Using the proposed method, the system identified 93 percent of correct products. On the other hand, only the remaining 7 percent of the products that were identified by the system to be correct (red color in graph), happened to be incorrect instead, and this was based on the analysis of the images as described in our method, see Fig.6. for a better understanding of the results. The results proved the detection method used was valid, and thus it could be used and become a significant asset in the classification, sorting and inclusion of car bodies into the relevant categories (categorization).

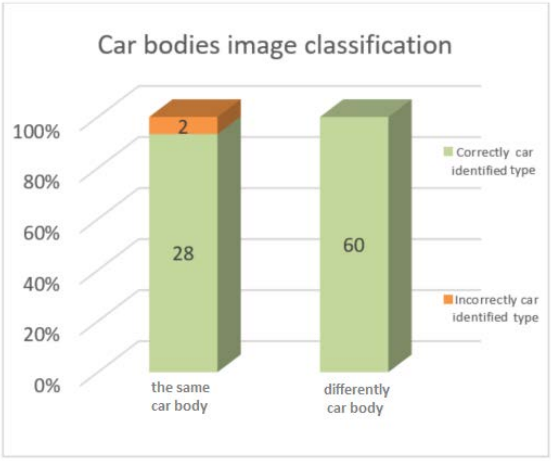


Fig.6. Results of the tested images.

The high success rate was achieved due to the use of a comparative set with rather small angle of rotation and scaling. The maximum correct recognition limit of this method has not been tested yet. Similarly, images have not been properly recognized in the case of the scale being too changed (0.7 and 1.3), and/or in combination with translation and rotation.

4. Conclusion

As mentioned before, the measurement process is a key part of any kind of production, what is especially important in the case of the automotive industry. One of the main outputs of any measurement process is the achievement of high levels of quality of the products. This represents the main premise for the company success and its positioning among the competitors. Besides, this all fully depends on the use of new technologies, automation and implementation of methods into the related manufacturing processes, modified based on [23], [24].

The verification of quality by automatic measurement systems provides an even higher chance to compete in the related field as well as the level of quality checks allows its further improvement. In the course of the manufacturing process or in the product completion various defects can occur, as it is stated in the ISO/TS16949, which is technical specification aimed at the development of a quality management system that provides for continual improvement, emphasizing defect prevention and the reduction of variation and waste in the supply chain.

The calculated metrics were compared to the verification threshold that determined the degree of correlation necessary for the comparison, issuing what was to be taken as match. Based on tests carried out, the threshold values of ***t*** were set, according to which the system decision is always regulated: images generating the result greater than or equal to ***t*** are evaluated as identical, in other words: It is the same car body

part; on the other hand, images generating results lower than ***t*** are evaluated as non-compliant, i.e.: these are not the same car body parts.

The decisions of system can be: match, mismatch, and without result, even though there are possible changing degrees of strong matches and mismatches.

The number of properly rejected and properly accepted images depends on the present threshold value of ***t***. The value is adjustable depending on the requirements, thus the system could be more or less accurate depending on this. In other words, it all depends on each particular case study and the levels of confidence and requirements needed for the verifications. In the case of the automotive industry and our case in particular, these values tend to be of 100 %, i.e.: the threshold should be established so as to have no errors in the process of detection.

5. Conclusions

The evolution of computing technology is progressing hand in hand with the evolution of other natural, technical or economic sciences. Adequate powerful computing technology enables high precision analysis of process models that more precisely describe the physical or social nature of things [25].

The paper describes the design of a method for the correct car body classification (car silhouettes) based on the principles of Fourier-Mellin transform and subsequent comparison and analysis of images using Fourier transform and phase correlation. The Fourier-Mellin transform proved to offer image transformation resistant to translation, rotation and scale. The proposed method uses the fact that integral transforms have their transformants in the case of translation, scaling and rotation in the frequency domain.

With the use of the automatic measurement system it was possible to compare silhouettes of car bodies and find out if these belonged to the same type or not. This was supported by the use of several key criteria for the matches, such as the phase correlation, the difference correlation, the correlation coefficient, the percentage comparison and the comparison of calculated values with the chosen threshold for the relevant criterion. In the frame of the experiments, threshold values have been set in a way that all the wrong couples of body works can be revealed.

Acknowledgment

This publication is the result of implementation of the project “University Scientific Park: Campus Mtf Stu – CAMBO” (ITMS: 26220220179) supported by the Research and Development Operational Program funded by the EFRR and by project KEGA MŠ SR 003TU Z-4/2016: Research and education laboratory for robotics and RFBR research project № 18-08-00772 A.

References

[1] Stylidis, K., Wickman, C., Söderberg, R. (2015). Defining perceived quality in the automotive industry: An engineering approach. *Procedia CIRP*, 36 (2015), 165–170.

[2] Bogue, R. (2013). Robotic vision boosts automotive industry quality and productivity. *Industrial Robot: An International Journal*, 40 (5), 415–419.

[3] Di Leo, G., Liguori, C., Pietrosanto, A., Sommella, P. (2017). A vision system for the online quality monitoring of industrial manufacturing. *Optics and Lasers in Engineering*, 89, 162–168.

[4] Ružarovský, R., Delgado Sobrino, D.R., Holubek, R., Košťál, P. (2014). Automated in-process inspection method in the flexible production system iCIM 3000. *Applied Mechanics and Materials*, 693, 50–55.

[5] Božek, P., Pivarčiová, E. (2013). Flexible manufacturing system with automatic control of product quality. *Strojarstvo*, 55 (3), 211–221.

[6] Mery, D., Jaeger, T., Filbert, D. (2002). *A review of methods for automated recognition of casting defects*. http://www.academia.edu/20111824/A_review_of_methods_for_automated_recognition_of_casting_defects.

[7] Świłło, S.J., Perzyk, M. (2013). Surface casting defects inspection using vision system and neural network techniques. *Archives of Foundry Engineering*, 13 (4).

[8] Dhillon, B.S. (2009). *Human Reliability, Error, and Human Factors in Engineering Maintenance*. CRC Press.

[9] Huang, S.-H., Pan, Y-Ch. (2015). Automated visual inspection in the semiconductor industry: A survey. *Computers in Industry*, 66, 1–10.

[10] Frankovský, P., Ostertag, O., Trebuňa, F., Ostertagová, E., Kelemen, M. (2016). Methodology of contact stress analysis of gearwheel by means of experimental photoelasticity. *Applied Optics*, 55 (18), 4856–4864.

[11] Kováč, J., Ďurovský, F., Hajduk, M. (2014). Utilization of virtual reality connected with robotized system. *Applied Mechanics and Materials*, 613, 273–278.

[12] Frankovský, P., Hroncová, D., Delyová, I., Hudák, P. (2012). Inverse and forward dynamic analysis of two link manipulator. *Procedia Engineering*, 48, 158–163.

[13] Abramov, I.V., Nikitin, Yu.R., Abramov, A.I., Sosnovich, E.V., Božek, P. (2014). Control and diagnostic model of brushless DC motor. *Journal of Electrical Engineering*, 65 (5), 277– 282.

[14] Jena, D.B., Kuma, R. (2011). Implementation of wavelet denoising and image morphology on welding image for estimating HAZ and welding defect. *Measurement Science Review*, 11, (4).

[15] Neogi, N. Mohanta, K.D., Dutta, K.P. (2014). Review of vision-based steel surface inspection systems. *EURASIP Journal on Image and Video Processing*, 2014 (50).

[16] Ito, K., Nakajima, H., Kobayashi, K., Aoki, T., Higuchi, T. (2004). A fingerprint matching algorithm using Phase-Only Correlation. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E87–A (3), 682–691.

[17] Carl Zeiss Ltd. (2018). *3D inline measuring technology from ZEISS*. https://www.zeiss.co.uk.

[18] Druckmüller, M., Antoš, M., Druckmüllerová, H. (2005). Mathematical methods for visualization of the solar corona. *Jemná mechanika a optika*, 10, 302–304.

[19] van den Dool, R. (2004). *Fourier and Mellin Transform. Image Processing Tools*. www.scribd.com/doc/9480198/Tools-Fourier-Mellin-Transform.

[20] Derrode, S., Ghorbel, F. (2001). Robust and efficient Fourier–Mellin transform approximations for gray-level image reconstruction and complete invariant description. *Computer Vision and Image Understanding*, 83 (1), 57–78.

[21] Gueham, M., Bouridane, A., Crookes, D. (2007). Automatic recognition of partial shoeprints based on phase-only correlation. In *IEEE International Conference on Image Processing*. IEEE, Vol. 4, 441-444.

[22] Chen, Q.S. (1993). *Image registration and its applications in medical imaging*. Dissertation work, Vrije University, Brussels, Belgium.

[23] Slížik, J., Harťanský, R. (2012). Metrology of electromagnetic intensity measurement in near field. *Quality Innovation Prosperity*, 17 (1), 57–66.

[24] Hallon, J., Kováč, K., Bittera, M. (2018). Comparison of coupling networks for EFT Pulses Injection. *Przeglad elektrotechniczny*, 94 (2), 17–20.

[25] Harťanský, R., Smieško, V., Rafaj, M. (2017). Modifying and accelerating the method of moments calculation. *Computing and Informatics*, 36 (3), 664–682.

Received April 17, 2018
Accepted July 18, 2018