Noise performance of textile-based dry ECG recording electrodes

Soumyajyoti Maji✉ and Martin J. Burke
Department of Electronic and Electrical Engineering, Trinity College, University of Dublin, Dublin, Republic of Ireland
✉E-mail: smaji@tcd.ie

This study reports attempts to characterise the noise properties of un-gelled, textile-based electrodes intended for use in ECG recording in wearable garments. Measurements were made using very low noise operational amplifiers and with mains power supply interference eliminated from the recorded signals so that they reflected only genuine noise. Measurements were carried out on four conductive fabric electrodes, one conductive rubber electrode and a standard self-adhesive, gelled electrode on three male and three female subjects. Curves of the form $y = K/f + C$ were fitted to the recorded noise spectral density functions. Coefficients obtained for the white noise spectral density ranged from 1.1 to 5.6 nV/$\sqrt{}$Hz and for the flicker ($1/f$) noise spectral density ranged from 8 to 80 nV$\sqrt{}$Hz.

Introduction: The skin-electrode interface noise performance is a crucial factor in high-resolution surface bio-potential measurements. Recorded electrocardiogram (ECG) signals have very low amplitudes of 50 μV–10 mV in the bandwidth of 0.5–250 Hz and are highly susceptible to contamination from unwanted interference and noise [1–3]. The use of un-gelled or dry electrodes in the ambulatory recording of heart rate and ECG has emerged in recent years and continues to gain popularity [4, 5]. More recently, textile-based electrodes employing conductive fabrics have been used in wearable garments for remote monitoring and telemedicine applications [6–11]. Dry, un-gelled electrodes have consistently been reported in the literature to be noisier than the standard gelled, self-adhesive electrodes used in hospital clinics [12–14]. This claim is not always supported with experimental evidence. The origin of the noise in surface electrodes has been investigated in the past by several researchers [1, 15–19]. However, there is little or no information available on the noise performance of un-gelled, fabric-based electrodes. This information is essential for the best design of garments using textile electrodes and for the optimisation of low-noise amplifiers to be used with them. This study reports the measurement of the noise voltages generated at the skin-electrode interface for four fabric-based electrodes and their comparison with those of a conductive rubber electrode and a standard self-adhesive electrode. The coefficients of a spectral noise density model having white and flicker noise components of the form $y = K/f + C$ are also established.

Methodology:

Electrode details: Details of the six different electrodes examined are given in Table 1. Measurements were carried out for each electrode on three male and three female volunteers aged between 22 and 64 years and having Irish and Indian nationalities. The electrodes were mounted on the arm of each subject in order to minimise the effects of mains power supply and other extraneous interference.

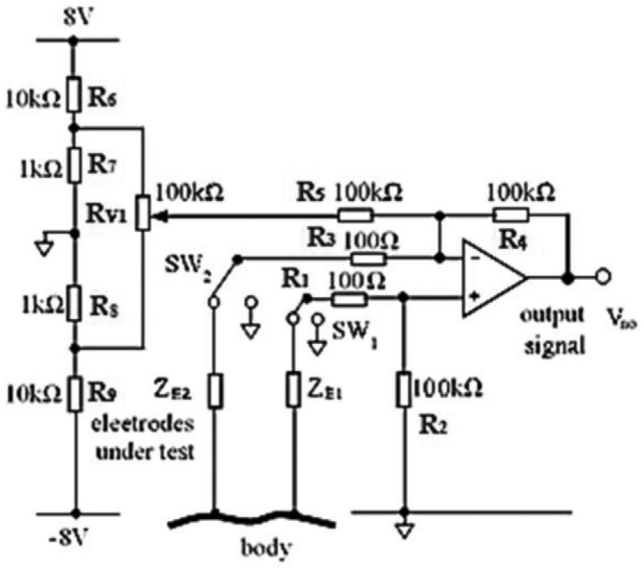


Fig. 1 *Schematic diagram of electrode noise measurement system*

Measurement circuit and experimental setup: A schematic diagram of the measurement circuit used is shown in Figure 1. A low-noise, battery-operated op-amp (LTC2058, Analog Devices Inc.) circuit structure providing a differential gain of 1000 provides the output noise voltage V_{no}. The white noise voltage and current noise spectral densities of this op-amp are extremely low and are quoted by the manufacturer as 9 nV/$\sqrt{}$Hz and 1 pA/$\sqrt{}$Hz, respectively. The op-amp exhibits practically no flicker ($1/f$) noise component and its input-referred noise is stated to be purely white. A resistive potential divider arrangement allows cancellation of electrode polarisation potentials and amplifier offset voltages within the range +800 to –800 mV without recourse to ac coupling. A battery-operated, high-resolution digital oscilloscope (PicoScope 4262) was used in conjunction with a laptop computer to digitise the output noise voltage obtained from the amplifier. This instrument featured a slow recording time-base combined with a high sampling rate and 16-bit amplitude resolution and had internal noise of $\leq$ 0.25 least significant bit (LSB) which was not available in conventional laboratory digital oscilloscopes. The use of battery-operated instrumentation within a Faraday cage and deactivation of the mains power supply in the measurement environment ensured an extremely low level of 50 Hz mains pick-up without the need for grounding the subject through a third electrode. Two identical electrodes were placed at a 3-cm separation on the subject's forearm without any manner of skin preparation. A flexible cotton bandage with an elastic clip fastener was wrapped snugly around the electrodes and arm to hold the electrodes in place, which were then left to stabilise for 5 min. Efforts were made to ensure the same degree of stretching and tension in the elastic clip in each case. During the measurements, subjects were asked to sit as still as possible to avoid

Table 1. Physical details of electrodes examined

Electrode no. and model	Manufacturer	Material type and construction	Dimension (cm)
(1) Red dot 50-2237	3M Healthcare Ltd., Canada	Self-adhesive foam, gel centre (17 mm)	Diam. 4
(2) Self-made	Robot-Shop, Norfolk, UK	Silver woven fabric on 1.5 mm polyester fibre cloth backed with 1.2 mm polyester	5 × 5
(3) Round cloth 50	Current Pleasures Inc., USA.	Silver fibre fabric with sponge backing	Diam. 5
(4) YD-C106	YD Strong Ltd., China	Silver fibre fabric with sponge backing	DIam. 3.5
(5) CGP83294	Medi-Stim Inc., USA.	Silver woven fabric with thick sponge backing and outer cotton fabric.	2.5 × 6.5
(6) WA45	Wandy Ind. Co. Ltd., Taiwan.	Conductive silicone rubber	4.5 × 4.5

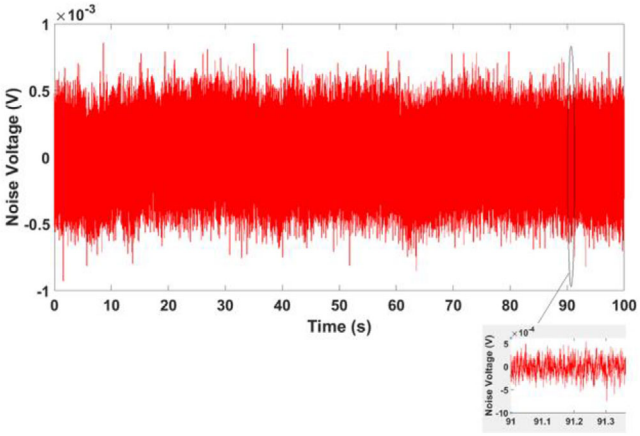


Fig. 2 *Peak-to-peak noise voltage recorded from the amplifier*

movement artefact in the recorded signals. The output noise data were recorded for 100 s at a sampling rate of 1M sample/s and stored in memory. Files were saved in .mat format for further processing in MATLAB (MathWorks 2019b).

Results and discussion: The amplifier noise voltage was measured by grounding the 100 Ω resistors using switches SW_1 and SW_2 shown in Figure 1. The effective resistance at each input terminal of the op-amp is therefore 100 Ω. The *rms* thermal noise voltage produced by this resistance is given as

$$V_{nR_g^-} = V_{nR_g^+} = V_{nR_g} = \sqrt{4kTBR_g} \tag{1}$$

where $k = 1.38 \times 10^{-23}$ JK^{-1} is Boltzmann's constant, $T = 300$ K is the ambient room temperature in Kelvins, $B = 250$ Hz is the ECG signal bandwidth of interest and R_g is the effective noise resistance seen at either input pin of the op-amp. The thermal noise voltage given by Equation (1) is $V_{nRg} = 20.34$ nV$_{\text{rms}}$. Including the semiconductor sources of noise in the op-amp, the total output noise is given as

$$V_{no,rms} = \left(1 + \frac{R_4}{R_3}\right) \times \sqrt{V_{na}^2 + 2V_{nR_g}^2 + 2i_{na}^2 R_g^2} \tag{2}$$

where V_{na} represents the op-amp noise voltage and i_{na} the op-amp noise current. Substituting the values of $V_{na} = 142.3$ nV and $i_{na} = 15.8$ pA obtained for the bandwidth of interest gives $V_{no,rms} = 0.145$ mV$_{\text{rms}}$. Applying a single-pole, low-pass filtre correction factor of 1.57 and a 3.3 scale factor, the peak-to-peak output noise voltage is finally given as:

$$V_{no,ptp} = 1.57 \times 6.6 V_{no,rms} \tag{3}$$

Applying these equations gives a calculated value of $V_{no} = 1.5$ mV$_{\text{ptp}}$. The instantaneous noise voltage recorded from a practical bench test is shown in Figure 2 and has a value of 1.66 mV$_{\text{ptp}}$ which verifies the calculations

Measurements of noise voltage were then conducted on the six volunteers using all of the electrodes listed in Table 1. For the two electrodes present in each test, the electrode noise power can be evaluated by subtracting the squared noise voltage measured with the amplifier inputs grounded from the squared noise voltage obtained with the electrodes placed on a subject's arm. In this case, the noise voltage for a single electrode is given as

$$V_{ne} = \sqrt{0.5\left(V_{nt}^2 - V_{ni}^2\right)} \tag{4}$$

where V_{ne} is the wanted electrode noise voltage, V_{nt} represents the amplifier total output noise voltage recorded with electrodes in place, and V_{ni} denotes the noise voltage with the amplifier inputs grounded. The resulting electrode noise voltage can then be input-referred by dividing by the amplifier gain.

This gave a noise spectral density across the bandwidth of interest for each electrode used on each subject. The noise spectral densities

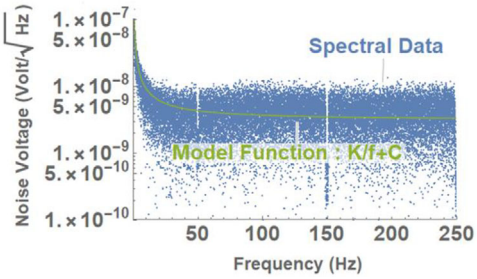


Fig. 3 *Input-referred noise voltage spectrum and model function for male 1, electrode 1*

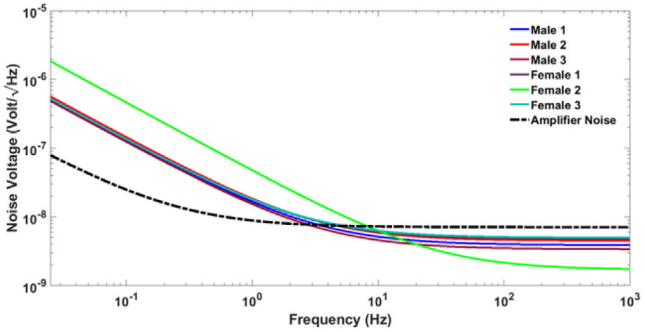


Fig. 4 *Noise spectral density function for electrode 2 across all subjects*

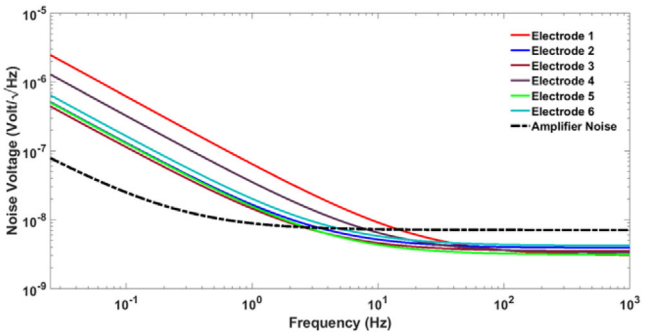


Fig. 5 *Noise spectrum density function for male 1 across all electrodes*

recorded were filtered with narrow-band infinite impulse response (IIR) filters in MATLAB in order to remove any residual 50 Hz mains interference and its third harmonic at 150 Hz. The curve-fitting function 'non-linearFit' in Mathematica (Wolfram Research) was used as shown in Figure 3 to fit a curve of the form of $y = K/f + C$ to each of the noise spectra recorded. This gave a noise model having a white component, C and a flicker or $1/f$ component with coefficient K. This was done for the noise spectrum obtained from each subject for each electrode studied. The input-referred noise spectral density function is shown plotted in Figure 4 for electrode 2 in all subjects and in Figure 5 for male subject 1 across all electrodes. In both cases, the spectral density function is also shown for the amplifier with the inputs grounded. It can be seen that a corner frequency for the amplifier noise can be identified but is extremely low in the region of 0.2 Hz, and the noise is purely white above this frequency rather than from dc. The plots also show that the amplifier noise voltage is higher above about 10 Hz than that of the electrode.

The coefficients of the noise spectral density functions, as well as the corner frequencies, are listed in Table 2 for all electrodes and subjects. Examining Table 2, it can be seen that the gelled adhesive electrode has lower levels of white noise. This is to be expected as this type of electrode provides reliable skin contact and lower skin interface impedance. It can also be seen that there is considerable intra-individual variation as the white noise voltage ranges from 1.1 to 4.6 nV/$\sqrt{\text{Hz}}$. Electrode 4, having the smallest area, exhibits the highest values of white noise. This is not surprising as for a conductive electrode the noise is inversely proportional to the square root of the cross-sectional contact area of the electrodes [15]. It is also noteworthy that the K/f noise contribution from

Table 2. Electrode voltage noise variations across all subjects

	Electrode no.	1	2	3	4	5	6
Male 1	fcv (Hz)	17	8	6.9	13.2	8.8	14
	C (nV/√Hz)	3	3.9	3.5	3.3	3.1	4.2
	K (nV√Hz)	61.4	12.7	11	32	12.5	16
Male 2	fcv (Hz)	13	5.9	8	9	10	12
	C (nV/√Hz)	1.4	4.5	3.4	3.1	3.3	2.9
	K (nV√Hz)	21	14	12	13	12.5	12
Male 3	fcv (Hz)	7	9	10	6	9.5	11
	C (nV/√Hz)	4.6	3.4	4.6	4.8	4.3	3.3
	K (nV√Hz)	14	12	13	11	14	12
Female 1	fcv (Hz)	9	9.3	16	10	7.8	7.8
	C (nV/√Hz)	4.2	4.8	4.3	4.9	4.5	4.3
	K (nV√Hz)	24.3	13	44.1	13	12.5	12
Female 2	fcv (Hz)	28	21	17	24	20	19
	C (nV/√Hz)	1.8	1.7	1.9	2.3	2.1	2.6
	K (nV√Hz)	80	46	45	47	50	44
Female 3	fcv (Hz)	30	8.3	8.5	7	9	12
	C (nV/√Hz)	1.1	5	4.4	5.6	4.5	3.2
	K (nV√Hz)	55	13	15	8	13	15

the gelled adhesive electrode is higher than most of the dry electrodes investigated in this study. This is particularly the case for male 1, female 2 and female 3 subjects and is due to the fact that these subjects had either body hair on the arm or very dry skin. It can also be seen that for similar measurement circumstances, female subjects often, but not always, exhibited higher values of white noise than male subjects. This is thought to be due to the fact that female subjects tend to exhibit higher values of skin-electrode impedance [20] and consequently higher values of intrinsic noise. The relationship between electrode noise and source impedance is the underlying mechanism and requires further investigation. The input noise currents of the op-amps used for biopotential measurements are generally considered to generate significant noise voltages when flowing through the impedance of the electrodes. The noise voltages can be significantly higher for the un-gelled fabric-based electrodes than for traditional gelled electrodes due to their higher impedance [20]. Therefore, while more noise may be present when using dry electrodes, the noise cannot be assumed to be generated intrinsically by the electrode itself, but rather by the input noise current of the amplifier flowing through the electrode as a source impedance. Many of the commercially available very low-power complementary metal-oxide-semiconductor (CMOS) op-amps popular in bio-amplifiers used in portable ambulatory recording applications have high-input noise currents as well as significant white noise voltages within the signal bandwidth.

Conclusion: The skin-electrode interface noise performance has been characterised for several un-gelled, textile-based electrodes and contrasted with a dry conductive rubber and a gelled adhesive electrode. The noise voltage generated by the electrodes themselves is found to be lower than that introduced by the amplifier. The high source impedance of the textile electrodes is seen to contribute to the noise generated at the input terminal of the amplifier due to input noise current of the latter. However, when an amplifier with low-input noise current is used, the evidence does not suggest that un-gelled fabric-based electrodes are intrinsically more noisy than others. This suggests that the design and development of wearable garments with inbuilt fabric-based electrodes for recording ECG and other biomedical signals should not be discouraged by the claim that such electrodes are intrinsically noisy.

Acknowledgments: This work was supported in part by a small stipend from Trinity College Dublin, Ireland.

Received: *18 September 2020* Accepted: *6 November 2020*
doi: 10.1049/ell2.12063

References

1 Fernandez, M., Pallas-Areny, R.: Electrode contact noise in surface biopotential measurements. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Paris, France, pp. 123–124 (1992)
2 Geddes, L.A.: *Electrodes and the Measurement of Bioelectric Events*. Wiley-Interscience, New York (1972)
3 Geddes, L.A., Baker, L.E.: *Principles of Applied Biomedical Instrumentation* (3rd ed.). Wiley, New York (1989)
4 Tasneem, N.T., et al.: A low-power on-chip ECG monitoring system based on MWCNT/PDMS dry electrodes. *IEEE Sens. J.* 1–1 (2020)
5 Asadi, S., et al.: Graphene elastomer electrodes for medical sensing applications: Combining high sensitivity, low noise and excellent skin compatibility enabling continuous medical monitoring. *IEEE Sens. J.* 1–1 (2020)
6 Rajanna, R.R., et al.: Performance evaluation of woven conductive dry textile electrodes for continuous ECG signals acquisition. *IEEE Sens. J.* **20**(3), 1573–1581 (Feb. 2020)
7 Gonzales, L., et al.: Textile sensor system for electrocardiogram monitoring. In: 2015 IEEE Virtual Conference on Applications of Commercial Sensors (VCACS), Raleigh, NC, USA, pp. 1–4 (Mar. 2015)
8 Cho, G., et al.: Performance evaluation of textile-based electrodes and motion sensors for smart clothing. *IEEE Sens. J.* **11**(12), 3183–3193 (2011)
9 Bu, Y., Hassan, M.F.U., Lai, D.: The embedding of flexible conductive silver-coated electrodes into ECG monitoring garment for minimizing motion artefacts. *IEEE Sens. J.* 1–1 (2020)

10 An, X., Stylios, G.K.: A hybrid textile electrode for electrocardiogram (ECG) measurement and motion tracking. *Materials (Basel)* **11**(10), 1887 (2018)
11 Arquilla, K., Webb, A., Anderson, A.: Textile electrocardiogram (ECG) electrodes for wearable health monitoring. *Sensors* **20**(4), 1013 (2020)
12 Puurtinen, M.M., et al.: Measurement of noise and impedance of dry and wet textile electrodes, and textile electrodes with hydrogel. In: 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, New York, NY, pp. 6012–6015 (2006)
13 An, X., Tangsirinaruenart, O., Stylios, G.K.: Investigating the performance of dry textile electrodes for wearable end-uses. *J. Text. Inst.* **110**, 151–158 (2019)
14 Meziane, N., et al.: Dry electrodes for electrocardiography. *Physiol. Meas.* **34**(9), R47–R69 (2013)
15 Huigen, E., Peper, A., Grimbergen, C.A.: Investigation into the origin of the noise of surface electrodes. *Med. Biol. Eng. Comput.* **40**(3), 332–338 (2002)
16 Novakov, E.: Evaluation of the electrode-amplifier noise in high resolution biological signal acquisition. In: Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. "Magnificent Milestones and Emerging Opportunities in Medical Engineering" (Cat. No.97CH36136), Chicago, IL, USA, vol. **4**, pp. 1452–1454 (1997)
17 Yacoub, S., et al.: Noise analysis of NASICON ceramic dry electrodes. In: Proceedings of 17th International Conference of the Engineering in Medicine and Biology Society, Montreal, Quebec, Canada, vol. **2**, pp. 1553–1554 (1995)
18 Fernández, M., Pallás-Areny, R.: Ag-AgCl electrode noise in high-resolution ECG measurements. *Biomed. Instrum. Technol.* **34**(2), 125–130 (2000)
19 Liu, X., Demosthenous, A., Donaldson, N.: On the noise performance of Pt electrodes. In: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Lyon, France, pp. 434–436 (2007)
20 Baba, A., Burke, M.: Measurement of the electrical properties of ungelled ECG electrodes. *Int. J. Biol. Biomed. Eng.* **2**(3), 89–97 (2008)

A high-linearity high-gain up-conversion mixer for 24 GHz automotive radar applications

A. Siddique, T. S. Delwar, and J. Y. Ryu✉
Department of Smart Robot Convergence and Application Engineering, Pukyong National University, Busan, South Korea
✉Email: ryujy3189@gmail.com

This paper presents a high conversion gain and highly linearised up-conversion mixer designed in 65 nm CMOS process, which translates intermediate frequency of 2.4 GHz and local oscillator frequency of 21.6–24 GHz in radio frequency. In the proposed mixer, the tranconductance stage consists of a *common source* route with *neutralisation technique* and *dual capacitive cross coupled-common gate* route to achieve high linearity and stability. In addition, it also includes dual negative resistance switches with improved inter-stage matching network to enhance the conversion gain and to decrease the noise figure. The mixer achieves the conversion gain of 4.7 dB, the measured output 1-dB compression point of 0.41 dBm and noise figure of 3.8 dB at 24 GHz. In an automotive radar application, the proposed mixer showed a high conversion gain, an excellent output 1-dB and a low noise figure with power consumption of approximately 5.2 mW at a 1.2 V supply.

Introduction: The ever-increasing wireless networking and radar processing industries continue to permit comprehensive efforts to integrate transceivers into CMOS technologies. Nevertheless, to our knowledge, there is no specifically published CMOS up-conversion mixer at 24 GHz frequency range in the literature, where high linearity and conversion gain are obtained simultaneously with low noise. The authors [1] proposed an up-conversion mixer at 24 GHz with conversion gain (CG) of −1.9 dB and an output 1-dB ($OP_{1\,dB}$) of 0.3 dBm, but this circuit showed a high power consumption of 22.8 mW. The authors [2] also showed a direct up-conversion mixer at 60 GHz, and the proposed circuit accomplishes a high CG of 4.5 dB, but it has poor linearity with a high power consumption 15.1 mW. To achieve low power, high CG and excellent linearity, we propose a low-power 24-GHz up-conversion mixer with admirable CG and high linearity fabricated in standard 65-nm Radio Frequency (RF) CMOS technology. The proposed mixer consists of a dual transconductance route which includes the common source *(CS)* route with *neutralisation technique* and the dual capacitive cross coupled-common gate (*DCCC-CG*) to improve linearity and stability. This circuit also contains the dual pMOS/nMOS switches along with a modified interstage matching to increase the conversion gain and to decrease the noise figure (NF).

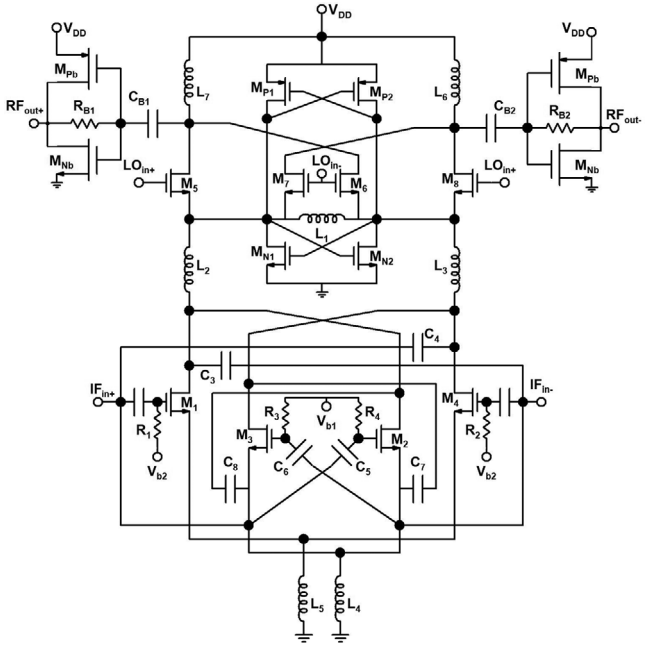


Fig. 1 *Schematic diagram of the proposed CMOS up-conversion mixer*

Table 1. Proposed mixer device component values

Element	Dimension
M_1, M_4	9 um/65 nm
M_2, M_3	27 um/65 nm
$M_5 - M_8$	28 um/65 nm
$M_{p1} - M_{p2}$	2 um/65 nm
$M_{n1} - M_{n2}$	3 um/65 nm
L_1, L_4, L_5	90 pH, 90 pH, 90 pH
$L_2 - L_3$	90 pH
$L_6 - L_7$	150 pH
$C_3 - C_4$	80 fH
$C_5 - C_6$	71 fH

Circuit design: The designed mixer is shown in Figure 1, and Table 1 represents the device component values of the mixer. In the proposed scheme, we have applied the neutralisation technique [3] formed by cross coupling capacitors C_3 and C_4 with *CS* transconductance route of transistors M_1, M_4. This route enriches the gain, noise performance as well as the stability of the mixer. The *DCCC-CG* route consists of transistors M_2-M_3 and dual cross-coupled capacitors C_5-C_6, and C_7, C_8. C_5 and C_6 are cross-coupled between gate terminal and source terminal of M_2-M_3, while C_7 and C_8 are cross-coupled between the drain terminal and the source terminal of M_2-M_3. The *DCCC-CG* also boosts the transconductance and gain, and it decreases NF and the signal strength of second-order harmonic signal.

M_5-M_8 are the switching stage transistors, and transistors M_{n1}, M_{n2} and M_{p1}-M_{p2} are dual nMOS and pMOS switches, respectively. L_1 represents a shunt inductor and L_2 is in series with the transconductance and the switching stage. The transistors M_{n1}-M_{n2} and M_{p1}-M_{p2} inject current to the M_1, M_4 during the on/off transition of switching stage, thus the switching performance of M_5-M_8 transistors is improved. This scheme helps in power consumption saving and conversion gain boosting as compared to single nMOS switch presented in [4]. The total current equation (1) injected the drain of the M_1, and M_2 can be written as

$$i_t = \left(g_{m1}1 + g_{m2}1\left(\frac{C_5 + C_8 - C_{gs2}}{C_5 + C_8 + C_{gs2}} + 1\right)\right)V_{\text{if}} + \left(g_{m1}2 + g_{m2}2\left(\frac{C_5 + C_8 - C_{gs2}}{C_5 + C_8 + C_{gs2}} + 1\right)^2\right)V_{\text{if}}^2 + \left(g_{m1}3 + g_{m2}3\left(\frac{C_5 + C_8 - C_{gs2}}{C_5 + C_8 + C_{gs2}} + 1\right)^3\right)V_{\text{if}}^3 + \cdots \quad (1)$$

The size of capacitor C_5 is selected in such a way as $C_5 \gg C_{gs2}$ while the capacitor C_7 is assumed to be $(\frac{1}{jwC_8}) \gg (\frac{1}{2g_{m2}})$. The transistor M_1-M_2 in *CS* route and *DCCC-CG* route are biased with the optimised voltages v_{b1} and v_{b2} so that it confirms the operation of M_1 in a weak inversion region and operation of M_2 in the saturation region. Under these biasing conditions, this circuit has optimised transistor sizes of M_1 and M_2, the third-order coefficients $(g_{m1}3 + g_{m2}3(\frac{C_5+C_8-C_{gs2}}{C_5+C_8+C_{gs2}} + 1)^3))V_{\text{if}}^3$ of M_1 and M_2 in Equation (1) have the opposite polarity and obviously cancelled third-order coefficient to provide better linearity.

The combination of inductor L_1-L_2 with parasitic capacitance of the transistors M_5-M_8, and $M_{n1,2}$-$M_{p1,2}$ at sources and drain terminals, respectively, provides an interstage matching, and so this interstage matching provides maximum power transfer and eradicates the leakage current between the transconductance stage and the switching stage. As the inductive padding effects on the circuit, the higher order harmonic decreases due to non-linear parasitic capacitances. The enhanced conversion gain of the designed mixer due to nMOS and pMOS switches can be written in Equation (2).

$$CG = \frac{2}{\pi}\left(\frac{g_{m5,7}}{g_{m5,7} - g_{mp1,2} - g_{mn1,2}}\right)g_{m1,2,3,4}\omega_{RF}L \quad (2)$$

where $g_{m1,2,3,4}$ is the transconductance of the transistors M_1, M_2, M_3 and M_4, and L is the equivalent load inductance. $g_{m5,7}$ represents the transconductance at the source terminals of the transistors M_5 and M_7, and $g_{mn1,2}$, and $g_{mp1,2}$ represent the transconductance of the transistors nMOS and pMOS, respectively. From Equation (2), it is evident that conversion gain can be enhanced by using the nMOS and pMOS switches.

Measurement results: The up-conversion mixer is fabricated in a standard 65-nm RF CMOS process. A chip micrograph of the designed mixer is illustrated in Figure 2. The chip size is 0.83×0.51 mm^2 without including pads. It is powered by 1.2 V supply with a power consumption of 5.2 mW. The measurements were carried out at Local Oscillator (LO) power of 1.5 dBm. Optimal results are achieved at 24 GHz. The calculated isolation for LO-to-RF and LO-to-IF (Intermediate Frequency) is −32.3 dB and −47.7 dB, respectively, as shown in Figure 3. Figure 4 shows the measured return losses of RF, LO and IF ports at 24, 21.6 and 2.4 GHz, respectively. The measured return losses of RF, LO and IF ports are −26.3, −31.4 and −33.2 dB, respectively.

Figure 5 depicts the output power as according to the input power at an IF port. The designed up-conversion mixer showed a high 1-dB compression point ($OP_{1\,dB}$) of 0.41 dBm at 24 GHz. The enhanced transconductance stage using *CS* route with the neutralisation technique and *DCCC-CG* route improves the linearity of the mixer.

The measured CG and NF for the RF frequency are shown in Figure 6. The proposed mixer showed high CG of 4.7 dB and low NF of 3.8 dB at 24 GHz using the nMOS and pMOS switches with enhanced transconductance stage.

The performance summary of the proposed up-conversion mixer is compared in Table 2. As can be seen in Table 2, this circuit showed a high

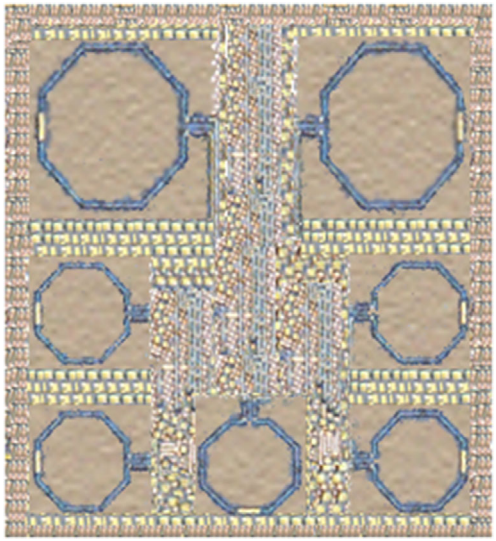


Fig. 2 *A chip micrograph of up-conversion mixer*

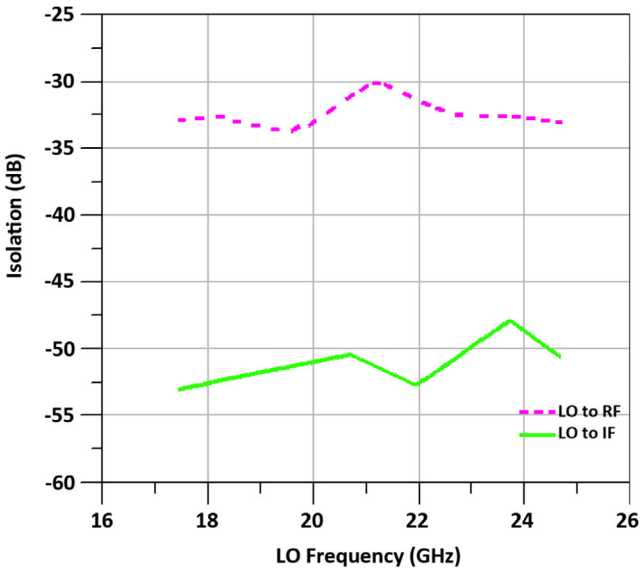


Fig. 3 *Measured isolations between LO-to-RF and LO-to-IF ports*

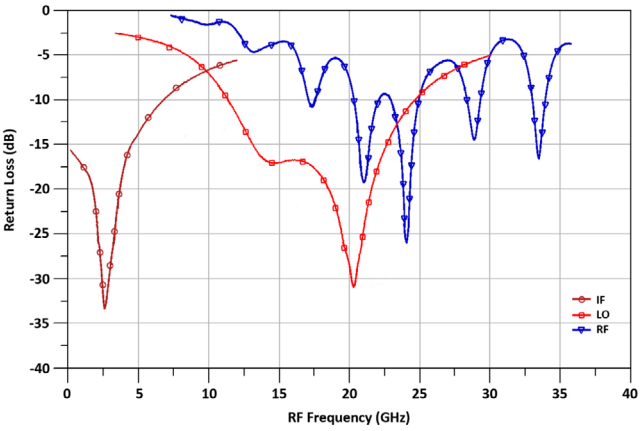


Fig. 4 *Measured return losses of RF, LO, and IF ports*

conversion gain of 4.7 dB, an excellent output 1-dB compression point of 0.41 dBm and a low noise figure of 3.8 dB with power consumption of only 5.2 mW at the 1.2 V supply as compared to recently reported research results [5–8].

Conclusion: This letter introduces a 24-GHz up-conversion mixer in a 65-nm RF CMOS process. To achieve a high conversion gain and an

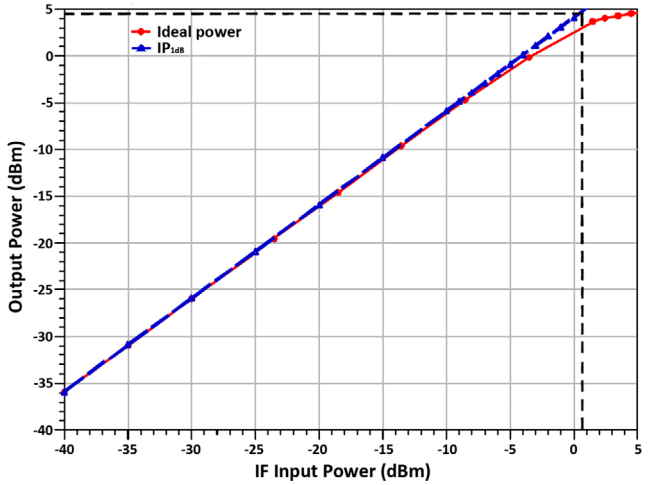


Fig. 5 *RF output power versus IF input power*

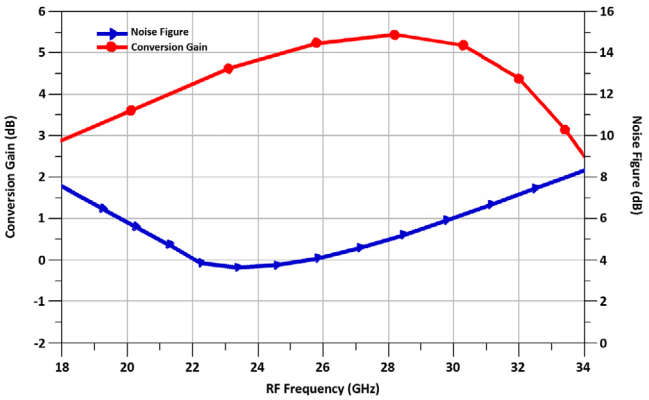


Fig. 6 *Measured conversion gain and noise figure versus RF frequency*

Table 2. Comparison summary for recently reported results

Ref.	[5]	[6]	[7]	[8]	This Work
Freq. (GHz)	27.5–43.5	20–26	19–31	20	24
Tech. (CMOS)	65 nm	90 nm	180 nm	130 nm	65 nm
CG (dB)	−5	2	−0.8	1	4.7
$OP_{1\,dB}$	−3 at 28 Ghz	−12.8	−8.6	−11	0.41
P_{DC} mW	38	11.1	38	8	5.2
Chip Area (mm^2)	0.686	–	1	0.52	0.4233

excellent linearity, an advanced transconductance stage which consists of a *CS* route with the neutralisation technique and *DCCC-CG* route for improvement of linearity, and dual cross-coupled switches with nMOS and pMOS for gain enhancement is proposed. We also proposed the interstage matching technique to eliminate the interstage leakage current and to suppress higher order harmonics. The proposed mixer only consumes 5.2 mW at 1.2 V. This mixer showed a high conversion gain of 4.7 dB and a low NF of 3.8 dB with an excellent $OP_{1\,dB}$ of 0.41 dB at the 24 GHz. We believe that the proposed mixer is suitable for an automotive radar application.

Acknowledgments: This research was supported by Basic Science Research Program through the *National Research Foundation of Korea* (NRF) funded by the Ministry of Education (2018R1D1A1B07043286).

Received: *23 June 2020* Accepted: *26 October 2020*
doi: 10.1049/ell2.12046

References

1 Won, Y.S., Kim, C.H., Lee, S.G.: A 24 GHz highly linear up-conversion mixer in CMOS 0.13 nm technology. *IEEE Microwave Wireless Compon. Lett.* **25**(6), 400–402 (2015)
2 Tsai, T.M., Lin, Y.S.: A 60-GHz double-balanced mixer with negative resistance compensation for direct up-conversion using 90-nm CMOS technology. *Microwave Opt. Technol. Lett.* **55**(3), 536–543 (2013)
3 Heller, T., Cohen, E., Socher, E.: Analysis of cross-coupled common-source cores for W-band LNA design at 28nm CMOS. *2013 IEEE International Conference on Microwaves, Communications, Antennas and Electronic Systems (COMCAS 2013)*, October 2013, pp. 1–5. IEEE, Piscataway, NJ (2013)
4 Tsai, T.M., Lin, Y.S.: A 60-GHz double-balanced mixer with negative resistance compensation for direct up-conversion using 90-nm CMOS technology. *Microwave Opt. Technol. Lett.* **55**(3), 536–543 (2013)
5 Chen, Z., et al.: A 27.5-43.5 GHz high linearity up-conversion CMOS mixer for 5G communication. *2017 IEEE Electrical Design of Advanced Packaging and Systems Symposium (EDAPS)*, December 2017, pp. 1–3. IEEE, Piscataway, NJ (2017)
6 Lai, I.C., Fujishima, M.: An integrated 20-26 GHz CMOS up-conversion mixer with low power consumption. *2006 Proceedings of the 32nd European Solid-State Circuits Conference*, September 2006, pp. 400–403. IEEE, Piscataway, NJ (2006)
7 Comeau, J.P., Cressler, J.D.: A 28-GHz SiGe up-conversion mixer using a series-connected triplet for higher dynamic range and improved IF port return loss. *IEEE J. Solid-State Circuits* **41**(3), 560–565 (2006)
8 Lin, Y.H., et al.: A Ka-band high linearity up-conversion mixer with LO boosting linearization technique. *2018 48th European Microwave Conference (EuMC)*, September 2018, pp. 259–262. IEEE, Piscataway, NJ (2018)

A fast learning approach for autonomous navigation using a deep reinforcement learning method

Muhammad Mudassir Ejaz,[1] Tong Boon Tang,[2] and Cheng-Kai Lu[2,✉]

[1]*Department of Electrical and Electronic Engineering, Universiti Teknologi PETRONAS, Seri Iskandar, Malaysia*
[2]*Institute of Health and Analytics, Universiti Teknologi PETRONAS, Seri Iskandar, Malaysia*

✉E-mail: chengkai.lu@utp.edu.my

Deep reinforcement learning-based methods employ an ample amount of computational power that affects the learning process. This paper proposes a novel approach to speed up the training process and improve the performance of autonomous navigation for a tracked robot. The proposed model named "layer normalization dueling double deep Q-network" has been trained in a virtual environment and then implemented it to a tracked robot for testing in a real-world scenario. Depth images have been used instead of RGB images to preserve the temporal information. Features are extracted using convolutional neural networks, and actions are derived using the dueling double deep Q-network. The input data has been normalized before each convolutional layer, which reduces the covariate shift by 69%. This end-to-end network architecture of the proposed model provides stability to the network, relieves the burden of computational cost, and converges in much less number of episodes. Compared with three Q-variant models, the proposed model demonstrates outstanding performance in terms of episodic reward and convergence rate. The proposed model took 12.8% fewer episodes for training compared to other models.

Introduction: Safe navigation plays a vital role in the autonomous mobile system. Autonomous navigating is the ability to perceive the dynamics of the environment and steer around without human intervention. In the past several years, navigation achieved by laser sensors [1], but due to lower sensing capability and inability to generalize the 3D environment,

they are not used often. Currently, the research tends to focus on vision sensors rather than on laser sensors. Vision sensors are rich in providing more information about the surroundings and generalize the environment better. In the past, several methods proposed by vision sensors for safe navigation [2]. However, they were handcrafted and required more time for training.

After the advancement in deep learning, the capability of autonomous robots has increased where they adapt and learn the problem more smartly [3, 4]. In recent years, autonomous visual navigation has gained popularity using machine learning methods. Supervised learning [5, 6] and deep reinforcement learning (DRL) [7] are the main concepts of machine learning widely proposed in the literature. However, the main limitation of using a supervised learning method is making a labelled dataset. A large dataset is needed to train an agent for a particular environment. Generally, DRL-based methods do not require explicit knowledge before training, and an agent learns from its behaviour using reward function. Deep Q-network (DQN) was the first DRL algorithm implemented by Mnih et al. [8], on Atari games, where actions were performed using a raw pixel. This approach has opened a new era of using DRL-based methods in many other applications, such as robotics [9], education [10], health [11], business [12], and others.

DRL-based methods have a significant covariate shift because of deploying deep neural networks as function approximators. In this paper, we proposed a novel approach that highlighted this issue. We named the proposed model as "layer normalization dueling double deep Q-network (LND-DDQN)", which accelerates a training process by reducing the covariate shift between the convolution layers. We implemented this method on a tracked robot for autonomous navigation by using depth images. For feature extraction, convolutional neural network was applied on four consecutive depth images. Depth images were used instead of RGB images because of the visual fidelity exhibited by the depth image are much better in the simulation environment due to their textureless quality [13]. Actions were derived by dueling double deep Q-network (D3QN). Before each of the convolution layers, layer normalization was performed that reduced 69% co-variance at the end of the last layer. The main contribution of the paper is summarized as follows:

- A novel end-to-end network architecture that accelerates the training process of a deep reinforcement learning-based method.

Method: An autonomous navigation problem has been formulated with Markov decision process (MDP) [14], which contains a set of tuples Y=(S, A, R, P, γ). S represents state space, A is the action space, R as an immediate reward, P is the transition probability, and γ is the discount factor. State space contained the depth images information, while action space has both linear and angular actions (0.4 m/s, 0.2 m/s, $\pi/12$, $\pi/6$, 0.0, $-\pi/6$, and $-\pi/12$). The following equation has been meticulously designed as the reward function, which included both linear and angular velocities:

$$r(s_t, a_t) = \begin{cases} -10, & \text{if collision} \\ a_1 \upsilon^2 \cos(a_2 \upsilon\omega) - a_3, & \text{otherwise.} \end{cases} \quad (1)$$

The value of constants a_1, a_2, and a_3 have been set as 2, 2, and 0.1, respectively. ω represents the linear velocity, and υ is the angular velocity. Reinforcement learning is a closed-loop learning mechanism, where an agent acts in a given state, receives a reward in return, and moves to the next state. For better learning, a penalty is also given to the agent if the selective action leads to the collision. We also included –10 as a penalty in our reward function assigned to the robot when it collides before 500 steps. Otherwise, it received the maximum reward calculated by Equation (1).

In this letter, we modified the network architecture of dueling double deep Q-network (D3QN) [15] by incorporating the layer normalization before each of the convolution layers, as shown in Figure 1. Layer normalization calculates each layer of the input data's moments without adding any dependencies within the network [16]. It calculates the statistics as follows for all hidden units in the same layer:

$$\mu^m = \frac{1}{K}\sum_{i=1}^{K} x_i^m, \qquad \sigma^m = \sqrt{\frac{1}{K}\sum_{i=1}^{K}(x_i^m - \mu^m)^2} \quad (2)$$

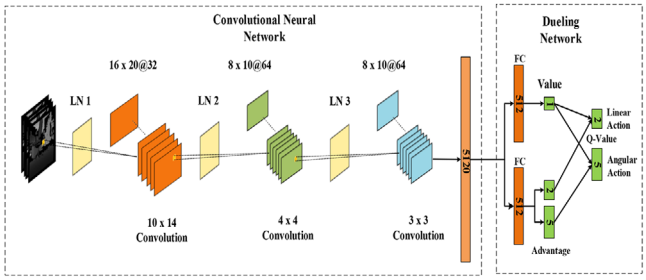


Fig. 1 The network architecture of LND-DDQN. Features were extracted from four consecutive depth image using three convolutional layers. Before each of the convolutional layers, layer normalization was applied that represents by LN to reduce the co-variate shift. After feature extraction, it is fed to a dueling neural network to compute the Q-values

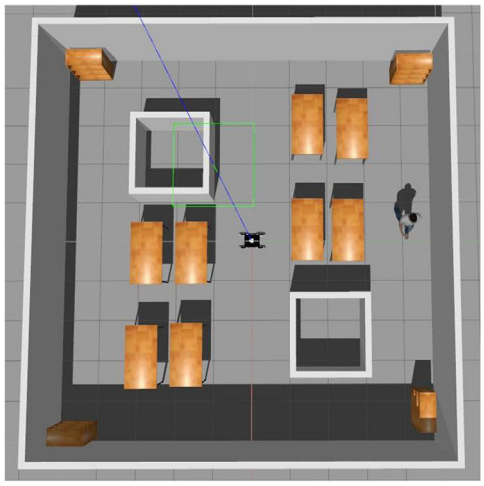


Fig. 2 A training environment designed in a Gazebo simulator includes obstacles and a walking person

where K represents the number of hidden layers, x is summed input vector, μ and σ are the mean and variance, respectively.

In forward-feed neural network, the output of the first layer becomes the input of the second layer, which increases the co-variance of that layer, notably when the ReLU activation function is used. In the proposed model, three convolution layers formed the convolutional neural network to extract features. We found that the co-variance reduced after normalization in all three convolutional layers as 32%, 55%, and 69%, respectively. This reduction in the co-variance not only accelerates the training speed but also reduces the computational cost. The extracted features developed a flatten array with a size of 5120. The flatten array passed to the dueling network to compute the Q-values. The flatten vector split into two fully connected layers, where one computed the state value and others calculated the action advantage. Again, these two layers combined to calculate the Q-values by the following equation:

$$Q_i(s, a) = LeakyReLU(V(s) + A_i(s, a) - \frac{1}{N}\sum_{a} A_i(s, a')) \quad (3)$$

where N represents the number of actions.

Experiments: A tracked robot is used for training in a Gazebo simulator [17]. A 10 × 10 m training environment, as shown in Figure 2, is designed, which includes obstacles and a walking person. The goal of a tracked robot is to avoid the obstacles and navigate autonomously. The hyper-parameters such as learning rate and batch size are set as 0.0001 and 64. However, the training performed for 0.5M steps using Tensorflow [18]. One episode composed of 500 steps. We evaluate the results in terms of episodic reward, as illustrated in Figure 3. During the training, online and target network functions parallel to each other were

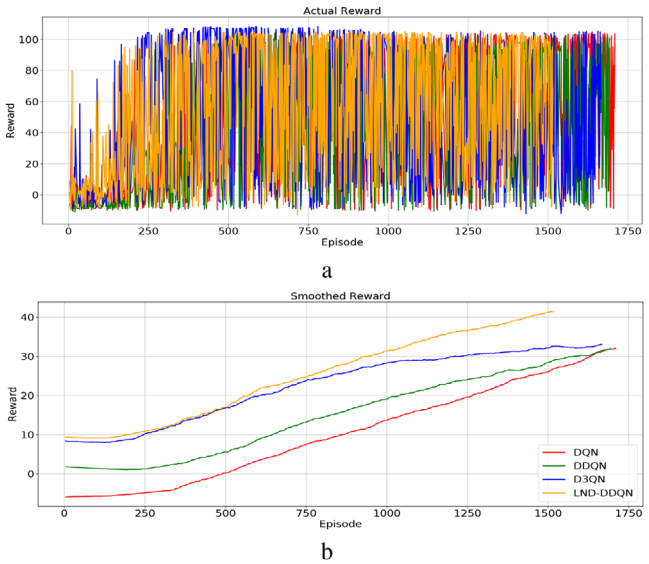


Fig. 3 Rewards received by all the models in the training environment: (a) actual reward and (b) smoothed reward

used to compute the Q-values and the target value. The target value is defined as

$$y_i = \begin{cases} r_t, & \text{if episode ends at step } t{+}1 \\ r_t + \gamma \mathrm{Q}(s_{t+1}, & \\ \mathrm{argmaxQ}(s_{t+1}, a_t + 1; \theta); \theta^-), & \text{otherwise.} \end{cases} \quad (4)$$

We used the mean squared error (MSE) as a loss function with Adam optimizer. The network parameters of an online network were updated using the back-propagation method, while the target network, which is a duplication of an online network, was updated by some time interval. The ϵ-greedily method used for exploration, where the initial and final epsilon values are set as 0.1 and 0.001, respectively.

The results of the proposed model, LND-DDQN, are compared with three Q-variant models, such as DQN, DDQN [19], and D3QN. Figure 3a shows the episodic reward obtained by each model, and episode reward after smoothing is depicted in Figure 3b. A smoothing is performed using the exponential moving average technique, where the smoothing factor (α) is set as 0.999. Figure 3b shows that the DQN model leads to poor performance from the starting, whereas the DDQN algorithm performed slightly better than DQN. However, at the end of the training, the reward value is similar. In contrast, the performance of D3QN is better than the other two models; nevertheless, it results in lower episodic rewards than the proposed model. Generally, the proposed model performed outstandingly throughout the training. Furthermore, the proposed model's total episodes were around 1500, whereas other models took more episodes to complete the training. In comparison, the suggested model took 12.8% fewer episodes for training compared to the other models. Layer normalization speeds up the training and converges in much less number of episodes. It is also worth pointing that the average time taken by the proposed model to steer a tracked robot from the input depth images was merely 0.002 s. It proves that the suggested model takes less computational cost.

After training in a simulation environment, we tested its feasibility in a real tracked robot. The tracked robot is equipped with two small PCs with Intel Real-sense camera, as shown in Figure 4. The trained model has been used to derive the robot in a real-world scenario. It is worth noting that no fine-tuning was performed during the experiments. Figure 5 shows the experimental scene where small boxes have been placed at random positions. The goal of a robot is to avoid obstacles and navigate autonomously.

The paths shown on the left and right side of Figure 5 are the suggested and actual paths in a 2D plane, respectively. The blue and green boxes represent the obstacles (cardboard box, basket, and table), and the star represents the chair. The small black box shows the robot position. A tracked robot starts moving from the middle and goes straight until it detects the obstacles nearby. It changes its orientation and takes a

Fig. 4 Tracked robot equipped with Intel Real-sense depth camera and Lidar sensor

Fig. 5 A real-world scenario for evaluation of a trained model in an indoor environment. The boxes are represented by blue and green rectangles, while a star represents a chair. The robot's starting position depicts a small black box: (a) suggested path and (b) actual path

left turn and goes forward. By following the control commands given by the proposed model, the tracked robot succeeds in performing safe autonomous navigation in a real-world scenario. It shows that our proposed model can adapt to real-world uncertainties after being trained in a simple simulation environment.

Conclusion: In this paper, an end-to-end network architecture is introduced that accelerates the training process of a deep reinforcement learning-based method. An autonomous visual navigation problem is addressed where actions are derived using the depth images. Four consecutive depth images passed to convolution neural networks for feature extraction and layer normalization applied before each of the convolution layers to reduce the co-variance. In particular, it reduced in each layer and reduced up to 69% in the last layer. This reduction in the co-variance accelerates the learning process, making the proposed model take 12.8% fewer episodes for training than the other models. The proposed model performed outstandingly among three Q-variant models (namely, DQN, DDQN, and D3QN) in terms of episodic reward. It is worth noting that the proposed model has been trained in a simple virtual environment, readily transferable, and able to deal with the uncertainties of the real world directly without any fine-tuning. Further research is required to develop exploration strategies for better performance.

Received: *10 July 2020* Accepted: *20 October 2020*
doi: 10.1049/ell2.12057

References

1 Sadeghi, F., Levine, S.: Cad2rl: Real single-image flight without a single real image in Proc. Robotics: Science and Syst., Massachusetts, USA, (July 2017). https://doi.org/10.15607/RSS.2017.XIII.034
2 Bonin-Font, F., Ortiz, A., Oliver, G.: Visual navigation for mobile robots: A survey. *J. Intell. Rob. Syst.* **53**(3), 263–296 (2008)
3 Wu, K., et al.: Tdpp-net: achieving three-dimensional path planning via a deep neural network architecture. *Neurocomputing* **357**, 151–162 (2019)
4 Codevilla, F. et al.: End-to-end driving via conditional imitation learning. *2018 IEEE International Conference on Robotics and Automation*, pp. 1–9, IEEE, Piscataway, NJ (2009)
5 Giusti, A. et al.: A machine learning approach to visual perception of forest trails for mobile robots. *IEEE Rob. Autom. Lett.* **1**(2), 661–667 (2016)
6 Wortsman, M. et al.: Learning to learn how to learn: self-adaptive visual navigation using meta-learning. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6750–6759. IEEE, Piscataway, NJ (2019)
7 Lobos-Tsunekawa, K., et al.: Visual navigation for biped humanoid robots using deep reinforcement learning. *IEEE Rob. Autom. Lett.* **3**(4), 3247–3254 (2018)
8 Mnih, V. et al.: Playing Atari with deep reinforcement learning. Preprint, arXiv:1312.5602 (2013)
9 Raffin, A., Stulp, F.: Generalized state-dependent exploration for deep reinforcement learning in robotics. Preprint arXiv: 2005.05719 (2014)
10 Jiang, Z., Liang, J.: Cryptocurrency portfolio management with deep reinforcement learning. In: *2017 Intelligent Systems Conference*, pp. 905–913 (2017)
11 Shen, Z. et al.: Learning to scan: A deep reinforcement learning approach for personalized scanning in CT imaging, Preprint arXiv:2006.02420 (2014)
12 Li, Z. et al.: Deep reinforcement learning based game decision algorithm for digital media education. In: *16th International Computer Conference on Wavelet Active Media Technology and Information Processing*, pp. 139–142 (2019)
13 Tai, L. et al.: Socially compliant navigation through raw depth inputs with generative adversarial imitation learning. *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1111–1117. IEEE, Piscataway, NJ (2018)
14 Sutton, R., Barto, A.: *Introduction to Reinforcement Learning*. MIT Press, A Bradford Book 55, Hayward Street, Cambridge, MA, United States (2018) https://mitpress.mit.edu/books/reinforcement-learning-second-edition
15 Wang, Z. et al.: Dueling network architectures for deep reinforcement learning. *International Conference on Machine Learning*, pp. 1995–2003 (2016)
16 Ba, J., et al.: Layer normalization, Preprint arXiv:1607.06450 (2014)
17 Xiao, J. et al.: Building software system and simulation environment for robocup msl soccer robots based on ros and gazebo. *Studies in Computational Intelligence*, pp. 597–631. Springer International Publishing, Berlin (2017)
18 Smilkov, D. et al.: Tensorflow. js: Machine learning for the web and beyond, Preprint arXiv:1901.05350 (2019)
19 Hasselt, H., et al.: Deep reinforcement learning with double q-learning. *13th AAAI Conference Artificial Intelligence*, Menlo Park, CA 2016

PotNet: Pothole detection for autonomous vehicle system using convolutional neural network

Deepak Kumar Dewangan✉ and Satya Prakash Sahu
Department of Information Technology, National Institute of Technology, Raipur, India
✉E-mail: dkdewangan.phd2018.it@nitrr.ac.in

Advancement in vision-based techniques has enabled the autonomous vehicle system (AVS) to understand the driving scene in depth. The capability of autonomous vehicle system to understand the scene, and detecting the specific object depends on the strong feature representation of such objects. However, pothole objects are difficult to identify due to their non-uniform structure in challenging, and dynamic road environments. Existing approaches have shown limited performance for the precise detection of potholes. The study on the detection of potholes, and intelligent driving behaviour of autonomous vehicle system is little explored in existing articles. Hence, here, an improved prototype model, which is not only truly capable of detecting the potholes but also shows its intelligent driving behaviour when any pothole is detected, is proposed. The prototype is developed using a convolutional neural network with a vision camera to explore, and validates the potential, and autonomy of its driving behaviour in the prepared road environment. The experimental analysis of the proposed model on various performance measures have obtained accuracy, sensitivity, and F-measure of 99.02%, 99.03%, and 98.33%, respectively, which are comparable with the available state-of-art techniques.

Introduction: The recent advancements in artificial intelligence techniques has opened the gateway towards the design and development of autonomous vehicle system (AVS) in automobile industries. However, real road environment requires strong safety assurance for human is still a challenging issue for these industries. World health organization (WHO) has marked high death rate in road accidents [1] and reason behind such deaths are random [2] which are mostly caused by casual driving and incapability to understand the driving scene. Driving scene understanding capability is being explored by deep learning (DL) and computer vision (CV) approach that mainly consists of feature extraction, classification, detection and tracking [3–5]. Various significant phases of an AVS such as detection of road, lane, pedestrian, vehicle, etc. have been assisted with the DL learning and CV approaches.

Likewise, mishandling with pothole objects can damage the mechanic assembly of the vehicle and may lead to accidents. Aparna et al. has utilized thermal camera to obtain the distinct characteristics of pothole objects especially in the night, but offered restricted and expensive performance for an AVS in day light [6]. Nienaber et al. used optical camera to identify pothole by eliminating non-road regions and targeted road portion where the potholes were present [7]. Ryu et al. have used three major phases in pothole detection [8]. In segmentation phase, histogram shape-based thresholding was applied to transform input into binary image. Then, modified histogram shape-based thresholding (HST) was applied to separate enhanced pothole candidate region. In final phase, ordered histogram intersection (OHI) computed the degree of resemblance between regions to identify potholes. Similarly, work in [9] includes the road surface structure analysis where smartphone accelerometers were used to determine the steadiness of driving vehicle on the road. Vehicle movements on potholes represents the fluctuations until the stability of vehicle occurs on the smooth road. These event sequences were recorded and modelled as multivariate time series to identify the different road objects based on the pattern of stability. Here, DL contains convolutional neural network (CNN), long short-term memory (LSTM) and reservoir computing (RC) to distinguish potholes and non-pothole along with other road objects. To minimize the computation, ROI were specified and trained to assess selected frames with deep CNN model. Inception v4, Inception ResNet v2 and MobileNet v1 were employed for the detection of potholes. LS-SVM and ANN are methods that were also used for pothole detection with classification accuracy rate of roughly 89% [10]. The existing practices in pothole detection have limited performance and can be extended a bit by designing much simpler and low-cost architecture that will be optimized for both software and hardware. Besides, AVS behaviour about taking intelligent decision is still not explored.

To overcome the challenges and issues revealed in the literature, a vision-based CNN model with an embedded vehicle prototype towards the design of AVS for pothole detection has been proposed to show its driving intelligence against detected potholes to avoid any mishap and is illustrated in Figure 1. The major contributions for the proposed work can be summarized as:

1. A low-cost computer vision and CNN-based model has been developed for the precise detection of potholes to overcome the limitations of various expensive sensors.
2. Proposed model utilizes monocular camera unit and Raspberry Pi computing device (vehicle prototype) to confirm the concept developed for pothole detection in prepared real-time scenario.

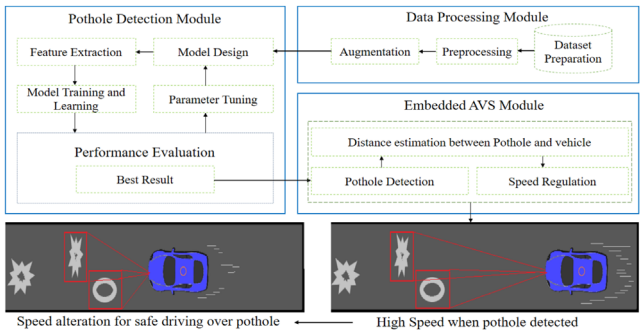


Fig. 1 *Proposed workflow for pothole detection module*

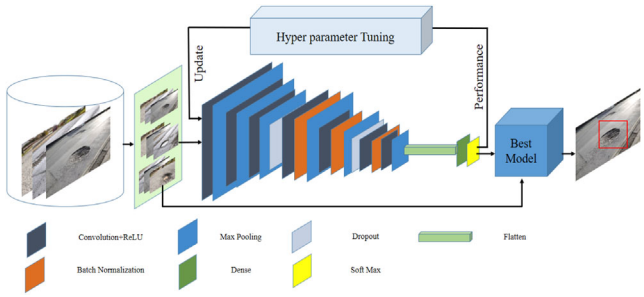


Fig. 2 *Pothole detection CNN architecture*

3. Optimum hardware resources are recognized to detect pothole objects and their operational conformity towards the intelligence of vehicle prototype is assured.

Proposed approach: Initially, the basic setup and CNN model are developed to learn the more intrinsic features of potholes and then the learned features are processed to determine the intelligent behaviour of vehicle prototype against the detected potholes. For this, various stages have been considered. A potential CNN model requires significant features of the target object through a large amount of data. To assist this, a relevant dataset is prepared by capturing the prepared pothole setup through smart phone and Pi camera unit which is mounted on the vehicle prototype. In this direction, images from different orientations and perspectives are considered and dataset size is increased from 783 to 3915 images through flipping, blurring, warping and rotation techniques.

The proposed CNN architecture obtains image as an input frame and delivers pothole object enclosed within a bounding box. For this, combination of convolutional layers has been applied to extract the significant features of potholes. Initially, a set of filters of size 7×7, 5×5 and 3×3 has been considered; and after intensive experiments 5×5 and 3×3 are found appropriate to acquire the necessary features for pothole objects. Layered approach enables the architecture to create feature map that is updated in every combination of convolutional pooling and fully connected layers. These feature maps represent the characteristics of intended objects. However, the modification in any of the layers and its parameters reflect the performance and hence behaviour of architecture is recorded and experimented thoroughly to extract the best version of the network. The schematic can be visualized from Figure 2 where distinct layers are represented having specific task according to the architectural configuration. Similarly, summarized details of operational layer are represented in Table 1 and their configurations are revealed and are represented in Table 2.

Training procedure includes adaptive moment estimation (Adam) optimizer that uses exponential weighted moving averaging to find the momentum and the second moment of the gradient. Similarly, different variations of learning rate (LR) has been experimented and 0.001 is found effective with batch size of 64 for the given input frames. Because, if it is very large, optimization diverges, and if it is very small, it takes extended time to train or end up with a trivial outcome. Moreover, activation function ReLU makes model training computationally effective because, the gradient in the positive interval is always 1. Hence, parameters are not accurately initialized, the sigmoid function might get a gradient of nearly

Table 1. Operational layers

Layers	Operational expression
Covolutional	$C(I*F)_{x*y} = \sum_{a=1}^{n_H} \sum_{b=1}^{n_W} \sum_{c=1}^{n_C} F_{a,b,c} I_{x+a-1,y+b-1,c}$
Pooling	$P_{Max}(C) = Max_T Activation_{score}$
Dense	$F_{Activation}(W * X + B)$
BatchNorm	$BN(Layer)_i = \gamma \odot (Layer_i - \hat{\mu})/\hat{\sigma} + \beta$
Learning Rate	$w_{a+1} = w_a + (\alpha * \delta * InputImage)$

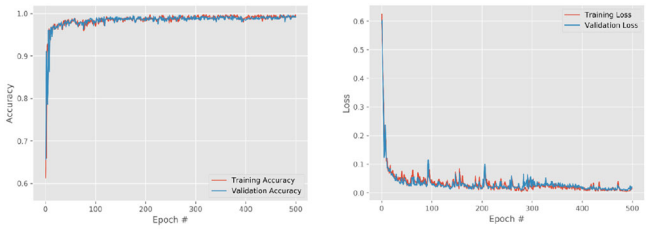
F = Filter, I = Input image, x,y = Resulting index, W = Weights of neuron for current and previous layer, X = Number of neurons (Previous layers), B = Bias, $\hat{\mu}$ = Sample mean for mini batch, $\hat{\sigma}$ = Standard deviation for mini batchmini batch gradient, γ and β = Scaling Coefficient, α = Learning rate, δ = Error.

Table 2. Configuration of layered architecture for pothole detection

Layer	Operational layer	Output shape	Filter size	Weights
0	Input	64, 64, 3	—	0
1	Convolution 2D	64, 64, 250	5, 5	19,000
2	Activation	64, 64, 250	—	0
3	Max pooling	32, 32, 250	2, 2	0
4	Convolution 2D	32, 32, 250	5, 5	1,562,750
5	Activation	32, 32, 250	—	0
6	Max pooling	16, 16, 250	2, 2	0
7	Convolution 2D	16, 16, 200	5, 5	1,250,200
8	Activation	16, 16, 200	—	0
9	Max pooling	8, 8, 200	2, 2	0
10	Dropout	8, 8, 200	—	0
11	Convolution 2D	8, 8, 200	3, 3	360,200
12	Activation	8, 8, 200	—	0
13	Batch normalization	8, 8, 200	—	800
14	Max pooling	4, 4, 200	2, 2	0
15	Convolution 2D	4, 4, 150	3, 3	270150
16	Activation	4, 4, 100	—	0
17	Batch normalization	4, 4, 150	—	600
18	Max pooling	2, 2, 150	2, 2	0
19	Dropout	2, 2, 150	—	0
20	Convolution 2D	2, 2, 150	3, 3	202,650
21	Activation	2, 2, 150	—	0
22	Batch normalization	2, 2, 150	—	600
23	Convolution 2D	2, 2, 100	3, 3	135,100
24	Activation	2, 2, 100	—	0
25	Max pooling	1, 1, 100	2, 2	0
26	Flatten	100	—	0
27	Dense	2	—	202
28	Activation	2	—	0

Total Parameters: 3,802,252.

0, and the model cannot be efficiently trained. Activation score in transitional layers may vary from the input to the output, over time due to the updation in the model's parameters and through nodes in the same layer. This accumulation in the activation distribution can delay the convergence of the network that too may lead to alterations in the LRs for every layer which is computationally expensive. Thus, normalizing the



a b

Fig. 3 *Training performance of pothole detection model*

Table 3. Quantitative evaluation of proposed model

Metrics	Accuracy	Specificity	Sensitivity	Precision	FPR	FScore
Score	99.02%	99.01%	99.03%	98.03%	0.80	98.33%

Table 4. Comparison with state-of-art methods

Method	Nienaber [7]	Aparna [6]	Ruy [8]	Varona [9]	Hoang [10]	Proposed
A	—	97.08%	73.50%	98.00%	89.00%	99.02%
P	81.80%	—	80.00%	–	—	99.03%
R	74.40%	—	73.30%	–	—	99.03%

A = Accuracy, P = Precision, R = Recall.

activations of every node through batch normalization has been applied to handle with the mentioned problem and it assists the model to converge at the early stages. Padding and stride are kept at unit movement and Softmax function is utilized in final layer.

Experimental results and discussion: Behaviour of pothole detection model can be visualized and assessed through training and validation learning graph represented in Figure 3. These learning curves which are gradually increasing at each step describes the model performance for which more than 3 million parameters are configured and trained. Having early convergence in accuracy for both curves indicate the potential of model and stability in the network portrays the optimal performance as shown in Figure 3a. Obviously, the chances of overfitting is completely void during their learning and likely to perform optimum against the test cases. Similarly, Figure 3b describes the learning loss which is initially unstable and started from 0.6 and soon it gets stabilized and reaches nearly 0. From these observations, both accuracy and loss graph represents the optimum behaviour of learning significant pothole features.

Besides, quantitative performance analysis [11] is also carried out that matches with the qualitative behaviour as discussed and shown in Figure 3. The proposed model is also tested using the public dataset named as 'Nienaber Potholes' accessed from Kaggle platform [7]. The obtained scores are encouraging and hence this model is considered to be integrated with vehicle prototype that intelligently changes its behaviour at run time. Standard metrics with their numerical score for this evaluation are determined and represented in Table 3. Comparing with the study in [6], which uses thermal camera and CNN-based method, our proposed model is capable of detecting potholes using simpler but effective CNN architecture without using thermal camera thus minimizing the overall working and design cost. Similarly, texture shape and dimension feature that uses image processing technique obtained 73.50% detection accuracy [8]. Proposed model obtained promising scores which outperforms the state-of-art methods and represented in Table 4. Using image processing techniques, 81.80% precision and 74.40% recall score is achieved in pothole detection [7].

Vehicle to pothole distance measurement: Assuming that mounted camera on vehicle is ideally calibrated and focuses on the front road portion. The distance measurement of detected pothole from the vehicle requires its position when vehicle moves closer to it. For this, studies in [12] and [13] has been taken into consideration, but applicability of these state-of-art techniques in distance measurement are somewhat computationally expensive and cross-ratio method which requires location of the reference pixels to be reestablished is memory intensive. Having these limitations and to measure distance effectively, bounding box boundaries that encloses potholes are utilized to measure the actual distance of object from vehicle position.

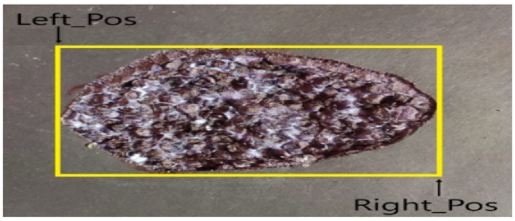


Fig. 4 *Distance measurement setup using bounding box boundary*

For this, dynamic pixel positions of left *Left_Pos* and right *Right_Pos* boundaries are computed and recorded. Outline of assumptions for bounding box boundary is represented in Figure 4. Using this positional information, difference between *Right_Pos* and *Left_Pos* are calculated (let 90 pixels) at a fixed position from the farther location (let *Distance_One* be 100 cm unit). Similarly, the difference between these positions (assuming 40 pixels) are recalculated at different and closer positions (let *Distance_Two* be 60 cm unit). Now, to simplify the distance calculation, linear equations $y = mx + c$ is applied and uses the above calculated difference between *Right_Pos* and *Left_Pos* for distances against closer and farther locations ($100 = 90m + c$ and $60 = 40m + c$). Values of m and c parameters will be used to compute the distance between object and vehicle on every movement (in centimetre) towards the pothole. The moment when vehicle gets closer to pothole object, its visual size gets enlarged and is computed with dynamic change in pixel positions *Right_Pos* and *Left_Pos*. It denotes the real distance which is associated to the focal length of camera in real-time scenario.

Intelligent behaviour of prototype vehicle against detected pothole objects: The proposed prototype model is incorporated in Raspberry pi which intelligently modifies its driving behaviour when pothole object is identified. The scheme of modifying its driving speed depends upon the distance calculation from vehicle to the detected potholes described in section 'Vehicle to pothole distance measurement' in this article. After the distance is calculated from pothole object, Raspberry Pi unit instantly transmits distance information to the Arduino unit. A threshold value (15 cm in the presented scenario) is set when the vehicle goes closer to the detected pothole objects in order to change the driving behaviour either by slowing down the speed or maintaining the constant speed. Upon receiving the distance information, Arduino unit adjusts its pulse width modulation (PWM) by modifying its duty cycle from 100% to 25% for all pair of motors. Reduction in duty cycle corresponds to decrease in vehicle speed and hence chances of hampering the vehicle assembly during high speed is minimized. Slowing down the vehicle's speed when it identifies the object is the early warning mechanism that a vehicle intelligently uses to change its driving. Figure 5a,b represents the change in driving pattern upon receiving the distance information from Raspberry unit.

Conclusion: This paper presents an improved CNN-based pothole detection model which is further utilized to validate the intelligent behaviour of proposed vehicle prototype on driving decision-making. Pothole detection model has undergone through several combinations of CNN layers and after tuning up design parameters in CNN, optimal version of pothole detection model is constructed. From the intensive experiment, both quantitative and qualitative experiments on the public and prepared environment dataset validate the performance of our dynamic pothole detection model and achieves 99.02% of detection accuracy, FScore of 98.33% and 99.03% of sensitivity which outperforms various state-of-art techniques in this domain. Further, this model is integrated with vision-based vehicle prototype to confirm its intelligent driving behaviour against detected potholes which provides low cost and

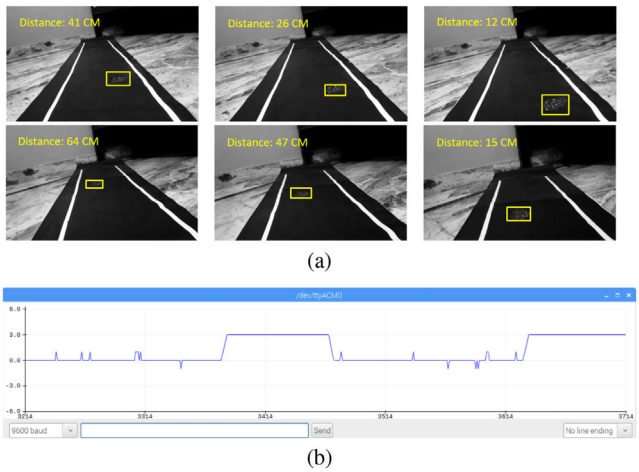


Fig. 5 *Run time performance of AVS. (a) Pothole detection with distance calculation from AVS. (b) Signal behaviour (small unit) in positive and negative directions indicates AVS's driving movement in left and right directions, respectively, while sudden change in signal (0.0–3.0) represents that speed of AVS is reduced to ignore the mechanical hazard during driving over pothole as the distance to pothole is close*

optimized solution towards the design of such AVS. A simple but effective mechanism of distance measurement of pothole from the vehicle is planned and found applicable to provide the early warning to the AVS. In the future work, evolutionary algorithm will be investigated against the pothole detection model and hence utilization of such a model into vehicle is expected to bring the intelligent driving behaviour to an optimized level.

Received: *19 September 2020* Accepted: *9 November 2020*
doi: 10.1049/ell2.12062

References

1 *Road traffic deaths, global health observatory data repository by world health organization*. (2019)
2 Dewangan, D., Sahu, S.: Real time object tracking for intelligent vehicle. *2020 First International Conference on Power, Control and Computing Technologies (ICPC2T)*, pp. 134–138 IEEE (2020)
3 Han, Z., et al.: Design of intelligent road recognition and warning system for vehicles based on binocular vision. *IEEE Access* **6**, 62880–62889 (2018)
4 Huang, L., et al.: Robust inter-vehicle distance estimation method based on monocular vision. *IEEE Access* **7**, 46059–46070 (2019)
5 Junaid, M., et al.: Multi-feature view-based shallow convolutional neural network for road segmentation. *IEEE Access* **8**, 36612–36623 (2020)
6 Bhatia, A.Y., et al.: Convolutional neural networks based potholes detection using thermal imaging. *J. King Saud Univ. -Comput. Inf. Sci.*, no. xxxx, 2019, doi: 10.1016/j.jksuci.2019.02.004
7 Nienaber, S., et al.: A comparison of low-cost monocular vision techniques for pothole distance estimation. In: *2015 IEEE Symposium Series on Computational Intelligence* IEEE, (2015)
8 Ryu, S-K., et al.: Image-based pothole detection system for its service and road management system. *Math. Prob. Eng.* **2015**, 1–10 (2015)
9 Varona, B., et al.: A deep learning approach to automatic road surface monitoring and pothole detection. *Pers. Ubiquitous Comput.*, 2019, doi: 10.1007/s00779-019-01234-z
10 Hoang, N.: An artificial intelligence method for asphalt pavement pothole detection using least squares support vector machine and neural network with steerable filter-based feature extraction. *Adv. Civ. Eng.* **2018**, 1–12 (2018)
11 Nieto-Hidalgo, M., et al.: Two-stage convolutional neural network for ship and spill detection using SLAR images. *IEEE Trans. Geosci. Remote Sens.* **56**(9), 5217–5230 (2018)
12 Huang, L., et al.: Robust inter-vehicle distance estimation method based on monocular vision. *IEEE Access* **7**, 46059–46070 (2019)
13 Danial, J., et al.: Position estimation of moving objects: Practical provable approximation. *IEEE Robot. Autom. Lett.* **4**(2), (1985)

Fusion of semantic and appearance features for loop-closure detection in a dynamic environment

Yan Xu and Jiani Huang✉
School of Electrical and Information Engineering, Tianjin University, Tianjin, People's Republic of China
✉Email: hjn@tju.edu.cn

Loop-closure detection is important in large-scale localisation system. However, it is still difficult in a dynamic environment. An online fast loop-closure detection algorithm based on Deeplabv3 with MobileNetV2 (DpMn2) as network backbone and local difference binary descriptor is proposed, and the algorithm is named as DpMn2-LDB. DpMn2 splits out common dynamic objects of images, and then uses visual geometry group network (VGG16) that is trained on place-centric data to extract global features for nearest neighbour image retrieval. The loop-closure matches are verified based on LDB descriptors and random sample consensus (RANSAC). Experimental results show that the proposed method can obtain a higher recall rate under 100% precision with less execution time per frame on several public datasets compared with other typical or state-of-the-art algorithms.

Introduction: Simultaneous localisation and mapping (SLAM) have attracted significant attention owing to its performance in localising moving agents. Loop-closure detection plays an important role in SLAM, which aims to determine whether the mobile agent has reached previously visited places. However, loop-closure detection is still a challenging problem because two images of the same place taken at different times might change dramatically due to the interference from viewpoints and dynamic objects such as people and cars. Several approaches have been implemented to address this problem. Zhang et al. [1] adopted regional maximum activation of convolutions to represent images. Han et al. [2] utilised log-spectral residual to extract salient regions of images. In addition, common methods are based on offline bag of words (BoW) such as Dorian's bag of word 2 (DBoW2) [3] and gridding place recognition (GPR) [4], which constructs a visual words dictionary offline by clustering feature descriptors extracted from a large-scale image dataset. The descriptors usually applied to BoW are local features such as speeded up robust features (SURF) and binary robust independent elementary features (BRIEF), which perform well in viewpoint changes but may fail to represent high-level features [1]. Thus, the authors in [5] proposed a fast and incremental loop closure detection (FILD) to extract SURF local features for geometrical verification while serving as a complement of the convolutional neural network (ConvNet)-based global features for online nearest neighbour search based on hierarchical navigable small world (HNSW) [6].

Inspired by the methods mentioned above and the structure of FILD, we propose to adopt a lightweight semantic segmentation network Deeplabv3 with MobileNetV2 as network backbone (named as DpMn2) to split out dynamic objects. And then ConvNet that is pre-trained on the place-centric data is used to extract the global feature for nearest neighbour search. The loop-closure matches are verified based on local difference binary [7] (LDB) descriptors matching and random sample consensus (RANSAC). The flowchart of our proposed method is shown in Figure 1.

Proposed algorithm: The proposed algorithm consists of five phases: Separation of dynamic objects, ConvNet-based global feature extraction, nearest neighbour image retrieval, geometrical consistency check and temporal consistency check. A detailed description is given as follows.

(i) *Separation of dynamic objects*: To reduce the impact of dynamic objects, we form a lightweight semantic segmentation network

DpMn2 to split out common dynamic objects from the input image I_i. DpMn2 is trained on the Pascal Visual Object Classes 2012 (Pascal VOC 2012) dataset. Fourteen kinds of objects such as bicycle, person, car, dog are considered dynamic objects and removed from the image I_i. The rest of the image I_i is retained as a static background for feature extraction.

(ii) *ConvNet-based global feature extraction*: For dynamic-object-free image I_i', VGG16 that performs well in appearance changes is used to extract global features. This network is trained on the places365 standard dataset [8], which is specially built for scene recognition and pays more attention to the characteristics of static objects than other popular datasets. The output of the penultimate fully connected layer of the network is considered as image global features and will be used in the following steps.

(iii) *Nearest neighbour image retrieval*: After obtaining the global descriptors of I_i', HNSW is employed to retrieve the nearest neighbour image of I_i'. Since the images used are continuous images taken by mobile agents, adjacent images may have strong similarities. It is easy to mistake adjacent images as search results, which will lead to false-positive loop closure. To avoid this situation, we exclude adjacent images from the search area U_{sa}. The search area U_{sa} is defined as

$$U_{sa} = U_{\text{before}} - U_{fr\times ct} \tag{1}$$

where U_{before} is set of all images before input image, fr is the frame rate of the camera, ct is time constant and $U_{\text{fr}\times\text{ct}}$ is the set of fr × ct frames before input image that is considered as adjacent images. As shown in Figure 1, HNSW is an online graph-based vector search algorithm and image features can be inserted into the graph gradually. The nearest neighbour image I_n retrieved by HNSW graph will be interpreted as the loop-closure candidate of the input image I_i.

(iv) *Geometrical consistency check with LDB descriptor*: Geometrical verification is applied to confirm whether I_i and I_n forms a true-positive loop-closure match. First, LDB local descriptors of I_i and I_n are extracted for descriptors matching. Then, RANSAC is used to remove mismatches and calculate a fundamental matrix. If the fundamental matrix cannot be calculated or the number of inlier points between I_i and I_n is less than threshold δ, the loop-closure match is rejected. The implementation of LDB local descriptors extracting and descriptors matching are described as follows:

a. LDB local descriptors: Keypoints of I_i can be detected by using an oriented fast and rotated brief (ORB) point detector. For each keypoint k_{ij}, we crop an image block B_{ij} with a size of $S\times S$ pixels centred around k_{ij} and then divide the block B_{ij} into M equal-sized grid cells c_{ij}^m ($m = 1,2,\ldots,M$). Moreover, averaged image intensities I_{avg} and image gradients d_x, d_y are calculated for each grid cell c_{ij}^m. Binary codes of keypoint k_{ij} can be calculated by performing a binary test as defined in Equation (2) on any two grid cells c_{ij}^m and c_{ij}^n in block B_{ij}:

$$T(f(m), f(n)) = \begin{cases} 1 & f(m) > f(n) \\ 0 & f(m) <= f(n), \end{cases} m \neq n \tag{2}$$

where f(m) and f(n) represent the value of I_{avg}, d_x and d_y in grid cell c_{ij}^m and c_{ij}^n, respectively. Notably, the number of grid cells M has great influences on LDB. Smaller M enhances the distinctiveness, while larger

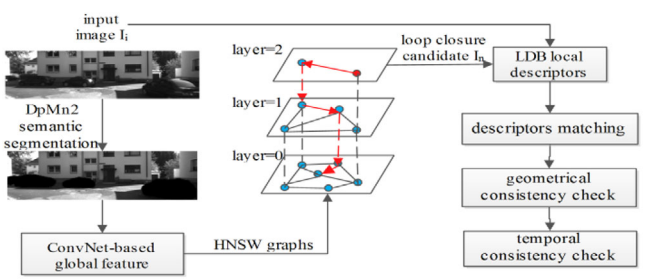


Fig. 1 *Flowchart of our proposed method*

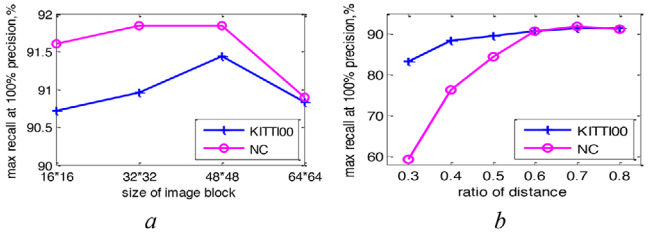


Fig. 2 *The maximum recall rate curves at 100% precision with different parameters on KITTI 00 and new college (NC) datasets. (a) The maximum recall rate curves with different sizes of image block S×S, (b) the maximum recall rate curves with different values of ε_d*

M increases robustness [7]. Thus, taking both distinctiveness and robustness into consideration, we take M as four values {4, 9, 16, 25} to calculate four binary codes and then concatenate them. Finally, a random bit selection method is applied to select 256 bits from the concatenated binary code to form the LDB descriptor of k_{ij}:

(a) Descriptors matching: For LDB descriptors of I_i and I_n, we employ hamming distance to match the descriptors. For d_i^a, which is defined as the descriptor of I_i, d_n^1 and d_n^2 are the two best-matched descriptors found in I_n. If d_i^a, d_n^1 and d_n^2 satisfy Equation (3), d_i^a and d_n^1 will be considered as a pair of good matches:

$$H(d_i^a, d_n^1) < \varepsilon_d \times H(d_i^a, d_n^2) \tag{3}$$

where H (d_i^a, d_n^1) and H (d_i^a, d_n^2) represent the hamming distance between d_i^a and d_n^1, d_n^2, respectively. ε_d is the ratio of distance. The good matches of descriptors between I_i and I_n are then fed into RANSAC to remove mismatches and calculate the fundamental matrix.

(v) *Temporal consistency check*: Finally, the images I_i and I_n will be accepted as a loop-closure match if the aforementioned conditions are satisfied by the two consecutive camera measurements as the method proposed in [9].

Experimental results and discussion: To evaluate the proposed algorithm, we perform experiments on four public datasets KITTI 00 [10], KITTI 05 [10], new college (NC) [11] and city centre (CC) [11] and compare the result with other typical or state-of-the-art methods. These datasets were taken outdoors and contained many dynamic scenarios. The frame rate fr of KITTI 00 and KITTI 05 datasets is 10 fps that is 20 times faster than CC and NC datasets. The ground truths were provided by authors in [9] and [11]. In this letter, we set the time constant ct to 40, the maximum number of connections for each node in HNSW to 48 as recommended in [5]. All algorithms are measured on an Intel(R) Core i7-9700k CPU@3.6 GHz machine with a GeForce GTX 2080Ti GPU card.

To obtain the optimal values of $S\times S$ and ε_d, we conduct experiments with different parameter settings and the results are shown in Figure 2. As shown in Figure 2a, when the size of the image block S increases from 16 to 48, the maximum recall rate rises gradually due to the enhancement of distinctiveness. However, the recall rate goes down when S increases to 64 due to the dimension limit of LDB. In Figure 2b, the maximum recall rate increases gradually and reaches its peak at 0.7 as ε_d increases. Therefore, we set S to 48 and ε_d to 0.7.

Besides, we conduct experiments to evaluate the performances of our method with different network backbones in Deeplabv3 and local feature descriptors. Table 1 shows the maximum recall rate and average running time of methods with different backbones and descriptors. The combination of Deeplabv3 with Xception65 as network backbone (DpXc65) and LDB descriptors obtains the highest recall rates due to DpXc65 owning higher semantic segmentation accuracy than DpMn2. However, the time consumed per frame in DpXc65 is two times longer than that of DpMn2. For models that apply DpMn2, the model with SURF yields similar results as LDB while the memory usage of floating-point format SURF is much higher than that of character format LDB. Experimental results show that our method with DpMn2 and LDB is the optimal combination considering recall rates and efficiency.

Table 1. The maximum recall rates and average running time of methods with different network backbones and local feature descriptors

Model	KITTI 05		CC	
	Recall (%)	Time (ms)	Recall (%)	Time (ms)
DpMn2+Brief	82.94	48.9174	69.34	45.6920
DpMn2+Surf	86.35	69.2330	72.55	53.3149
DpMn2+LDB	**86.35**	**53.5476**	**73.98**	**49.2461**
DpXc65+LDB	86.35	127.8594	74.86	124.3898

Notes: DpMn2 is Deeplabv3 with MobileNetV2 as the network backbone, DpXc65 is Deeplabv3 with Xception65 as the network backbone, LDB is local difference binary.

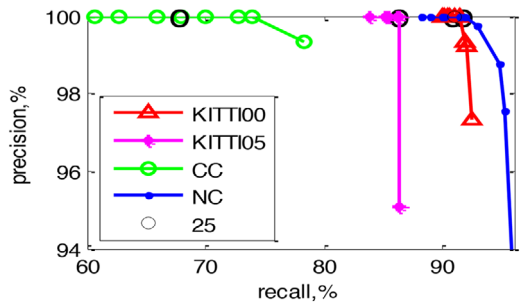


Fig. 3 *Precision-recall curves for four datasets*

Table 2. The recall rates of different algorithms

	KITTI00 (%)	KITTI05 (%)	CC (%)	NC (%)
DBoW2 [3]	–	–	43.03	70.29
Zhang et al. [1]	**95.37**	71.01	60.86	48.79
Han et al. [2]	54.90	62.50	27.70	20.90
FILD [5]	91.23	85.15	–	–
GPR [4]	–	–	50.36	71.28
Proposed method	90.95	**86.35**	**67.91**	**91.85**

Table 3. Execution time (ms/per frame)

	KITTI 00	KITTI 05
FILD [5]	337.478	339.949
Proposed method	56.698	53.5476

Figure 3 shows the precision-recall curves for four datasets with the threshold δ decreasing from 50 to 5 with step of 5. It can be observed that δ needs to be set to 25 to ensure a 100% precision for all four datasets.

With the optimal parameters determined by the above experiments, we compare the performances of our method with some typical or state-of-the-art methods. Table 2 shows the recall rate of each algorithm under 100% precision. In this table '–' means the comparison methods do not perform experiments on the dataset. Notably, our algorithm achieves the highest recall rates in KITTI 05, CC and NC datasets. Zhang et al. achieve the highest recall rate for KITTI 00 dataset because this dataset contains some small dynamic objects that are difficult to split out.

Finally, to evaluate the efficiency of our method, we compare the execution time between FILD, which has the second-best result, and our method. Table 3 shows the time consumed per frame on KITTI 00 and KITTI 05 datasets. The average running time per frame of our method is only one-fifth of the time consumed by FILD.

Conclusion: In this letter, we propose a fast online loop-closure detection algorithm based on DpMn2 and LDB, which splits out dynamic objects and extracts ConvNet-based features to retrieve the nearest neighbour image based on HNSW. LDB descriptors and RANSAC are used for geometrical verification. Experimental results and comparisons show that our method can achieve a good recall rate with high efficiency.

Acknowledegments: This work was supported by the National Natural Science Foundation of China (grant 61372145), the Independent Innovation Fund of Tianjin University (grant 2015XZC-0005) and the Fundamental Research Project of Qinghai Province (grant 2017-ZJ-753). The authors would like to appreciate Dr. Konstantinos A. Tsintotas for kindly offering ground truth information for the datasets.

Received: *6 September 2020* Accepted: *22 October 2020*
doi: 10.1049/ell2.12052

References

1 Zhang, X. et al.: Graph-based place recognition in image sequences with CNN Features. *J. Intell. Robot. Syst.* **95**(8), 389–403 (2019)
2 Wang, H. et al.: Fast loop closure detection via binary content. In: 2019 IEEE International Conference on Control and Automation (ICCA), Edinburgh, the United Kingdom, pp. 1563–1568 (2019)
3 Mur-Artal, R., Tardos, J.D.: Fast relocalisation and loop closing in keyframe-based SLAM. In: 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, pp. 846–853 (2014)
4 Yang, Z. et al.: Gridding place recognition for fast loop closure detection on mobile platforms. *Electron. Lett.* **55**(17), 931–933 (2019)
5 An, S. et al.: Fast and incremental loop closure detection using proximity Graphs. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, pp. 378–385 (2019)
6 Malkov, Y.A., Yashunin, D.A.: Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**(4), 824–836 (2020)
7 Yang, X., et al.: LibLDB: A library for extracting ultrafast and distinctive binary feature description. In: 22nd ACM International Conference on Multimedia (ACM), New York, USA, pp. 671–674 (2014)
8 Zhou, B. et al.: Places: A 10 million image database for scene recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(6), 1452–1464 (2018)
9 Tsintotas, K.A. et al.: Assigning visual words to places for loop closure detection. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, pp. 5979–5985 (2018)
10 Geiger, A., et al.: Are we ready for autonomous driving? The KITTI vision benchmark suite. In: 2012 IEEE International Conference on Computer Vision and Pattern Recognition (CPVR), Providence, Rhode Island, pp. 3354–3361 (2012)
11 Cummins, M., Newman, P.: FAB-MAP: probabilistic localization and mapping in the space of appearance. *Int. J. Robot. Res.* **27**(6), 647–665 (2008)

Will people like your illustration? Popularity assessment of illustrations

Jiachen Yang, Yao Men, and Yang Zhao✉
School of Electrical and Information Engineering, Tianjin University, Tianjin, China
✉Email: yangzhao321@tju.edu.cn

Illustrations frequently appear in advertisements, films, books, etc. Research into the popularity of illustrations shows practicality and necessity already, which plays an important role in the improvement of illustrations and advertising decisions. In this letter, an Illustration Popularity Assessment Network (IPANet) is designed, inspired by the illustration creation process. Atrous convolutional layers are densely

concatenated to extract image composition information and Gram matrix is used to measure the image style. We also introduce a scoring model based on 'views' and 'likes' to score the popularity of collected illustrations. The IPANet is proved to be effective according to experimental results on the scored illustrations. As far as we know, we are the first to explore the popularity of illustrations and practice successfully.

Introduction: An illustration is a visual interpretation and decoration of a text, concept, or process (e.g. picture book, infographics, and technical and scientific illustration). Previously, most illustrations appeared in newspapers, magazines, or books. Nowadays, in addition to traditional media, illustrations also can be seen in posters, movies, animations, video games, etc. If we understand which kind of illustrations people would prefer, we can guide illustrators to modify their works, develop targeted advertising strategies, and help games and movies attract more customers. Therefore, research on the popularity of illustrations is of great significance. Early illustrations are commonly produced in pencil, watercolour, oil paints, and acrylics. Nowadays illustrators use computers for creation or post-processing. They share their works, view other images, and express their opinions on online platforms. When relevant information could be collected and analysed, we will infer users' preferences to a certain extent.

The illustrations are often used to explain the text or attract people's attention. Therefore, illustrations dose not mind authenticity but emphasise a sharp and exaggerated style, which is very different from photographs. In this letter, we tailor an Illustration Popularity Assessment Network (IPANet) based on the illustration creation process. As shown in Figure 2, the network design focuses on image composition and image style. Densely connected Atrous Spatial Pyramid Pooling (DenseASPP) [1] containing multiple atrous convolutions [2] is used to learn the composition information. Multiple Gram blocks [3] based on the Gram matrix are used to learn the style information [4]. Moreover, we develop a method to score the collected images based on 'views' and 'likes'. Because most platforms provide these two indicators, the scoring model has a wide application. The experimental results on the scored illustrations show that IPANet can evaluate the popularity of illustrations effectively.

Fig. 1 *Examples of illustrations and the corresponding scores and predictions. The scores are obtained by the new scoring model. The predictions are obtained by the IPANet*

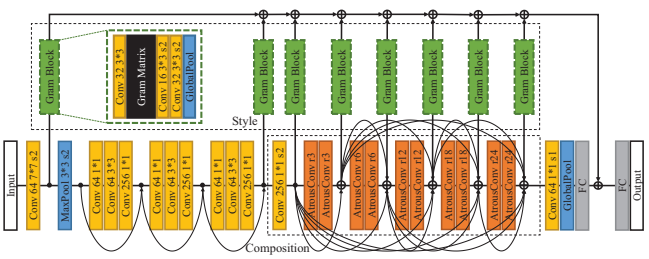


Fig. 2 *The architecture of IPANet. '⊕' is the concatenation operation*

Figure 1 shows some examples of illustrations and the corresponding scores and predictions.

Popularity prediction of illustrations: As shown in Figure 2, we design the IPANet to automatically evaluate the popularity of illustrations. The composition and style of images are the two most important factors in the process of illustration creation, so they are the two main considerations in the design of IPANet. When creating an illustration, the spatial composition must be clearly structured, including the size and position of objects and the perspective. The composition is not restricted by the real scene in illustration creation, so the size and position may vary greatly, and the perspective is always exaggerated. It is necessary to extract features with different receptive field sizes. We use DenseASPP to extract image composition information. Multiple atrous convolutional layers with different dilation rates form a pyramid structure. For atrous convolution [2], a large dilation rate means a large receptive field:

$$y[i,j] = \sum_{t=0}^{K-1}\sum_{k=0}^{K-1} x[i+rk, j+rt]w[k,t], \tag{1}$$

where r is the dilation rate, $x[i, j]$ and $y[i, j]$ represent the input and output values at $[i, j]$, respectively, $w[k, t]$ denotes the parameter of filter at position $[k, t]$, and K is the filter size. The outputs of different layers with different dilation rates are densely concatenated. The dilation rates are 3, 6, 12, 18, and 24. In this way, a denser feature pyramid can be obtained, owing to denser scale sampling and denser pixel sampling [1]. We finally get 512-dimension image composition features.

Natural images with different styles are still similar, however, illustrations of different styles are not the same at all, such as ink-and-wash painting and oil painting. Therefore, the extraction of image style features is a focus of IPANet. Each layer of the Convolutional Neural Network (CNN) network has multiple local feature extractors. Each extractor has some features of its interest, such as vertical edge, horizontal edge, colour, and texture. The relationship between different local features can reflect the image style. We use the Gram matrix to calculate the correlations between local features [4]. If the Frobenius norm of the difference between two Gram matrices is small, the feature maps are similar in style. Suppose there are $K^{[l]}$ feature maps of size $W^{[l]} \times H^{[l]}$ in layer l. The Gram matrix at $[i, j](i = 0, 1, \ldots, H^{[l]} - 1, j = 0, 1, \ldots, W^{[l]} - 1)$ is defined as

$$G^{[l]}_{i,j} = \begin{bmatrix} a^{[l]T}_{i,j,1}a^{[l]}_{i,j,1} & a^{[l]T}_{i,j,1}a^{[l]}_{i,j,2} & \cdots & a^{[l]T}_{i,j,1}a^{[l]}_{i,j,k} \\ a^{[l]T}_{i,j,2}a^{[l]}_{i,j,1} & a^{[l]T}_{i,j,2}a^{[l]}_{i,j,2} & \cdots & a^{[l]T}_{i,j,2}a^{[l]}_{i,j,k} \\ \vdots & \vdots & \ddots & \vdots \\ a^{[l]T}_{i,j,k}a^{[l]}_{i,j,1} & a^{[l]T}_{i,j,k}a^{[l]}_{i,j,2} & \cdots & a^{[l]T}_{i,j,k}a^{[l]}_{i,j,k} \end{bmatrix} \tag{2}$$

where $a^{[l]}_{i,j,k}$ is the activation of the kth$(k = 0, 1, \ldots, K^{[l]} - 1)$ feature map at $[i, j]$ of layer l. The Gram matrix of layer l is

$$G^{[l]} = \sum_{i=0}^{H^{[l]}-1}\sum_{j=0}^{W^{[l]}-1} G^{[l]}_{i,j}, \tag{3}$$

where $G^{[l]} \in \mathbb{R}^{K^{[l]} \times K^{[l]}}$. The diagonal elements provide information about feature maps themselves, while the information between different feature maps can be found in remaining elements. Gram matrix not only reflects what features there are but also reflects the tightness between different features. Therefore, it is very suitable for describing the style of images. Generally the higher the layer, the more abstract the features. We use eight Gram blocks with Gram matrix as the core to extract features from different layers. Each Gram block outputs 32-dimension features. Concatenating together all the outputs, we get 256-dimension image style features.

Finally, all the style features and the composition features are concatenated together for the final prediction.

Scoring model: There is no public dataset for popularity prediction. Following the previous researchers [5–7], we crawl illustrations directly from online platforms to reflect the preferences of a wider class of participants from different backgrounds more credibly. A time-independent

scoring model that takes into account the total amount is proposed. Each image is associated with the number of 'views' n and the number of 'likes' k. 'Views' is the number of times an image has been watched, while 'likes' is the number of people who like the image among all the people who have viewed it. Event E is defined as 'a total of n people have viewed the given image, of which k people like it'. The probability of E occurring is p. We use p to measure the popularity of the image. E can be described by a binomial distribution. A beta distribution with two positive shape parameters α_0 and β_0 is used to model the prior distribution of p. The posterior probability can be derived from Bayes' theorem:

$$P(p|k,n) = \frac{P(k,n|p)P(p)}{P(k,n)} = \frac{C_n^k p^{\alpha_0+k-1}(1-p)^{\beta_0+n-k-1}}{B(\alpha_0,\beta_0)P(k,n)}, \quad (4)$$

where the beta function B is a normalisation constant to ensure that the total probability is 1. Let $\alpha = \alpha_0 + k$, $\beta = \beta_0 + n - k$. We ignore the very small numbers α_0 and β_0. The beta distribution with parameters k and $n-k$ can be used to model the posterior distribution of p.

Due to the total number of samples, beta distributions with the same expected value may have great differences. So we consider the probability p_{bound} at the lower boundary of the $1/e$ of the maximum point at Probability Density Function (PDF). This point corresponds to the maximum Wilson entropy of the distribution. The probability of the maximum point at the PDF is $p_{\text{appro}} = \frac{\alpha-1}{\alpha+\beta-2}$. We denote the ratio of the PDFs as a function $F(p_{\text{bound}})$. Let $\alpha_0{=}\beta_0{=}1$ to simplify the calculation. An approximate equation for $F(p_{\text{bound}})$ is

$$F(p_{\text{bound}}) = \frac{P(p_{\text{appro}}|k,n)}{P(p_{\text{bound}}|k,n)} \approx \exp\left\{\frac{n}{2}\left[1 - \frac{p_{\text{bound}}^2}{p_{\text{appro}}} - \frac{(1-p_{\text{bound}})^2}{1-p_{\text{appro}}}\right]\right\}. \quad (5)$$

Solving equation $F(p_{\text{bound}}) = \frac{1}{e}$, we obtain the probability at the lower boundary:

$$p_{\text{bound}} = p_{\text{appro}} - \sqrt{\frac{2p_{\text{appro}}(1-p_{\text{appro}})}{n}} + r\frac{p}{n}, \quad (6)$$

where $r\frac{p}{n}$ is the correction for approximation errors, $r = 2.8$ is obtained from the empirical value. Considering the data accuracy in the code, the scores are multiplied by 100.

Experimental results and discussion: We evaluate the performance of the IPANet on illustrations. We download 600 images from Behance (www.behance.net), which is a social media platform with over 10 million members. All images are scored according to the proposed scoring model. The maximum score is 40.16, the minimum is 0.83, and the average is 14.40. Note that 80% of the images are randomly selected for training, and the remaining are for testing. We shrink the input image to a short side length of 576. Square patches with a size of 512×512 are cropped randomly. During the training process, Adam optimiser is applied. The mini-batch size is 32. The initial learning rate is $1e-4$. If no improvement is seen for five epochs, the learning rate will be reduced to $1/10$ of the current learning rate. We use Pearson Linear Correlation Coefficient (PLCC) and Spearman Rank-Order Correlation Coefficient (SROCC) to evaluate the performance of different methods following previous researches.

We compare with other state-of-the-art image popularity assessment methods and modern deep networks (VGG16, MobileNet, and Inception-v3). As shown in Table 1, the good performance of IPANet on the illustration dataset proves its superiority over other image popularity assessment methods. IPANet also performs better than the deeper or more modern networks (the Inception-v3 cannot even converge). Its good performance documents that the IPANet is indeed tailored to illustrations rather than being simply a larger deep network architecture. Further, the performance of the IPANet on 10, 000 natural images from Flicker [5] is still stable but not much more effective than other methods. This is because the IPANet is customised for illustrations.

The DenseASPP for image composition and the Gram blocks for image style are the most important modules for IPANet. For showing which modules can help when dealing with illustrations rather than natural images, we further do ablation tests as shown in Table 2. The 'name' in the table is just for the convenience of describing the structure of the corresponding rows. The performance of IPANet-None without composition and style modules is quite poor. After adding the module for image composition, the IPANet-C performs better. By adding the module for image style, the performance has been greatly improved (IPANet-S and IPANet-CS8). When we gradually add Gram blocks (IPANet-CS2-IPANet-CS8), the network performance gets better and better. So the style features have a great influence on the popularity of illustrations, which confirms the huge difference between illustrations drawn by computers and photographs taken by cameras.

Table 1. Comparison results of different methods

	Illustration		Natural image	
Method	PLCC	SROCC	PLCC	SROCC
Visual-LibLinear [5]	0.3768	0.3228	0.4910	0.4373
Visual-LibSVM [5]	0.4125	0.3499	0.5187	0.5002
Visual [6]	0.8484	0.8232	0.8120	0.8062
UHAN (w/o u) [7]	0.8376	0.8201	0.8431	0.8207
VGG16	0.8314	0.8198	0.8402	0.8013
MobileNet	0.7481	0.7119	0.6911	0.6735
Inception-v3	0.2429	0.2138	0.2441	0.2403
Proposed	0.8791	0.8471	0.8405	0.8192

Table 2. Ablation tests of IPANet modules

	Module			Prediction	
Name	Composition	Style	Blocks	PLCC	SRCC
IPANet-None	No	No	0	0.4508	0.4812
IPANet-S	No	Yes	8	0.7532	0.7432
IPANet-C	Yes	No	0	0.8093	0.7798
IPANet-CS2	Yes	Yes	2	0.8100	0.8081
IPANet-CS4	Yes	Yes	4	0.8153	0.7957
IPANet-CS8	Yes	Yes	8	0.8791	0.8471

Conclusion: In this letter, we discuss the popularity assessment of illustrations for the first time. We design a novel IPANet, inspired by the illustration creation process. DenseASPP and Gram blocks are used to extract composition and style features of images, respectively. We also develop a scoring model to score the images based on the number of 'views' and 'likes'. It takes total influence factors into account and is widely adapted to various Internet platforms. Experimental results confirm the superiority of the IPANet over the state-of-the-art methods for image popularity prediction. Our work is quite a novel exploration of the illustration popularity assessment.

Acknowledgments: This work was partially supported by the National Natural Science Foundation of China (No. 61871283), the Foundation of Pre-Research on Equipment of China (No. 61400010304), and the Major Civil-Military Integration Project in Tianjin, China (No. 18ZXJMTG00170).

Received: *18 September 2020* Accepted: *26 October 2020*
doi: 10.1049/ell2.12054

References

1 Yang, M., et al.: Denseaspp for semantic segmentation in street scenes. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3684–3692. IEEE, Piscataway, NJ (2018)
2 Chen, L.-C., et al.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(4), 834–848 (2017)
3 Liu, Z., Qi, X., Torr, P.H.: Global texture enhancement for fake face detection in the wild. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8060–8069. IEEE, Piscataway, NJ (2020)
4 Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2414–2423. IEEE, Piscataway, NJ (2016)
5 Hu, J., Yamasaki, T., Aizawa, K.: Multimodal learning for image popularity prediction on social media. *2016 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, pp. 1–2. IEEE, Piscataway, NJ (2016)
6 Zhang, Z., et al.: How to become instagram famous: Post popularity prediction with dual-attention. *2018 IEEE International Conference on Big Data (Big Data)*. pp. 2383–2392. IEEE, Piscataway, NJ (2018)
7 Zhang, W., et al.: User-guided hierarchical attention network for multimodal social image popularity prediction. *Proceedings of the 2018 World Wide Web Conference*, pp. 1277–1286. ACM, New York, NY (2018)

Privacy-preserving evaluation for support vector clustering

J. Byun,[1] J. Lee,[1] and S. Park[2,✉]

[1]*Department of Industrial Engineering, Seoul National University, Korea*

[2]*Department of Convergence Security Engineering, Sungshin University, Korea*

✉E-mail: psr6275@sungshin.ac.kr

The authors proposed a privacy-preserving evaluation algorithm for support vector clustering with a fully homomorphic encryption. The proposed method assigns clustering labels to encrypted test data with an encrypted support function. This method inherits the advantageous properties of support vector clustering, which is naturally inductive to cluster new test data from complex distributions. The authors efficiently implemented the proposed method with elaborate packing of the plaintexts and avoiding non-polynomial operations that are not friendly to homomorphic encryption. These experimental results showed that the proposed model is effective in terms of clustering performance and has robustness against the error that occurs from homomorphic evaluation and approximate operations.

Introduction: Recently, machine learning technologies have successfully solved real-world problems in various fields, including biomedical and financial applications. Sensitive data such as personal data, biometric data, medical information, and financial information can be used to construct high-quality machine learning models. Therefore, moral and legal issues about data protection and privacy use have received attention, and the conflict between data utilization and protection needs to be addressed.

Fully homomorphic encryption (FHE), which enables numerical operations on encrypted data, is considered to be a promising direction that satisfies data utilization and protection [1, 2]. Privacy-preserving machine learning algorithms with FHE have been proposed to train supervised models and evaluate the models on the encrypted domain, where encrypted data can be transmitted to the model without revealing the original data [3, 4]. However, implementing machine learning algorithms with FHE involves much slower computations and much larger data storage than implementing the same algorithms on the plaintext domain. Implementing machine learning algorithms without considering the operational characteristics of FHE can worsen the problems and degrade the performance.

Clustering is a representative unsupervised learning task widely used in areas including image segmentation, information retrieval, and marketing. Clustering algorithms partition given instances into a set of subgroups called clusters depending on their similarity (or distance). Clustering on the encrypted data can be more complicated than classification or regression because the shape and the number of clusters are unknown. Clustering algorithms such as k-means clustering and mean shift clustering methods have been implemented on encrypted data but lack the performance of complex and non-convex data [5, 6]. In contrast, the support vector clustering (SVC) algorithm can capture the complex shape of clusters by labelling the support of data distribution based on the support vector domain description (SVDD) [7, 8].

In this letter, we propose a privacy-preserving evaluation for SVC. Our work aims to implement an efficient SVC inference on the encrypted domain, where it can capture the complex data distribution with a support function and assign the most appropriate cluster for a new test data. Our algorithm enables robust SVC labelling for test data on an encrypted domain without decryption by configuring the entire procedure as a homomorphic operation. In the experiments, six datasets were used to evaluate the performance of clustering algorithms on the encrypted domains.

FHE and HEAAN scheme: An FHE scheme aims to construct a homomorphic encryption (HE) scheme that supports an unbounded number of operations for evaluating any function f on encrypted data without decryption [1]. An FHE scheme consists of four procedures: KeyGen, Encrypt, Decrypt and Evaluate. The correctness condition for evaluated ciphertexts ensures that for any function f, a plaintext m and its encryption c, the decryption result of the ciphertext $c' = \mathsf{Evaluate}(f, c)$ is the same as $f(m)$. In FHE, any evaluation function is represented as a composition of homomorphic additions and multiplications. Thus, homomorphic evaluations of non-polynomial operations can consume a much longer time and have faster error propagation than the evaluations on the plaintext domain.

In this letter, we used a HEAAN scheme that supports the approximate computation of real numbers, where a small error is added to the plaintext vector after decryption [2]. HEAAN can provide efficient floating-point operations at the expense of a bounded loss of precision. A complex vector (plaintext) can be encoded into a ring element and encrypted into a single ciphertext, and the slot rotation of the vector on the encrypted domain enables an efficient parallel computation of the ciphertexts. Because of efficient computation, many machine learning algorithms based on HEAAN have been proposed for real-world applications [4, 6].

Initializing the HEAAN scheme, some parameters are determined to achieve a targeted level of security, including N, initial ciphertext modulus $\log q_L$ and scaling factor p, where N is related to the ciphertext space and the number of plaintext slots ($= N/2$). The parameters $\log q_L$ and p determine the precision of the calculations and the number of operations without bootstrapping. Since the error in the evaluated ciphertext grow rapidly with multiplication compared to other operations, it is necessary to rescale the ciphertext after multiplication to control the magnitude of the error with decreasing the ciphertext modulus.

The bootstrapping procedure allows the evaluated ciphertext to be refreshed by homomorphically evaluating the Decrypt function with increasing the ciphertext modulus. In the case of HEAAN, the computational cost of bootstrapping increases with the number of plaintext slots with $O(\log N/2)$ [9]. Although bootstrapping of HEAAN is more efficient than other FHE schemes, it is still the most expensive part of HEAAN. Therefore, it is essential to consider the trade-off between the efficiency of parallel operation and the cost of bootstrapping. For detailed information about the scheme, we refer the readers to [2, 9].

SVC: Support-based clustering starts with estimating the support function of data distribution obtained by the SVDD, Gaussian process clustering, or kernel density estimation. In this study, we used SVDD with the Gaussian kernel to obtain the support function as follows:

$$s(\mathbf{x}) = 1 - 2\sum_{i=1}^{N_v} \beta_i e^{-\delta\|\mathbf{x}-\mathbf{x}_i\|^2} + \sum_{i=1}^{N_v}\sum_{j=1}^{N_v} \beta_i\beta_j e^{-\delta\|\mathbf{x}_i-\mathbf{x}_j\|^2}, \quad (1)$$

ALGORITHM 1 HE friendly evaluation of support-based clustering

Input: Test data $\{\mathbf{x}_i\}_{i=1}^{N_{test}}$, SEVs $\{\mathbf{s}_i\}_{i=1}^{N_S}$, SVs $\{\mathbf{v}_i\}_{i=1}^{N_V}$, support function s, number of iteration T, learning rate η

Output: Clustering label $\{l_i\}_{i=1}^{N_{test}}$

1: **for** $i = 1$ to N_{test} **do**

2: $\quad \mathbf{x}_i^0 \leftarrow \mathbf{x}_i$

3: $\quad$ **for** $t = 1$ to T **do**

4: $\quad\quad \mathbf{x}_i^t \leftarrow \mathbf{x}_i^{t-1} - \eta \cdot \nabla s(\mathbf{x}_i^{t-1})$

5: **end for**

6: Find the nearest SEV $(\mathbf{s}_i)$ of $\mathbf{x}_i^T$, and label $\mathbf{x}_i$ as l_i, the label of $\mathbf{s}_i$.

7: **end for**

where $\delta > 0$ is the width parameter for the Gaussian kernel and N_v denotes the number of support vectors (SVs). The support function (1) can be used to partition the data space into basin cells by constructing the following dynamical system:

$$\frac{d\mathbf{x}}{dt} = -\nabla s(\mathbf{x}). \tag{2}$$

A stable equilibrium vector (SEV) of the system (2) is an equilibrium state where all the eigenvalues of Hessian $\nabla^2 s(\mathbf{x})$ are positive. Then, the basin cell $B(\mathbf{s}_i)$ is defined as the closure of the set of all the data points that converge to a SEV $\mathbf{s}_i$ following the system (2). Some SEVs are connected to constitute clusters using the characteristic of the dynamical system (2), and the points in a basin cell are labelled with the cluster label of the corresponding SEV [7, 8, 10]. SVC can be inductive and have an efficient and stable inference phase using the system (2) [7, 10].

HE-friendly evaluation of SVC: In user-server scenarios, the inference phase of the clustering algorithm with FHE can provide to partition the given encrypted instances while protecting user's privacy from curious servers. In this letter, we present the inference phase of the SVC algorithm with FHE because SVC is naturally inductive and has a stable inference. For SVC, most inference methods are based on the system (2) that a test point follows to find the SEV. In particular, the simplest inference method is to allocate the test points to the cluster of the closest SEV. However, this inference cannot capture the complex clusters because it ignores the intrinsic data distribution. Therefore, we present an elaborate inference algorithm of SVC which utilizes the support function (1) of the data distribution. We addressed the challenge of dealing with HE-unfriendly operations of SVC inference while balancing efficiency and accuracy.

We propose a HE-friendly evaluation that utilizes the basin induced from the system (1) to stabilize the inference. Our algorithm consists of two parts. In the first part, we use the dynamical system (2) to move the given test data to more stable points. We can postulate that the boundary regions between two different basins are the most unstable, whereas the SEVs are the most stable points. Thus, our algorithm makes the points evade the boundary region by applying the system (2) which converges to SEVs. In the second part, our algorithm assigns the cluster label by finding the SEV closest to the point obtained after the first part. The procedure of our proposed method is presented in Algorithm 1.

In Algorithm 1, we need to implement the calculation of the gradient of the support function $\nabla s(\mathbf{x})$ homomorphically. When using the Gaussian kernel, $\nabla s(\mathbf{x})$ can be directly calculated as:

$$\nabla s(\mathbf{x}) = 4\delta \sum_{j=1}^{N_v} \beta_j e^{-\delta\|\mathbf{x}-\mathbf{v}_j\|^2} \cdot (\mathbf{x} - \mathbf{v}_j). \tag{3}$$

HEAAN is used to efficiently implement the approximate computation of real numbers on encrypted data, which supports homomorphic additions and multiplications. However, it requires polynomial approximations for non-polynomial operations. The gradient (3) contains the exponential function and the distance between a test point and the SVs.

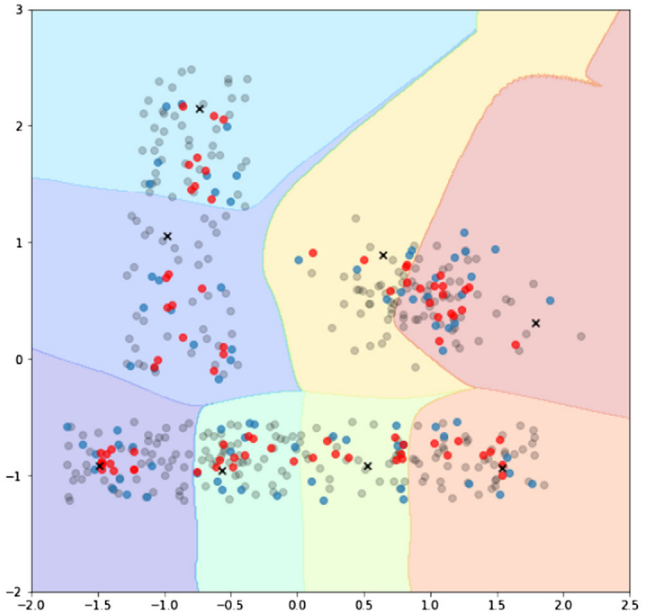


Fig. 1 *Illustration of gradient step for Algorithm 1, where the gray points represent training data, the blue points represent test data, the red points represent the data point after each gradient step, and each "x" point represents the corresponding SEV of the basins region with different colors*

After the first part has been found without decryption, finding the nearest SEV consists of two components: calculating the distances between a test point and the SEVs and finding the minimum distance between them. The exponential function is approximated with a Taylor expansion. To obtain the min-index of the distances, we used the iterative algorithm in [11]. However, the appropriate packing strategy is needed to improve the computational efficiency of calculating the distances and finding the min-index.

We design the plaintext packing to avoid bootstrapping. For simplicity, we assume that all vectors, including test data, SVs, and SEVs, are d-dimensional row vectors. Figure 2 shows the structures of the packed plaintexts, where a plaintext vector will be encrypted into a ciphertext. Note that HEAAN supports addition, element-wise multiplication, and right-shift and left-shift rotations for plaintext vectors on the encrypted domain. We designed the plaintext vector to efficiently calculate the distances using these operations because both parts of Algorithm 1 involve the calculation of the distances. The test samples $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_{test}}$ are repeated as many as N_V, while each instance of SVs $\mathbf{v}_i$ and SEVs $\mathbf{s}_i$ are repeated as many as the number of test data N_{test} as in Figure 2. For SEVs, N_{test} times repeated vectors are additionally repeated N_V/N_S times. The plaintexts become $dN_{test}N_V$-dimensional vectors, which consist of d-dimensional sub-vectors. In this way, subtracting all SVs or SEVs from all test data can be done with only one operation. In addition, $\beta = [\beta_1, \ldots, \beta_{N_V}]$ were packed in the same way as SVs, where β_j's were repeated d times to constitute a vector like $\mathbf{v}_j \in \mathbb{R}^d$. Depending on the number of test samples N_{test} and SVs N_V, we can use multiple ciphertexts for a plaintext vector and parallelize the homomorphic operations.

Finally, we can efficiently compute the difference $\mathbf{x}_i - \mathbf{v}_j$ for obtaining the gradient (3) and $\mathbf{x}_i - \mathbf{s}_j$ for finding the closest SEV. After constituting all necessary ciphertexts, we can compute the gradient using homomorphic operations without decryption by replacing the entire operations with the polynomial operations. When finding the nearest SEV, we use the min-index algorithm whose output has a reciprocal of the number of minimal elements for the indices of the minimal elements and 0 for the other indices as in [6]. Because the min-index algorithm on the encrypted domain involves an approximate error, the outputs of the minimal elements can be indistinguishable from 0s with the errors when the number of minimal elements is large. However, the gradient phase of our algorithm induces the test samples near the boundary to move to the SEV, resulting in more accurate homomorphic results for the min-index algorithm. Figure 1 illustrates the change of test points after one gradient

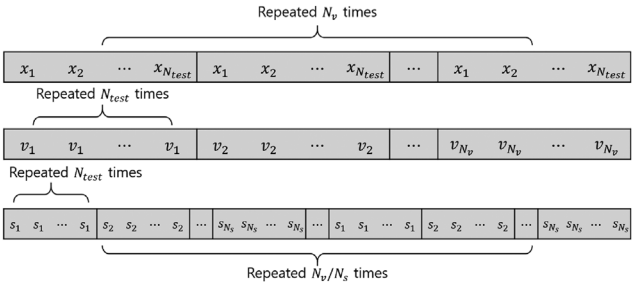


Fig. 2 *Description of the packing method*

Table 1. Summarization of the datasets

Dataset	Instances	Attributes	Clusters	Convexity
Hepta	212	3	7	convex
Tetra	400	3	4	convex
Lsun	400	2	3	convex
Two diamonds	800	2	2	convex
Target	758	2	2	non-convex
Chainlink	1000	3	2	non-convex

descent iteration. We can notice that the test points move toward the SEVs of their basins and away from the other SEVs simultaneously. It can affect the clustering performance for complex data distribution.

Experiments: We evaluated the proposed method on six datasets shown in Table 1 from the fundamental clustering problems suite [12]. We compared clustering performance in terms of the adjusted Rand index (ARI) metric and computation time with HE-friendly mean shift clustering (Meanshift) and k-means clustering (KMeans), which are currently the state-of-the-art clustering models used with FHE. The ARI measures the similarity between two data partitions and has a value between 0 and 1, where 1 represents a perfect agreement between two data partitions. To verify how the error included in ciphertext affects the clustering results, we compared the results of Algorithm 1 to encrypted data (ARI-enc) and unencrypted data (ARI-no). We implemented the evaluation of KMeans and Meanshift, following [6].

We used an Intel Xeon CPU E5-2660 v3 @2.60GHZ processor. We set $\log q_L = 1200$, $\log p = 30$, $N = 2^{16}$ and $\delta = 2$ for all datasets. We conducted the experiments with different test rates in $\{0.2, 0.5, 0.9\}$, which means the ratio between the number of test data and the number of training and test data. Training data was sampled five times for each experiment to measure the average performance. For Algorithm 1, we used one iteration step of (2), where the learning rate was set to 0.5 for the Chainlink dataset and 0.8 for the other datasets.

Table 2 shows the clustering results. KMeans showed almost the same computation time for all experiments since a plaintext vector is encrypted into a single ciphertext. In contrast, except Hepta and Lsun, the proposed method showed the longest execution time because it needed to split a plaintext into multiple ciphertexts. The total execution time is highly dependent on the number of ciphertexts, which is related to the number of SVs for the proposed method, to the number of modes for Meanshift, and to the number of clusters for Kmeans. Therefore, we can reduce the computational cost if the SVC algorithm obtains a sparse support function (1) with few support vectors.

For Hepta and Tetra datasets, which have simple and convex distributions, all the models showed good performance. For all the other datasets, our method showed the highest ARI value regardless of the test rate. The other algorithms showed poor performance, especially for the Target and Chainlink datasets that have non-convex data distributions, while our model achieved high ARIs. For the Lsun dataset that consists of rectangular clusters with different lengths and widths, the ARI values of the KMeans and Meanshift algorithms for encrypted and unencrypted data are different significantly, whereas SVC always showed consistent results even after encryption. These results demonstrate that by pulling the data point at the boundary toward the center of the basin cell, SVC can attain robustness against the error that can occur from homomorphic operations.

Table 2. Comparison of the results of three algorithms on FCPS datasets. For each experiment, the highest ARI value is highlighted in bold

Dataset	Test rate	KMeans		Meanshift		SVC	
		ARI-enc	time (s)	ARI-enc	time (s)	ARI-enc	time (s)
		ARI-no		ARI-no		ARI-no	
Hepta	0.2	1	16.967	1	17.332	1	30.402
		1		1		1	
	0.5	1	17.001	1	16.994	1	28.360
		1		1		1	
	0.9	1	17.025	1	17.028	1	27.239
		1		1		1	
Tetra	0.2	**1**	16.162	0.957	16.962	0.927	57.585
		1		0.970		0.927	
	0.5	**1**	16.505	0.973	17.383	0.947	113.896
		1		0.973		0.947	
	0.9	**0.994**	16.110	0.915	17.033	0.961	109.437
		0.994		0.915		0.961	
Lsun	0.2	0.351	15.386	0.629	16.543	**0.875**	25.673
		0.417		0.613		**0.875**	
	0.5	0.346	15.328	0.611	16.494	**0.931**	25.837
		0.400		0.651		**0.931**	
	0.9	0.342	15.409	0.663	16.171	**0.993**	25.368
		0.400		0.734		**0.993**	
Two diamonds	0.2	0.897	14.978	0.812	15.476	**0.995**	26.147
		0.897		0.812		**0.995**	
	0.5	0.891	15.085	0.842	15.404	**0.934**	52.469
		0.891		0.842		**0.934**	
	0.9	0.879	14.640	0.777	15.534	**0.923**	102.129
		0.879		0.777		**0.923**	
Target	0.2	0.105	14.694	0.619	17.055	**1**	26.375
		0.105		0.623		**1**	
	0.5	0.119	14.822	0.620	17.331	**1**	53.773
		0.119		0.620		**1**	
	0.9	0.121	14.857	0.627	17.185	**1**	104.359
		0.121		0.627		**1**	
Chain-link	0.2	0.094	15.639	0.235	18.890	**0.853**	111.958
		0.094		0.231		**0.853**	
	0.5	0.098	15.465	0.228	37.062	**0.887**	223.896
		0.098		0.229		**0.887**	
	0.9	0.068	15.548	0.318	15.513	**0.967**	220.380
		0.068		0.318		**0.967**	

Figure 3 shows the result of our investigation of the difference in clustering performance by presenting the clustering results for three data sets. Figure 3a,b illustrates that KMeans and Meanshift failed to capture the clusters, especially for the data points at the boundary of the divided region. However, our method correctly allocated the clusters because of the robustness of our method, as stated in the above paragraph.

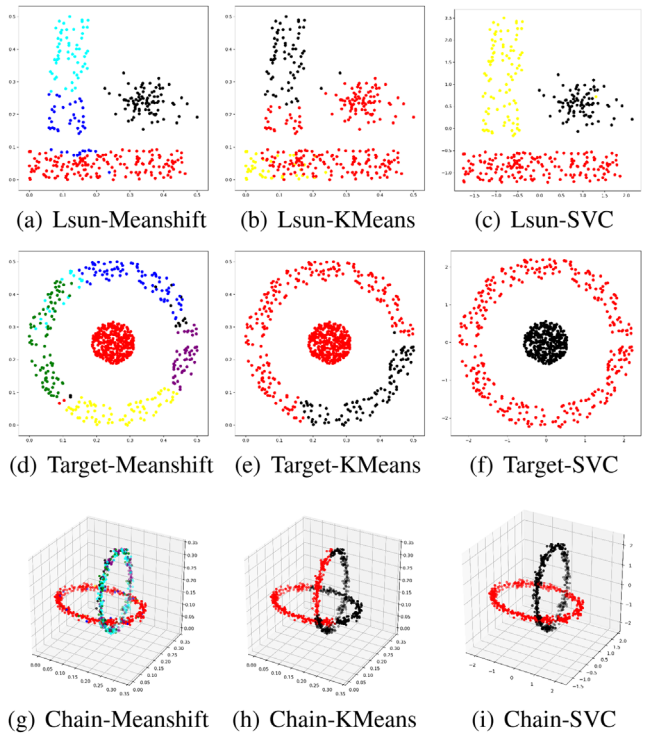


Fig. 3 *Visualization of clustering results for Lsun, Target and Chainlink datasets. Each estimated cluster is color-coded*

In the case of datasets with non-convex distributions such as Target and Chainlink, we found that KMeans essentially did not reflect the non-convex distribution, and Meanshift was not able to connect the clusters into a single cluster. On the other hand, SVC completely clustered the Target dataset, and for the Chainlink dataset, almost all data points were properly clustered except for the points located where two clusters are adjacent to each other.

Conclusion: In this letter, we proposed a privacy-preserving evaluation algorithm of SVC with fully homomorphic encryption. Our model enables the allocation of the cluster label for new test data without decryption, improving the clustering performance for non-convex data, and provides robustness of data points near the boundary. The experimental results show that the proposed method effectively clusters encrypted data with various distributions in a realistic amount of time. In the future, we can improve the computational cost of our algorithm by training a sparse SVC model and parallelizing computationally expensive operations when using multiple ciphertexts.

Acknowledgements: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2018R1D1A1A02085851), and in part by the NRF Grant funded by the Korean Government (MSIT) (NRF-2019R1A2C2002358).

Received: *23 September 2020* Accepted: *27 October 2020*
doi: 10.1049/ell2.12047

References

1 Gentry, C.: Fully homomorphic encryption using ideal lattices. In: *Proceedings of the 41st ACM symposium on Theory of Computing–STOC*, pp. 169–178. ACM, New York (2009)
2 Cheon, J.H., et al.: Homomorphic encryption for arithmetic of approximate numbers. In: *International Conference on the Theory and Application of Cryptology and Information Security*, pp. 409–437. Springer, Cham (2017)
3 Graepel, T., Lauter, K., Naehrig, M.: ML confidential: Machine learning on encrypted data. In: *International Conference on Information Security and Cryptology*, pp. 1–21. Berlin, Heidelberg (2012)
4 Park, S., et al.: HE-friendly algorithm for privacy-preserving SVM training. *IEEE Access* **8**, 57414–57425 (2020)
5 Almutairi, N., Coenen, F., Dures, K.: K-means clustering using homomorphic encryption and an updatable distance matrix: Secure third party data clustering with limited data owner interaction. In: *International Conference on Big Data Analytics and Knowledge Discovery*, pp. 274–285. Springer, Cham (2017)
6 Cheon, J.H., Kim, D., Park, J.H.: Towards a practical cluster analysis over encrypted data. In: *International Conference on Selected Areas in Cryptography*, pp. 227–249. Springer, Cham (2019)
7 Lee, J., Lee, D.: Dynamic characterization of cluster structures for robust and inductive support vector clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(11), 1869–1874 (2006)
8 Kim, K., Son, Y., Lee, J.: Voronoi cell-based clustering using a kernel support. *IEEE Trans. Knowledge Data Eng.* **27**(4), 1146–1156 (2014)
9 Cheon, J.H., et al.: Bootstrapping for approximate homomorphic encryption. In: *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pp. 360–384. Springer, Cham (2018)
10 Park, S., Hah, J., Lee, J.: Inductive ensemble clustering using kernel support matching. *Electron. Letters* **53**(25), 1625–1626 (2017)
11 Cheon, J.H., et al.: Numerical method for comparison on homomorphically encrypted numbers. In: *International Conference on the Theory and Application of Cryptology and Information Security*. Springer, Cham (2019)
12 Ultsch, A.: Fundamental clustering problems suite (FCPS). Technical report, University of Marburg (2005)

Enhancement of sensing characteristics of Polydimethylsiloxane-based capacitive force sensor by introducing conductive polymer to dielectric layer

Yasumin Siangkhio,[1] Adirek Rangkasikorn,[1] Narin Tammarugwattana,[2] Navaphun Kayunkid,[1,✉] Sukittaya Jessadaluk,[1] Sakon Rahong,[1] Supamas Wirunchit,[1] and Jiti Nukeaw[1]

[1]*College of Nanotechnology, King Mongkut's Institute of Technology Ladkrabang, Ladkrabang, Bangkok, Thailand*

[2]*Faculty of Engineering, Department of Instrumentation and Control Engineering, King Mongkut's Institute of Technology Ladkrabang, Ladkrabang, Bangkok, Thailand*

✉Email: navaphun.ka@kmitl.ac.th

A capacitive force sensor is one of the electronics components used in several electronic devices and applications. An improvement of sensing characteristics of the sensor, for example sensitivity and response time, becomes an interesting research topic. The alternative approach to enhance the sensitivity and response time of polydimethylsiloxane-based capacitive force sensors is proposed by introducing poly(3,4-ethylenedioxythiophene) polystyrene sulphonate, a conductive polymer, into polydimethylsiloxane active layer. Two sensors using different active layers, (i) polydimethylsiloxane (conventional sensor) and (ii) poly(3,4-ethylenedioxythiophene) polystyrene sulphonate mixed polydimethylsiloxane (modified sensor), were fabricated and characterised to reveal the sensing enhancement. Interestingly, the modified sensor shows the significant increase in the sensitivity from 0.7 to 1.14 kPa^{-1} (+62.86%) and the shortening response time from 1.55 to 0.43 s (−72.26%) with respect to the conventional sensor. In addition, the deterioration in elastic behaviour and the faster charge–discharge behaviour observed from the poly(3,4-ethylenedioxythiophene) polystyrene sulphonate mixed polydimethylsiloxane film indicate the better deformation and charge transport than that from polydimethylsiloxane film. Therefore, it can be concluded that the conductive poly(3,4-ethylenedioxythiophene) polystyrene sulfonate

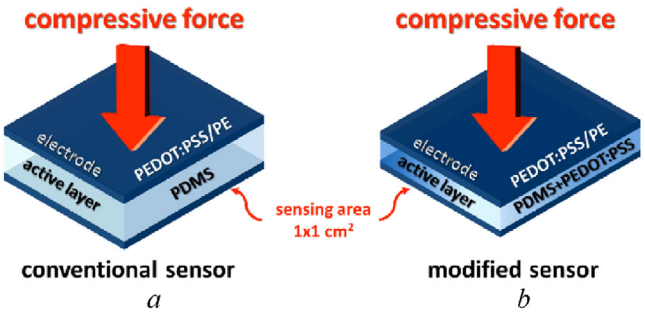


Fig. 1 *Information of materials and structure used to fabricate capacitive force sensors prepared by different active layers: (a) pure PDMS (conventional sensor) and (b) PEDOT:PSS-mixed PDMS (modified sensor)*

additive plays the role of mechanical and electrical modification of the polydimethylsiloxane active layer leading to the enhancement in sensitivity and response time of the polydimethylsiloxane-based capacitive force sensor.

Introduction: Capacitive force sensing device, the device changed its capacitance due to the exerting force, commonly uses in several fields of applications, for example health care monitoring, electronics as well as robotics [1]. The improvement of sensing characteristics, for example sensitivity and response time, of such device is one of the keys to achieving high efficiency in selective applications.

In the sensitivity point of view, the change in capacitance with the applying force directly relates to an active layer's deformation. Therefore, the modification of the active layer's mechanical properties is a crucial approach to enhance the sensitivity of the capacitive force sensing devices. One of the materials famously used as an active layer of such a sensor is polydimethylsiloxane (PDMS) due to its excellent properties, for example high flexibility, good thermal and chemical stability, and low toxicity. Several methods have been proposed to modify the mechanical properties of a PDMS-active layer, for example controlling the numbers of cross-linkage in the PDMS layer [2] as well as introducing a porous structure into the PDMS layer [3]. However, the above-mentioned approaches bring the consequential drawbacks in uniformity and durability of the sensing devices.

From the response time viewpoint, the capacitive force sensor's rapid response is related to both mechanical and electrical properties of the active layer. To obtain the short response/recovery time, the active layer's high elastic behaviour and good electrical conductivity are required. However, the PDMS layer with high elasticity consequently exhibits a reduction in the sensitivity of the sensor. Moreover, PDMS that is generally an insulator cannot provide good electrical conductivity. Therefore, it is a challenging topic to find a way to improve, at the same time, both sensitivity and response/recovery time of the capacitive force sensor.

Therefore, an alternative approach to enhance the sensing characteristics of capacitive force sensing devices is proposed by introducing the conductive polymers (poly(3,4-ethylenedioxythiophene) polystyrene sulfonate), PEDOT:PSS) into the cross-linked PDMS active layer. The sensitivity and response/recovery time of the capacitive force sensing device are used to compare the performance of sensors prepared by different active layers; pure PDMS (conventional device) and PEDOT:PSS mixed PDMS (modified device). Additionally, the changes in viscoelastic mechanical and electrical properties of both PDMS and PEDOT:PSS-mixed PDMS films are investigated in order to reveal the origin of the enhancement in sensing characteristics of the sensor.

Fabrication of PDMS-based capacitive force sensors: The simple structure of a capacitor, consisting of a dielectric layer sandwiched by two parallel electrodes, as shown in Figure 1, is employed to fabricate the force-sensing device. The conductive polymer PEDOT:PSS (PH 1000, Ossila Ltd.) coated on polyester film was used as flexible electrodes, while dielectric polymer PDMS (Sylgard 184 silicone elastomer, Dow Inc.) was used as the active layer of the sensor. For the conventional sensor, the PDMS elastomer ratio to curing agent is controlled at 30:1 by weight. For the modified sensor, the PEDOT:PSS was added to PDMS

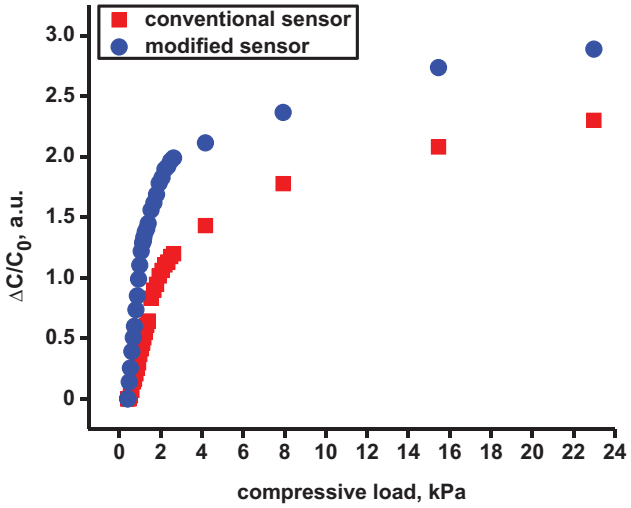


Fig. 2 *Relation between the change in capacitance of conventional (pure PDMS) and modified (PEDOT:PSS-mixed PDMS) sensors with an applied compressive load*

Table 1. Comparison of sensitivity obtained from the capacitive force sensor fabricated with respect to some literatures

Dielectric layer	Sensitivity (kPa^{-1})	Compressive load (kPa)
Pyramid PDMS [4]	0.19	0–3
Nanofiber (0.2% CNT) [5]	0.32	0–1.6
30:1 PDMS	0.70	0–2.5
30:1 PDMS mixed PEDOT:PSS	1.14	0–2.5

elastomer with a proportion of 10 μL:1 g and then mixed to the curing agent with a similar ratio used to prepare the conventional sensor. The active layer was assembled between two parallel electrodes via spin coating (1000 rpm, 60 s) then both sensors were annealed in a heating oven at the temperature of 100°C for 24 h to allow the active layer to be fully cured. The active area of fabricated sensors is fixed at 1 cm × 1 cm.

Sensitivity and response/recovery time of conventional and modified capacitive force sensors: Sensing characteristics of sensors prepared by PDMS and PEDOT:PSS-mixed PDMS as the active layers were measured by Precision LCR Meter (E4980A, Agilent Technologies Inc.). The correlation between the change in capacitance and the applied compressive load obtained from the sensors is exhibited in Figure 2. It should be noted that the sensitivity of the capacitive force sensor is calculated via the following equation:

$$S = \frac{\Delta C/C_0}{\Delta L_c}, \tag{1}$$

where S is sensitivity, ΔC is a change in capacitance, C_0 is an initial capacitance without the compressive load and ΔL_c is a change in compressive load exerting on the sensor. Therefore, the sensitivity can be obtained from a slope in Figure 2. Although both sensors present two regimes of the sensitivity, (i) 0 to 2.5 kPa and (ii) 2.5 to 25 kPa, the sensitivity in the low compressive load is of interest due to the high sensitivity enhancement of modified sensor with respect to the conventional sensor. From Figure 2, the sensitivity of the conventional sensor is 0.70 kPa^{-1}. By adding PEDOT:PSS into PDMS active layer, the sensitivity is increased to 1.14 kPa^{-1}, equivalent to the increment of 62.86%. Additionally, the sensitivity obtained from the capacitive force sensor used PEDOT:PSS-mixed PDMS as the active layer is comparable to those proposed by several modificative approaches, that is, introducing either periodic structure or additive materials to the active layer (see Table 1).

For the response/recovery time measurement, the compressive load of 0.75 kPa was applied to the sensors, and then the load was removed

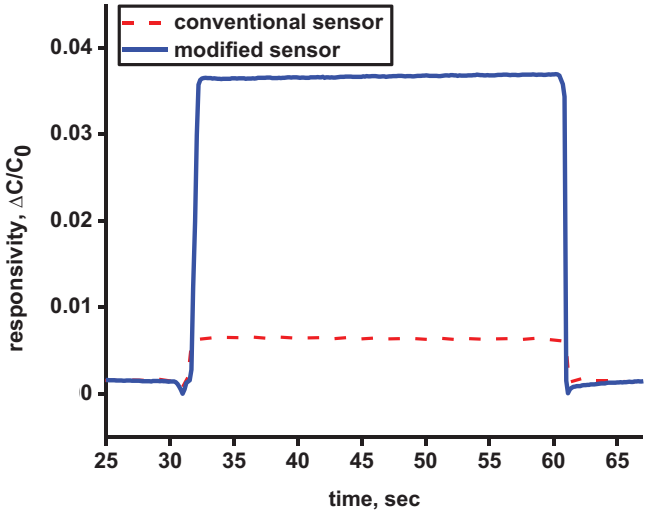


Fig. 3 *Response behaviour of conventional and modified capacitive force sensors induced by a compressive load of 0.75 kPa*

Table 2. Response and recovery time obtained from capacitive force sensors used PDMS (conventional sensor) and PEDOT:PSS-mixed PDMS (modified sensor) as the active layers

Dielectric layer	Response time (s)	Recovery time (s)
PDMS	1.55	0.88
PDMS mixed PEDOT:PSS	0.43	0.34

from the sensors after 30 s. Note that the response and recovery times are determined from an interval that the sensor requires to rise (or fall) the output signal from 10% to 90% (or 90–10% in case of recovery time) of the full scale. The time–response behaviour obtained from conventional and modified sensors with compressive load and their response and recovery time are shown in Figure 3 and Table 2, respectively. In Figure 3, modified sensor exhibits faster sensing response than the conventional one. The significant reduction in response time from 1.55 to 0.43 s (equivalent to the reduction of 72.26%) and recovery times from 0.88 to 0.34 s (equivalent to the reduction of 61.36%) are observed in the modified sensor compared to the conventional sensor (see Table 2). The enhancements in both sensitivity and response/recovery time observed in the modified sensor confirm that the introduction of PEDOT:PSS, the conductive polymer, into the PDMS active layer becomes an alternative approach to modify sensing characteristic of PDMS-based capacitive force sensor.

Viscoelastic mechanical properties of PDMS and pedot:PSS mixed PDMS films: Since the sensitivity of capacitive force sensor is directly related to the deformation of the active layer induced by an exerting force, understanding the mechanical properties of the active layer is a key to reveal the origin of the sensitivity enhancement observed in the modified sensor. The viscoelastic properties of PDMS and PEDOT:PSS-mixed PDMS films with the same preparation condition as used in the sensor were investigated by strain sweep measurement (HAAKE MARS rheometer, Thermo Fisher Scientific Inc.). The compressive force of 2 N (equivalent to the compressive load of 2 kPa) was applied to the sensor during the measurement. The viscoelastic parameters, for example elastic modulus (G'), loss modulus (G'') and tanδ, of PDMS and PEDOT:PSS-mixed PDMS films are presented in Figure 4(a) and (b). It is worth noting that a value of G' directly represents elastic behaviours of viscoelastic materials including an ability to deform by applying force [6, 7]. The material with higher value of G' infers the lower shape deformity. On the other hand, a value of G'' presents viscous behaviours of viscoelastic materials such as an ability to flow by applied force. Figure 4(a) shows the significant decrease in G' value observed in PEDOT:PSS-mixed PDMS film compared to that of PDMS film. The reduction in G' reveals that, with the same applying force, the larger shape deformation takes place in the modified film than in the conventional film [8]. Therefore, the larger deformation of active layer leads to the larger change in capacitance of the capacitive force sensor. As consequence, the sensitivity of the sensor prepared by PEDOT:PSS-mixed PDMS as the active layer is enhanced. The decrease in G′can be explained by the presence of PEDOT:PSS in active layer that disturbs the cross-linking formation in PDMS. Therefore, the declination of cross-link density leading to the deterioration of elastic modulus is observed in PEDOT:PSS-mixed PDMS film [9].

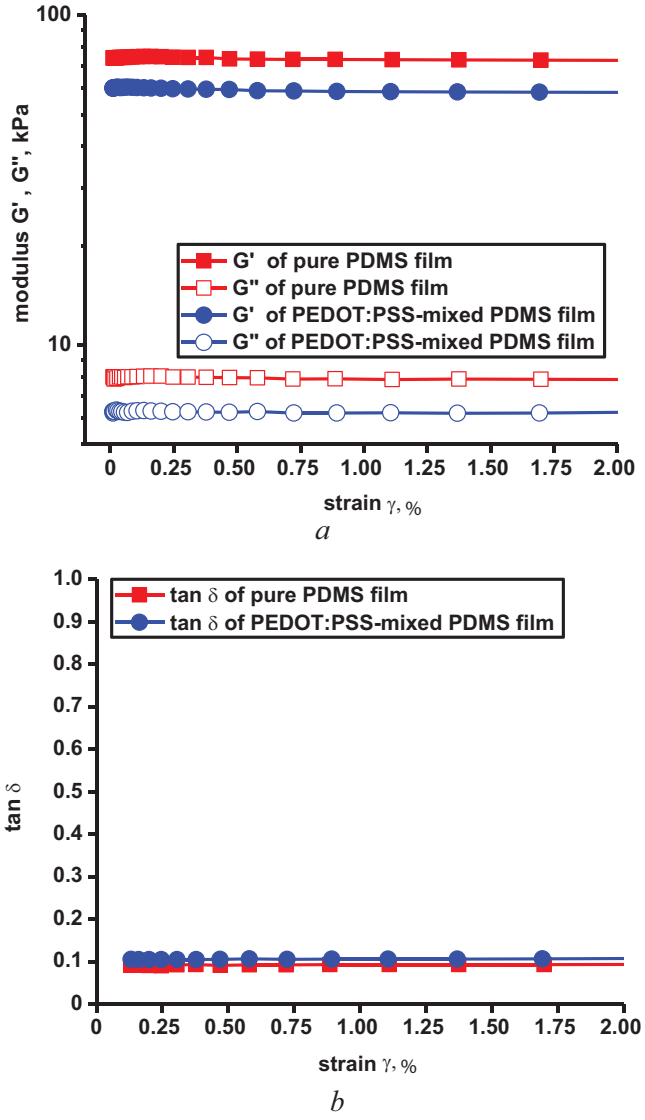


Fig. 4 *Strain sweep measurements are providing viscoelastic parameters: (a) elastic modulus (G′) and loss modulus (G″) of PDMS and PEDOT:PSS-mixed PDMS films; (b) tan δ obtained from PDMS and PEDOT:PSS-mixed PDMS films*

In contrast to the sensitivity of the capacitive force sensor that principally relies on the deformation of the active layer, the response/recovery time of such a sensor depends on how fast the deformation of the active layer reacts after a force is applied. In viscoelastic materials, the rapid-deforming response is observed in the material with elastic-dominant behaviour. Therefore, tanδ, representing the ratio of G'' to G', is employed to anticipate the dominant behaviour of PDMS and PEDOT:PSS-mixed PDMS films [10, 11]. The low value of tanδ indicates the elastic-dominant material, whereas the high value of tanδ refers to the viscous-dominant material. Figure 4(b) exhibits that both PDMS and PEDOT:PSS-mixed PDMS films behave as elastic-dominant materials. Moreover, the insignificant difference of tanδ obtained from each film (Δtanδ < 0.02) suggests that the mechanical time–response behaviour of both PDMS and PEDOT:PSS-mixed PDMS films are indistinguishable. However, the enhancement in response/recovery times is evidently obtained from the force sensors used PEDOT:PSS-mixed PDMS film as the active layer. Therefore, it should have another key factor that

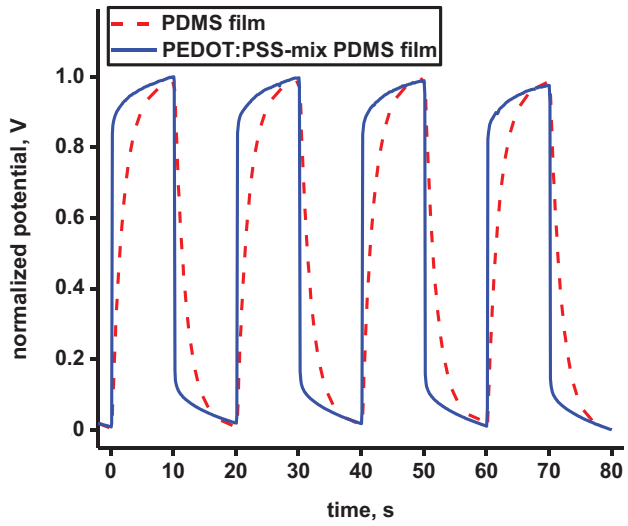


Fig. 5 *Charge–discharge characteristics obtained from PDMS film and PEDOT:PSS-mixed PDMS film*
Note*: Applied current and charge–discharge intervals are fixed at 1 nA and 10 s, noting that the data were normalised to clearly see the difference in charge–discharge characteristics.*

directly effects on the time-response behaviour of the capacitive force sensor.

Charge–discharge behaviour obtained from conventional and modified sensors: Two time-dependent factors that determines the response time of capacitive force sensor are (i) the mechanical-related response time, the interval that the active layer uses to deform to the final thickness, and (ii) electrical-related response time, the interval that the capacitor uses to change to the final capacitance. Since the mechanical response behaviour obtained from PDMS and PDMS-mixed PEDOT:PSS films demonstrate an insignificant difference, the major factor of the enhancement in response/recovery time should be related to the electrical aspect. To investigate the intrinsically charge transport behaviour of PDMS and PDMS-mixed PDMS films, a charge–discharge measurement was carried out without any applying compressive load by a potentiostat (Autolab PGSTAT302N, Metrohm AG Ltd.). The applied current and charge–discharge switching time during the measurement are fixed at 1 nA and 10 s, respectively. Figure 5 exhibits significantly faster charge–discharge characteristics in PEDOT:PSS mixed PDMS film than in PDMS film. This result suggests that the ability to change the capacitance of PEDOT:PSS-mixed PDMS film is much faster than the conventional PDMS film. As a consequence, the shortening in the response/recovery time is obtained from the modified sensor. The improvement of charge transportation in PEDOT:PSS-mixed PDMS film is originated from the existence of conducting charge pathway by PEDOT:PSS dispersed in PDMS film. Therefore, the roles of PEDOT:PSS in PDMS active layer are (i) to disturb the density of cross-linkage in PDMS, leading to better deformation of the modified layer and (ii) to increase the conductivity of PDMS, providing the enhancement of charge transportation in the modified layer.

Conclusions: The sensitivity and response/recovery time of the PDMS-based capacitive force sensor are significantly improved by adding PEDOT:PSS into PDMS active layer. The existence of PEDOT:PSS in PDMS active layer leads not only to the deterioration in elastic modulus but also to the improvement of a charge–discharge characteristic of PEDOT:PSS-mixed PDMS layer. As a consequence, the increase in sensitivity of 62.86% as well as shortening in response time of 72.26% with respect to the conventional sensor are obtained from the modified sensor. Therefore, the introduction of conducting polymer into the PDMS active layer could become an alternative approach to enhance the sensing characteristics of the PDMS-based capacitive force sensor.

Acknowledgements: The first author was supported by the master's degree scholarship from King Mongkut Institute of Technology Ladkrabang Research Fund (grant KREF016326). This research was supported by the Program Management Unit for Human Resources & Institutional Development, Research and Innovation (grant B05F630020). The HAAKE MARS rheometer used in viscoelastic measurements was supported by LMS Instruments Co., Ltd., with the valuable suggestions from Dr. Preeyanuch Prasopnatra.

Received: *23 September 2020* Accepted: *23 November 2020*
doi: 10.1049/ell2.12064

References

1 Bauer, S., et al.: 25th anniversary article: a soft future: from robots and sensor skin to energy harvesters. *Adv. Mater.* **26**(1), 149–162 (2014)
2 Wang, Z., Volinsky, A.A., Gallant, N.D.: Crosslinking effect on polydimethylsiloxane elastic modulus measured by custom-built compression instrument. *J. Appl. Polym. Sci.* **131**(22) (2014), https://doi.org/10.1002/app.41050
3 Kang, S., et al.: Highly sensitive pressure sensor based on bioinspired porous structure for real-time tactile sensing. *Adv. Electron. Mater.* **2**(12) (2016), https://doi.org/10.1002/aelm.201600356
4 Pignanelli, J., et al.: A comparative analysis of capacitive-based flexible pdms pressure sensors. *Sens. Actuator A Phys.* **285**, 427–436 (2019)
5 Yang, X., Wang, Y., Qing, X.: A flexible capacitive sensor based on the electrospun Pvdf nanofiber membrane with carbon nanotubes. *Sens. Actuator A Phys.* **299** (2019), https://doi.org/10.1016/j.sna.2019.111579
6 Hirschl, C., et al.: Determining the degree of crosslinking of ethylene vinyl acetate photovoltaic module encapsulants – a comparative study. *Sol. Energy Mater. Sol. Cells* **116**, 203–218 (2013)
7 Reinitz, S.D., et al.: Dynamical mechanical analysis as an assay of cross-link density of orthopaedic ultra high molecular weight polyethylene. *Polym. Test.* **45**, 174 (2015)
8 Hwang, J., et al.: Unveiling viscoelastic response of capacitive-type pressure sensor by controlling cross-linking density and surface structure of elastomer. *ACS Appl. Polym. Mater.* **2**(6), 2190–2198 (2020)
9 Luo, R., et al.: A simple strategy for high stretchable, flexible and conductive polymer films based on PEDOT:PSS-PDMS blends. *Org. Electron.* **76** (2020), https://doi.org/10.1016/j.orgel.2019.105451
10 Chim Ya Hua, Mason Louise M., Rath Nicola, Olson Michael F., Tassieri Manlio, Yin Huabing A one-step procedure to probe the viscoelastic properties of cells by Atomic Force Microscopy. *Sci. Rep.* **8**, (1), (2018). https://doi.org/10.1038/s41598-018-32704-8
11 Dalla Monta, A., et al.: Using thickness-shear mode quartz resonator for characterizing the viscoelastic properties of PDMS during cross-linking, from the liquid to the solid state and at different temperatures. *Sens. Actuator A Phys.* **280**, 107–113 (2018)

A unipolar thin-film transistor-based amplifier with enhanced DC offset suppression

Yuming Xu,[1] Bin Li,[1,2] Wei Zhong,[1,2] Sunbin Deng,[3] Houbo Fan,[1,2] Zhaohui Wu,[1,2] Fion Sze Yan Yeung,[3] Hoi Sing Kwok,[3] and Rongsheng Chen[1,2,✉]
[1]*School of Microelectronics, South China University of Technology, Guangzhou, China*
[2]*Pazhou Lab, Guangzhou, China*
[3]*State Key Laboratory on Advanced Displays and Optoelectronics Technologies, Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong*
✉Corresponding author: Rongsheng Chen, Email: chenrs@scut.edu.cn

This letter proposes a unipolar thin-film transistor (TFT)-based amplifier design that can provide enhanced DC offset suppression. The amplifier consists of an input AC coupling network and a capacitor bootstrap

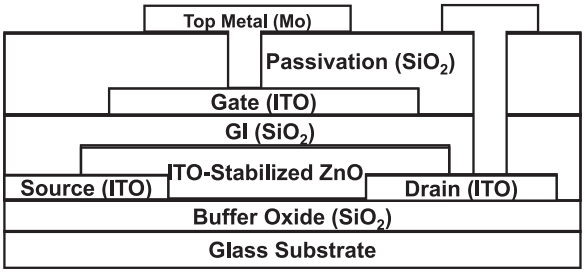


Fig. 1 *Structure of the metal-oxide thin-film transistor (TFT) used*

amplifier. By making the first pole of the input AC-coupling network coincide with that of the capacitor bootstrap amplifier, a DC offset suppression capability of 40 dB/dec, which is 2× higher than that of the prior arts, can be obtained. An implementation is also proposed and a Rensselaer Polytechnic Institute (RPI) TFT model whose parameters are turned to fit the measured data of our metal-oxide TFT is used to simulate the circuit. Simulation results show that the implementation has 31 dB midband gain, 0.5 Hz–5 kHz bandwidth, 300 μW power consumptions, and the expected 40 dB/dec DC offset suppression. This design is suitable for applications that need to block large DC interference, such as sensor readout amplifiers.

Introduction: Thin-film transistor (TFT) technologies provide attractive characteristics including flexibility, transparency, lightweight, ultra-thin dimensions, stretchability, and capability for large-area low-cost fabrication compared to the traditional silicon complementary metal–oxide–semiconductor (CMOS) technology [1]. Thanks to these advantages, they are promising in flexible/wearable sensor systems [2–4]. Using TFT technology, the readout circuit can be fabricated together with the flexible sensor on the same ultra-thin flexible substrate that can be attached to the surface of everyday objects (including the human body). This is beneficial to improve patient comfort and reduce motion artefacts [2, 4].

The sensor output signal usually contains a large DC offset on which a useful small signal is superimposed. Therefore, the sensor readout amplifier needs to have a DC signal suppression capability. In the existing TFT-based works [2–4], this DC suppression capability is realised by the input AC-coupling network (which will be detailed in the next section). The input AC-coupling network has one pole that provides a DC suppression capability of 20 dB/dec, which is slightly insufficient in some demanding applications and can be further improved. To solve this problem, this letter proposes a new design that is composed of an input AC-coupling network and a capacitor bootstrap amplifier and can increase the DC suppression capability by two times to 40 dB/dec.

It is worth noting that TFT technology generally lacks high-performance complementary devices. In this letter, the circuits are based on *n*-type only transistors because a metal-oxide TFT technology is used.

Device and simulation model: To verify the proposed design, SmartSpice simulation based on the Rensselaer Polytechnic Institute (RPI) Level = 36 TFT model whose parameters are turned to fit the measured data of our metal-oxide TFT is carried out. The structure of the device is shown in Figure 1. Fabrication process and device characteristics are detailed elsewhere [5–7]. Figure 2 shows the measured and simulated DC and capacitance characteristics of the device. Root mean square (RMS) error between measured and simulated curves is calculated as 1.22%. This mismatch comes from the triode region and is due to the poor contact of the source/drain electrode. It can be solved by adjusting electrode materials and optimising fabrication process. All TFTs in the proposed circuits, however, are biased in the saturation and cut-off region. So the triode region mismatch will not affect the credibility of simulation. Capacitors with metal-insulator-metal (MIM) structure can be formed by sandwiching the source/drain, the gate SiO_2, the gate metal, the passivation SiO_2, and the top metal layers. It has the typical capacitance of 40 nF/cm^2.

Conventional designs: In prior arts, the input AC-coupling network is followed by a resistor load amplifier [2], diode-load amplifier [3], and positive feedback amplifiers [4]. Their circuit concepts and gain

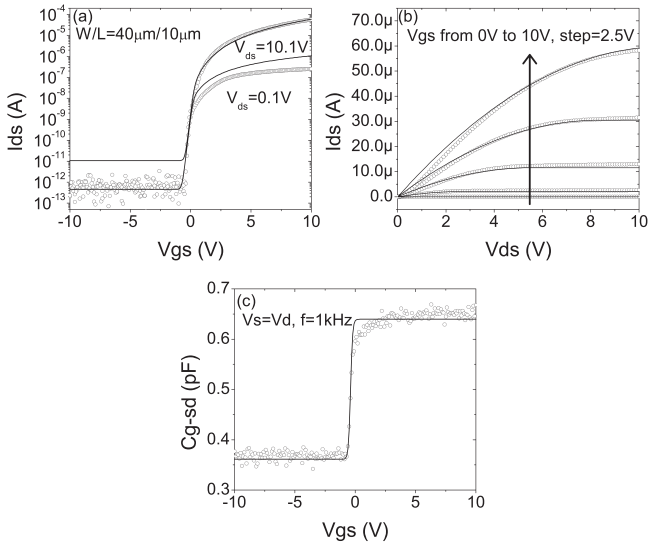


Fig. 2 *Typical measured (hollow dots) and simulated (lines) (a) transfer, (b) output, and (c)* C–V *characteristics of the TFT used*

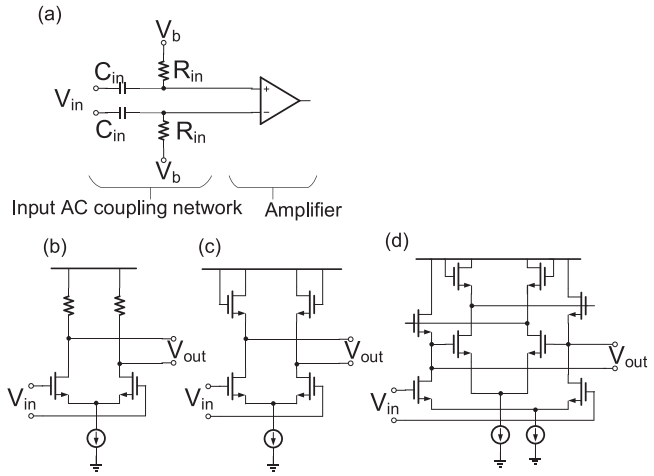


Fig. 3 *Circuit concepts of conventional designs: (a) The overall amplifier, (b) resistor load amplifier [2], (c) diode-load amplifier [3], and (d) positive feedback amplifier [4]*

response are, respectively, shown in Figures 3 and 4. The input AC-coupling network acts as a first-order resistor-capacitor (RC) high-pass filtre. Resistor R_{in} is used to provide bias voltage V_b to the amplifier. Capacitor C_{in} is used to couple input small signals. Transfer function (gain) of the coupling network is given by

$$|H_1(s)| = \frac{SR_{in}C_{in}}{1+SR_{in}C_{in}}. \quad (1)$$

Note that Equation (1) has a zero $f_{z0} = 0$ and a pole $f_{p0} = \frac{1}{2\pi R_{in}C_{in}}$. When $f < f_{p0}$, the gain increases by 20 dB/dec. Due to the limited output impedance of the amplifier, a second pole $f_{p2} = \frac{1}{2\pi r_O C_L}$ is introduced at the output node, where r_O is the output impedance of the amplifier and C_L is the load capacitance of the measurement setup. In these designs, DC suppression capability is achieved by the single pole f_{p0} and thus limited to 20 dB/dec.

Schematic of the proposed amplifier: Figure 5 shows the schematic of the proposed amplifier. It consists of an input AC-coupling network and a capacitor bootstrap amplifier [8]. The capacitive bootstrap amplifier consists of T1 (input drive transistor), T2 (load transistor), bootstrap capacitor C_f, and resistor R_f. C_f and R_f form a positive feedback network to increase the gain. Its transfer function can be given by

$$|H_2(s)| = g_{m1}\left\{\frac{1}{(1-A_f)g_{m2}} // r_{O1} // r_{O2}\right\}. \quad (2)$$

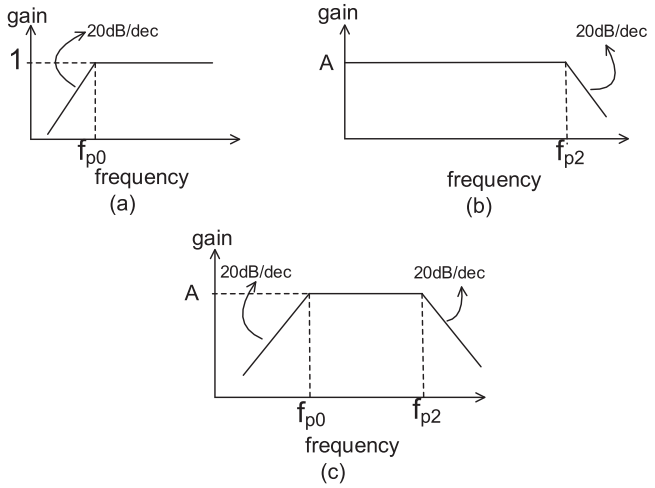


Fig. 4 *Gain response of the conventional designs: (a) Input AC-coupling network, (b) main amplifiers, and (c) overall amplifier*

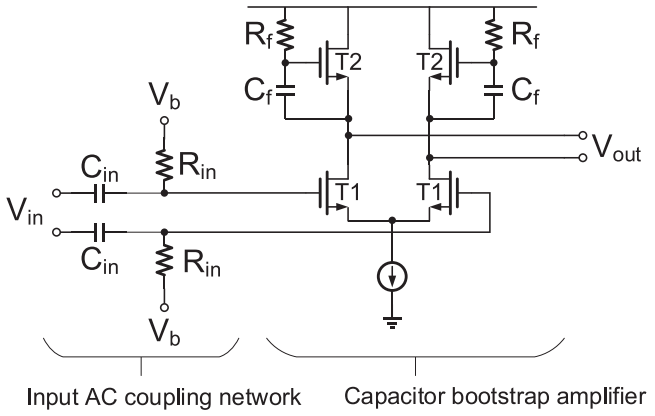


Fig. 5 *Schematic of the proposed amplifier*

Where A_f is the factor of the positive feedback network,

$$A_f = \frac{SR_fC_f}{1+SR_fC_f}. \tag{3}$$

At low frequencies, $A_f = 0$, bootstrap is failed and the gain is g_{m1}/g_{m2}. At high frequencies, $A_f = 1$, bootstrap is enabled and the gain is boosted to $g_{m1}(r_{o1}//r_{o2})$. Note that the gain at high frequencies is the same as that of the CMOS amplifier with current source load. This indicates the significant gain improvement of the bootstrap structure. Combining Equations (2) and (3), we can get

$$|H_2(s)| = \frac{g_{m1}r_{O1}r_{O2}\left(1+sR_fC_f\right)}{g_{m2}r_{O1}r_{O2}+(r_{O1}+r_{O2})\left(1+sR_fC_f\right)}. \tag{4}$$

Note that Equation (4) has a zero $f_{z1} = \frac{1}{2\pi R_fC_f}$ and a pole $f_{p1} = \frac{g_{m2}(r_{O1}//r_{O2})+1}{2\pi R_fC_f}$. When $f_{z1} < f < f_{p1}$, the gain increases by 20 dB/dec. The gain response of the capacitor bootstrap amplifier is shown in Figure 6(b). Through proper design, f_{p0} and f_{p1} can be matched, and the amplifier can provide 40 dB/dec DC signal suppression capability. At this case, it needs to meet

$$\frac{1}{R_{in}C_{in}} = \frac{g_{m2}(r_{O1}//r_{O2})+1}{R_fC_f}. \tag{5}$$

Figure 6(c) shows the overall gain response of the proposed amplifier. By designing appropriate circuit dimensions with the help of simulation, Equation (5) can be satisfied. We utilise the low-frequency pole of the capacitive bootstrap amplifier (f_{p1}) to improve DC suppression capability. To the best of our knowledge, this method is described in this letter for the first time.

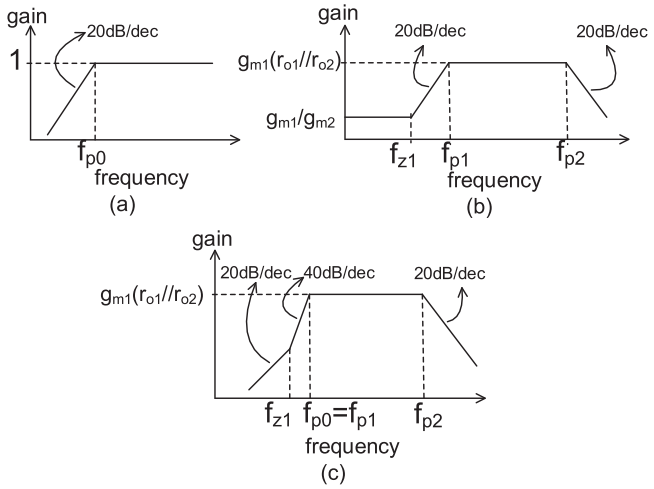


Fig. 6 *Gain response of the proposed amplifier: (a) Input AC coupling network, (b) capacitor bootstrap amplifier, and (c) overall amplifier*

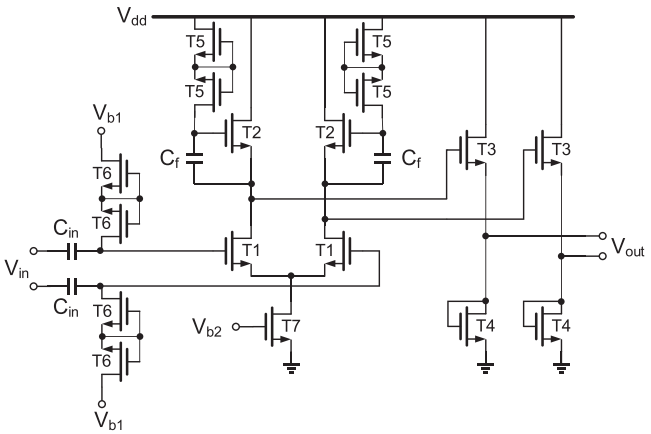


Fig. 7 *Implementation of the proposed amplifier*

Table 1. Circuit dimensions of the proposed amplifier. The units of transistors and capacitors are μm/μm and pF, respectively

Device	Value	Device	Value
T1	500/10	T6	10/50
T2	10 /10	T7	1000/10
T3	1000 /10	C_{in}	160
T4	100 /10	C_f	650
T5	10/100		

Implementation of the proposed amplifier: Figure 7 shows an implementation of the proposed amplifier. Circuit dimensions are shown in Table 1. In sensor applications, the poles f_{p0} and f_{p1} are usually located at low frequencies. This requires large R_{in} and R_f. However, in TFT technologies, resistors are usually realised by the indium tin oxide (ITO) layer with a sheet resistance less than 100 Ω [9]. So it is a challenge to realise large resistors in a small area. To solve this, we use the pseudo-resistor structure [5], where M5/M6 are 0-Vgs connected (cut-off) and stacked in opposite polarity. This structure can provide large (typically larger than 100 MΩ) and symmetrical resistance. Capacitors are realised by the MIM structure. C_{in} and C_f occupy an area of 0.4 and 1.625 mm^2, respectively. Source followers (T3 and T4) are added to drive the measurement setup. Biasing voltages V_{dd}, V_{b1}, and V_{b2} are, respectively, 10, 2.5, and 1 V. Output of the amplifier is connected to a 1 MΩ//12 pF load which mimics the loading effect of an oscilloscope.

Simulation results: Gain responses of the input AC-coupling network, the capacitor bootstrap amplifier, and the overall amplifier are shown

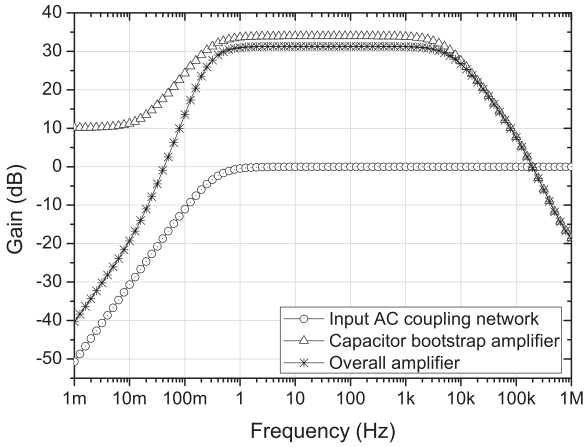


Fig. 8 *Simulated gain responses of the input AC-coupling network, the capacitor bootstrap amplifier, and the overall amplifier*

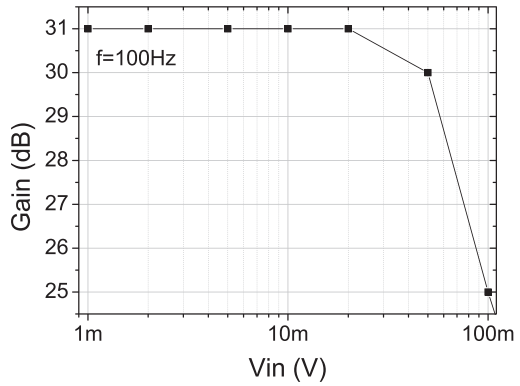


Fig. 9 *Simulated gain versus input magnitude plot of the proposed amplifier*

Table 2. Comparison to the state-of-the-art counterparts

	[2]	[3]	[4]	Proposed method
DC suppression (dB/dec)	20	20	20	40
Gain (dB)	20	22	25	31
Bandwidth (Hz)	0.05–200	3k	3.7–5.4k	0.5–5k
Power (μW)	11000	32	1300	300

in Figure 8. Note that C_{in} and the parasitic gate capacitance of the input differential pair (C_{g1}) form a voltage divider, which makes the gain of the overall amplifier slightly lower than that of the capacitor bootstrap amplifier. The high-pass corner, which is determined by the poles f_{p0} and f_{p1}, is designed to ∼0.5 Hz. The low-pass corner, which is determined by the pole f_{p2}, is ∼5 kHz. Thus, this implementation is suitable for applications such as bio-potential measurements in terms of bandwidth [10]. As expected, a 40 dB/dec slope of the gain curve is seen at the frequencies below 0.5 Hz. The gain is attenuated to –20 dB at 10 mHz (∼0 dB in conventional designs). This implementation exhibits 31 dB midband gain and consumes 30 μA at 10 V supply. To investigate the large-signal characteristics, Figure 9 shows the gain versus input magnitude plot. When input magnitude increases to 50 mV, the gain reduces by 1 dB, which refers to the output range of 3.16 V V_{pp}. Table 2 shows comparison between the proposed and conventional designs. The proposed design significantly outperforms previous works in terms of DC offset suppression capability and also achieves reasonable gain, bandwidth, and power consumption.

Conclusion: A unipolar TFT-based amplifier design composed of an input AC-coupling network and a bootstrap amplifier is introduced in this letter in order to improve the DC offset suppression capability of the existing works. It is found that by making the first pole of the input AC-coupling network coincide with that of the capacitor bootstrap amplifier, we can achieve a DC offset suppression capability of 40 dB/dec, which is 2× higher than that of the existing works. Simulation results of an implementation are consistent with the theoretical analysis. The proposed design can find potential applications in sensor readout circuits that need to block large DC interference.

Acknowledgements: This work was supported in part by the Key-Area Research and Development Program of Guangdong Province under Grant 2019B010140002, in part by the National Natural Science Foundation of China under Grant 61571196, in part by the Science and Technology Program of Guangdong Province under Grant 2019A050503005, in part by the Natural Science Foundation of Guangdong under Grant 2019A1515011820, in part by the Program for Guangdong High-level Talents, in part by the State Key Laboratory of Advanced Displays and Optoelectronics Technologies under Grant ITC-PSKL12EG02, and in part by the China Postdoctoral Science Foundation under Grant 2020M672628.

Author contributions: Conceptualisation, writing-original draft, writing-review and editing: Yuming Xu. Project administration: Bin Li. Project administration: Wei Zhong. Project administration: Sunbin Deng. Project administration: Houbo Fan. Project administration: Zhaohui Wu. Project administration: Sze Yan Fion Yeung. Project administration: Rongsheng Chen.

Received: *30 October 2020* Accepted: *18 December 2020*
doi: 10.1049/ell2.12066

References

1 Myny, K.: The development of flexible integrated circuits based on thin film transistors. *Nat. Electron.* **1**(1), 30–39 (2018)
2 Moy, T. et al.: An EEG acquisition and biomarker-extraction system using low-noise-amplifier and compressive-sensing circuits based on flexible, thin-film electronics. *IEEE J. Solid-State Circuits* **52**(1), 309–321 (2017)
3 Zulqarnain, M. et al.: A 52 μW heart-rate measurement interface fabricated on a Flexible Foil with A-IGZO TFTs. In: IEEE 44th European Solid State Circuits Conference (ESSCIRC), Dresden, pp. 222–225 (2018)
4 Garripoli, C. et al.: A fully integrated 11.2-mm^2 a-IGZO EMG front-end circuit on flexible substrate achieving up to 41-dB SNR and 29-MΩ input impedance. *IEEE Solid-State Circuits Lett.* **1**(6), 142–145 (2018)
5 Xu, Y. et al.: A novel envelope detector based on unipolar metal-oxide TFTs. *IEEE Trans. Circuits Syst. II Express Briefs* **67**(11), 2367–2371 (2020)
6 Xu, Y. et al.: The implementation of fundamental digital circuits with ITO-stabilized ZnO TFTs for transparent electronics. *IEEE Trans. Electron Devices* **65**(12), 5395–5399 (2018)
7 Qin, Y. et al.: Low-power design for unipolar ITO-Stabilized ZnO TFT RFID code generator using differential logic decoder. *IEEE Trans. Electron Devices* **66**(11), 4768–4773 (2019)
8 Marien, H. et al.: Analog building blocks for organic smart sensor systems in organic thin-film transistor technology on flexible plastic foil. *IEEE J. Solid-State Circuits* **47**(7), 1712–1720 (2012)
9 Tarn, Y. et al.: An amorphous-silicon operational amplifier and Its application to a 4-Bit Digital-to-Analog converter. *IEEE J. Solid-State Circuits* **45**(5), 1028–1035 (2010)
10 Yazicioglu, R.F. et al.: A 60 μW 60 nV/$\sqrt{\text{Hz}}$ readout front-end for portable biopotential acquisition systems. *IEEE J. Solid-State Circuits* **42**(5), 1100–1110 (2007)

A simplified zero-voltage switching detection circuit and switching loss dynamic optimisation method for synchronous buck converter

Sizhao Lu, Xiaoting Deng, and Siqi Li✉
Department of Electrical Engineering, Kunming University of Science and Technology, Kunming, P. R. China
✉E-mail: lisiqi@kust.edu.cn

This letter proposes a simplified zero-voltage switching (ZVS) detection circuit and switching loss dynamic optimisation method. The proposed circuit and tuning method can automatically adjust the switching frequency of a synchronous buck converter to work in critical current mode to realise ZVS under a wide load range. The advantages of the proposed circuit and method are that ZVS detection and switching loss dynamic optimisation can be realised by only one ZVS detection circuit, which simplifies the detection circuit, further improves the power density and reduces the number of analogue-to-digital conversion channels of the digital signal processor. A prototype of the synchronous buck converter with a rated input voltage of 100 V is established to verify the effectiveness of the proposed ZVS detection circuit and switching loss dynamic optimisation method. By using the proposed detection circuit and optimisation method, the switching loss of the MOSFETs has been greatly reduced and an efficiency improvement of nearly 1% at peak output power has also been achieved.

Introduction: As a widely used step-down converter, synchronous buck converter requires high power density and high efficiency in many applications. It is well known that increasing the switching frequency of the power devices, such as the MOSFETs, can reduce the output inductance and capacitance [1], thus reducing the volume and weight of the converter. However, with the increase of switching frequency, the switching loss of the MOSFETs would be inevitably increased, so zero-voltage switching (ZVS) of the MOSFETs should be guaranteed, which can significantly reduce the switching loss and thus greatly improve the efficiency of buck converter [2]. ZVS operation of the buck converter is usually realised by adding additional resonant circuit [3–5], however, the disadvantages of this method are that the selection of resonant element is complicated [6], and the use of resonant inductors and capacitors induces a large amount of reactive power exchange in the circuit, which increases the circuit conduction loss. In [7], it is stated that the critical current mode (CRM) of the synchronous buck converter is the simplest and most effective method to achieve ZVS since ZVS obtained by CRM does not use any auxiliary resonant components [8, 9]. However, these methods are based on the calculation of component parameters to obtain ZVS operation, which highly depends on the accuracy of component parameters. Once there is a deviation in the component parameter, ZVS may not be guaranteed. In [10], a voltage slope detection circuit is designed for the wireless power transfer system, and ZVS operation is tuned according to the voltage slope detection circuit, which is not affected by the variation or deviation of component parameters. However, the method proposed in [10] is designed for the wireless power transfer system whose tuning method depends on the compensation network, which cannot be directly applied to the buck converter. Moreover, its detection circuit cannot be fully utilised. Its detection circuit is divided into two similar parts, and each has different functions. One part is responsible for sampling the voltage slope of the switch during the turn-on process so that the digital signal processor (DSP) can judge whether ZVS is realised or not. The second part is responsible for the optimisation of the turn-off loss. It is the second part circuit that cannot be fully utilised because the detected values of the second part are invalid until ZVS is realised, which hinders the improvement of the power density. In addition, the sampled analogue signals corresponding to the two detection circuits need to be converted into digital signals, so the detection circuit proposed in [10] requires two DSP analogue-to-digital conversion (ADC) channels. For the multiphase converters, more detection circuit and more ADC channels are required, which make the method proposed in [10] is not preferred.

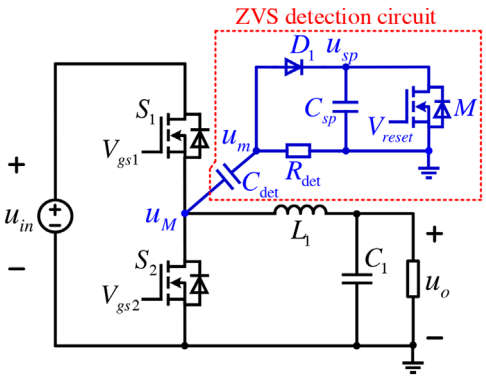


Fig. 1 *The simplified zero-voltage switching (ZVS) detection circuit for synchronous buck*

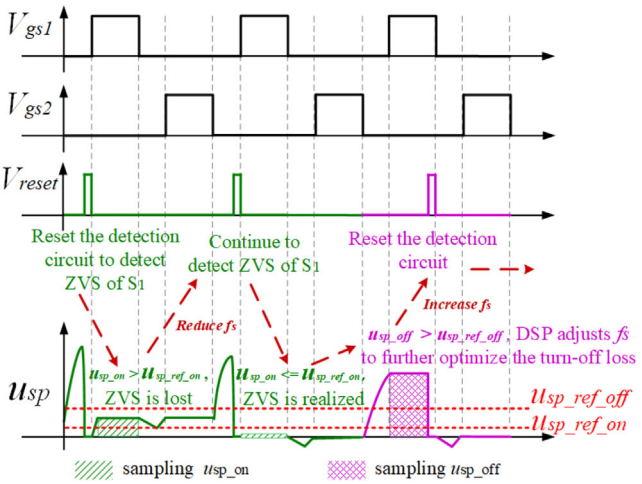


Fig. 2 *A basic ZVS tuning and turn-off loss optimisation process*

In this letter, a simplified ZVS detection circuit and switching loss dynamic optimisation method for the synchronous buck converter is proposed, which can realise ZVS by automatically adjusting the switching frequency to make synchronous buck converter work in CRM and is not dependent on the accuracy of component parameters. By reducing the detection circuit components by half, the power density of the circuit can be improved, and the number of the DSP ADC channels can be reduced. Meanwhile, it is known from [11, 12] that when ZVS is realised in CRM, the turn-on loss of the MOSFETs can be eliminated, but the turn-off loss of the MOSFETs is still high. Therefore, it is necessary to optimise the turn-off loss in order to further improve the system efficiency.

Proposed ZVS detection circuit: The simplified ZVS detection circuit for the synchronous buck converter is shown in Figure 1. It consists of capacitor C_{det}, C_{sp}, resistor R_{det}, diode D_1 and the MOSFET M. Compared with the detection circuit proposed in [10], this new detection circuit reduces the components by half. In the proposed ZVS detection circuit, when a reset signal is applied to the MOSFET M that connected in parallel with the capacitor C_{sp}, the charges stored in the capacitor C_{sp} are discharged to ensure that the voltage slope of S_1 or S_2 can be truly updated on the capacitor C_{sp}. The time of applying the reset signal to the MOSFET M is carefully selected to obtain both ZVS state and the turn-off loss information, so only one circuit is needed to perform ZVS detection and switching loss dynamic optimisation.

The principles of ZVS tuning and switching loss optimisation: A basic ZVS tuning and turn-off loss optimisation process is shown in Figure 2. When detecting ZVS state, the reset signal of the MOSFET M needs to be applied for a short time before driving the main switch S_1 in order to reset the voltage u_{sp} to the initial voltage 0. Then, the ZVS state of S_1 can be judged according to the voltage u_{sp}, which is referred as u_{sp_on}. If the voltage slope is 0 during the turn-on process of S_1, the capacitor C_{sp} will not be charged, and its voltage u_{sp_on} remains at the initial voltage 0, which indicates that ZVS operation of S_1 is realised and vice versa.

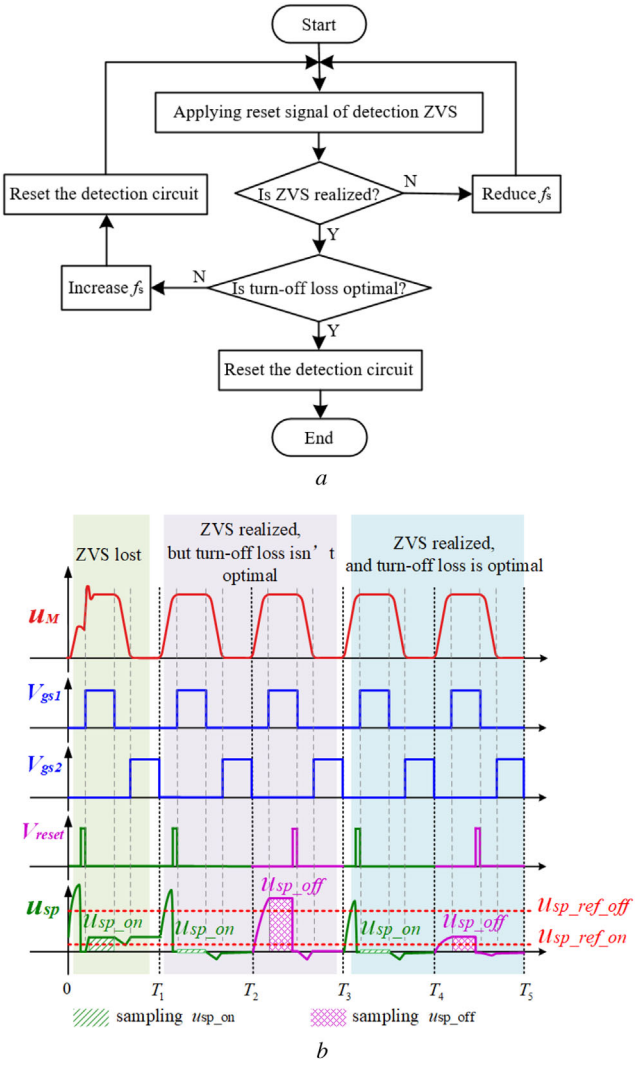


Fig. 3 *The proposed method of ZVS detection and optimisation. (a) The algorithm block diagram of the proposed method, (b) a specific detection and optimisation process of the proposed method*

Considering that there may be slight interference in the circuit, a reference voltage $u_{sp\_ref\_on}$ is set, which is approximately equal to 0. When $u_{sp\_on} <= u_{sp\_ref\_on}$, it is considered that ZVS has been realised.

After ZVS is realised, the turn-off loss should be further optimised. The optimisation principles are as follows: When S_2 is turned off, the negative inductor current flows through the capacitor C_{det} and diode D_1 to charge the capacitor C_{sp}. Obviously, the voltage u_{sp} of the capacitor C_{sp} is proportional to the negative inductor current, and this voltage is held for a period of time in order to sample its value. The holding time starts from the dead time and ends before S_1 is turned off, then the DSP applies reset signal to the MOSFET M so that the detection circuit is reset once per period. Since the ZVS of S_1 has been realised first, C_{det} and C_{SP} will not be charged or discharged during the turn-on process of S_1, and there is almost no voltage superimposed on u_{sp} and remains unchanged during the turn-on process of S_1; so the sampled voltage u_{sp} is referred as u_{sp_off} at this time, which only reflect negative inductor current. For synchronous buck converter, the lower the switching frequency, the larger the negative inductor current, the more charges stored in the capacitor, and the higher the sampled voltage u_{sp_off}. For the purpose of minimising the switching loss, the large negative inductor current is not preferred, so the switching frequency should not be too low. While ensuring the ZVS state, the switching frequency should be increased appropriately in order to obtain a small negative inductor current. There is a voltage $u_{sp\_ref\_off}$, and when the sampled voltage u_{sp_off} is lower than this value, the negative inductor current is small, and the turn-off loss is considered to be optimised.

ZVS Tuning and switching loss dynamic optimisation: The algorithm block diagram of the proposed method is shown in Figure 3(a), and since the realisation of ZVS is the primary goal, the tuning of ZVS should be carried out first. If the sampled voltage u_{sp_on} is greater than $u_{sp\_ref\_on}$, the main switch S_1 is considered to be in the hard-switching state, and then the DSP reduces the switching frequency properly until ZVS is realised again. If the sampled voltage u_{sp_on} is not greater than $u_{sp\_ref\_on}$, it will trigger the DSP to optimise the turn-off loss. In the next period, the DSP first samples the voltage u_{sp_off}. If the voltage u_{sp_off} is greater than $u_{sp\_ref\_off}$, DSP increases the switching frequency appropriately to reduce the negative inductor current and minimise the turn-off loss.

A specific ZVS tuning and switching loss dynamic optimisation process of the proposed method is shown in Figure 3(b). In the first period $0 \sim T_1$, the first reset signal is set to detect ZVS state. The reset signal V_{reset} is applied before the main switch S_1 is driven, then it can be seen that $u_{sp\_on} > u_{sp\_ref\_on}$ during $0 \sim T_1$, which indicates that the ZVS of main switch S_1 is lost. Then, the DSP reduces the switching frequency appropriately to regain ZVS. In the second period $T_1 \sim T_2$, ZVS is realised by the switching frequency adjustment, which will trigger the DSP to sample the voltage u_{sp_off} during $T_2 \sim T_3$. In the third period $T_2 \sim T_3$, the sampled voltage u_{sp_off} is higher than $u_{sp\_ref\_off}$, and it can be judged that the turn-off loss is not optimal and will trigger the DSP to increase the switching frequency appropriately. Whether the switching frequency adjustment results in hard switching, it will be detected during $T_3 \sim T_4$. $u_{sp\_on} <= u_{sp\_ref\_on}$ shows that ZVS is still maintained. In the fifth period $T_4 \sim T_5$, it can be detected that $u_{sp\_off} <= u_{sp\_ref\_off}$, and it is known from the above analysis that when u_{sp_off} satisfies this condition, the negative inductor current is small and the turn-off loss is optimised. Because the realisation of ZVS is the primary goal, ZVS should be detected once after the switching frequency adjustment.

Experimental verification: An experimental prototype of the synchronous buck converter is built to verify the effectiveness of the proposed circuit and method. The input voltage of the buck converter is 100 V, the output voltage is 50 V, and the maximum output power is 70 W. Two MOSFETs BSC12DN20NS3G are adopted as main switches S_1 and S_2 of synchronous buck, they are driven complementarily by DSP TMS320F28027, and their switching frequency fs can be adjusted to maintain ZVS state and minimise switching loss. The adjusting range of switching frequency fs is 150~260 kHz. Inductor L_1 is 50 uH, capacitor C_1 is 44 uF/50 V. Considering the maximum permissible voltage of DSP and sampling accuracy, the capacitance C_{sp} of ZVS detection circuit is set as 47 pF, C_m is 10 pF, the resistance R_m is 120 Ω, MOSFET DMN63D8LW is utilised as switch M, and $u_{sp\_ref\_on} = 10$ mV, $u_{sp\_ref\_off} = 1.2$ V in the DSP program.

Figure 4 shows the experimental waveforms of S_1 drain-to-source voltage u_{ds1}, driving signal u_{gs1} and inductor current i_L at 45 and 65 W output power, respectively. As mentioned above, the DSP adjusts the switching frequency according to the feedback from the detection circuit so that each output power case corresponds to an optimal switching frequency, which guarantees ZVS and minimises switching loss.

In Figure 4, 45 W output power corresponding to optimal switching frequency f_s is 215 kHz, 65 W output power corresponding to optimal switching frequency f_s is 160 kHz. ZVS operation is always guaranteed and the negative peak value of inductor current i_L is 210 mA in both two cases.

Figure 5 shows the negative peak value of inductor current i_L under two comparison cases. In the first case, the proposed circuit and method are utilised. As can be seen from Figure 5, when the proposed circuit and method are adopted, the negative peak value of inductor current i_L under the different load conditions fluctuates very little and stays within −0.3 to −0.2 A. While in the second case, the switching frequency is fixed at 200 kHz when the proposed circuit and method are not employed, and the negative peak value of inductor current i_L changes greatly over the entire load range. Since the negative peak value of inductor current is proportional to the circuit circulating power, the proposed method has an outstanding advantage in reducing switching loss.

The efficiency curves for these two comparison cases are shown in Figure 6. In the first case, the proposed circuit and method are applied to automatically adjust the switching frequency to adapt different loads. In the second case, the switching frequency is fixed at 200 kHz. It can be seen that when the output power is 50 W, the efficiencies of these two cases are approximately equal because 200 kHz is the optimal switching

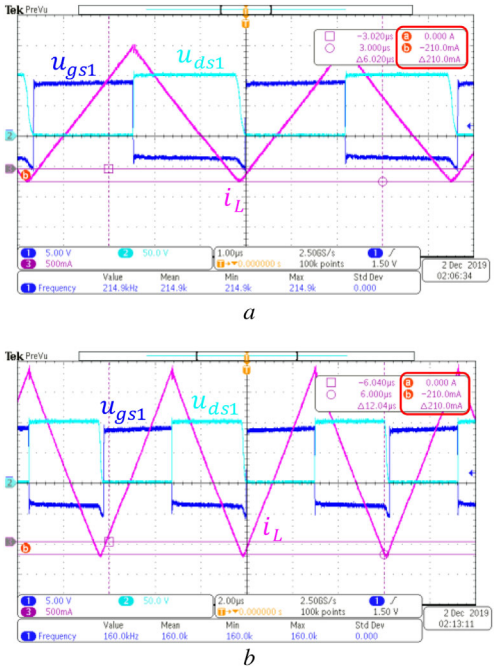


Fig. 4 *The experimental waveforms of S_1 drain-to-source voltage u_{ds1}, driving signal u_{gs1} and inductor current i_L at 45 and 65 W output power. (a) 45 W output power, (b) 65 W output power*

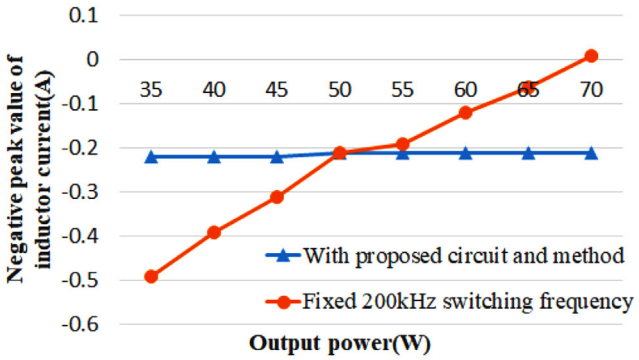


Fig. 5 *The negative peak value of inductor current i_L in two comparison cases*

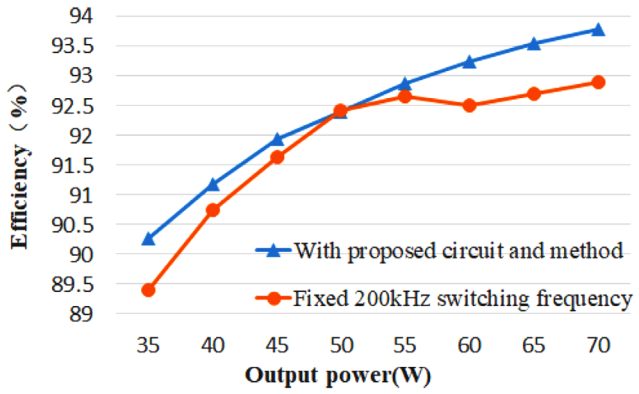


Fig. 6 *The efficiency of two comparison cases*

frequency for 50 W output power. When the output power is less than 50 W, ZVS operation of the buck converter with the fixed switching frequency is always guaranteed, but the negative inductor current is larger and the circuit circulating power is higher, resulting in lower efficiency. The converter employing the proposed method can dynamically adjust the switching frequency according to the feedback results from the detection circuit so as to adapt to the current load condition, which not only guarantees ZVS but also minimises the turn-off loss. Within the range of output power less than 50 W, the efficiency improvement indicates that the proposed method has remarkable effect in optimising turn-off loss, and it also can be seen that the lower the output power, the greater the efficiency improvement. When the output power is greater than 55 W, the efficiency of the fixed 200 kHz switching frequency method falls because the switching frequency gradually deviates from the optimal point with the increase of output power, and its ZVS may be lost. While ZVS operation of buck converter using the proposed method is always been maintained, at peak output power, its efficiency is improved by nearly 1%.

Conclusion: In this letter, a simplified ZVS detection circuit and switching loss dynamic optimisation method for synchronous buck converter has been proposed, which ensure ZVS state and minimise switching loss in a wide load range. A synchronous buck converter experimental prototype with rated input voltage 100 V and output power 70 W is established to verify the effectiveness of the proposed circuit and method. The experimental results show that the proposed method has a great effect in maintaining ZVS and optimising the switching loss. Finally, the efficiency has been promoted by nearly 1%, the detection circuit components and the number of ADC channels of the DSP are reduced by half, which can further improve the power density of the converter.

Acknowledgments: This work was supported by the National Natural Science Foundation of China under Grant 51707088 and Grant 51607081.

Received: *23 September 2020* Accepted: *16 November 2020*
doi: 10.1049/ell2.12065

References

1 Bing, L., et al.: Optimal design methodology for LLC resonant converter. In: 2006 IEEE Applied Power Electronics Conference and Exposition(APEC), Dallas, TX, USA, pp. 533–538 (2006)
2 Zhang, Y., et al.: Analysis and research of a soft-switching bidirectional DC–DC converter without auxiliary switches. *IEEE Trans. Ind. Electron.* **65**(2), 1196–1204 (2018)
3 Oh, I.: A soft-switching synchronous buck converter for Zero Voltage Switching (ZVS) in light and full load conditions. In: 2008 IEEE Applied Power Electronics Conference and Exposition(APEC), Austin, TX, USA, pp. 1460–1464 (2008)
4 Lee, S.-S.: Step-down converter with efficient ZVS operation with load variation. *IEEE Trans. Ind. Electron.* **61**(1), 591–597 (2014)
5 Do, H.: Zero-voltage-switching synchronous buck converter with a coupled inductor. *IEEE Trans. Ind. Electron.* **58**(8), 3440–3447 (2011)
6 Chuang, Y.-C., Ke, Y.-L.: A novel high-efficiency battery charger with a buck zero-voltage-switching resonant converter. *IEEE Trans. Energy Convers.* **22**(4), 848–854 (2007)
7 Huang, X., et al.: High-frequency high-efficiency GaN-based interleaved CRM bidirectional buck/boost converter with inverse coupled inductor. *IEEE Trans. Power Electron.* **31**(6), 4343–4352 (2016)
8 Lee, J.-J., Shin, J.-W., Cho, B. H.: A digital predictive critical conduction mode buck converter control method. In: 2012 IEEE Applied Power Electronics Conference and Exposition (APEC), Orlando, FL, USA, pp. 709–714 (2012)
9 Yeh, C., Zhao, X., Lai, J.: An investigation on zero-voltage-switching condition in synchronous-conduction-mode buck converter. In: 2017 IEEE Energy Conversion Congress and Exposition (ECCE), Cincinnati, OH, USA, pp. 1728–1732 (2017)
10 Li, S., Shu, W., Lu, S.: Voltage slope-based zero voltage switching detection method for wireless power transfer systems. *Electron. Lett.* **54**(12), 775–777 (2018)
11 Huang, X., et al.: Characterization and enhancement of high-voltage cascode GaN devices. *IEEE Trans. Electron Dev.* **62**(2), 270–277 (2015)
12 Huang, X., et al., Evaluation and application of 600 V GaN HEMT in cascode structure. *IEEE Trans. Power Electron.* **29**(5), 2453–2461 (2014)

Investigation of a hybrid converter with 16:1 wide voltage operation

Bor-Ren Lin ✉ and Mao-Cheng Ko
Department of Electrical Engineering, National Yunlin University of Science and Technology, Taiwan
✉ E-mail: linbr@yuntech.edu.tw

This study investigates a hybrid zero-voltage switching converter for dc–dc converters on railway vehicle applications or solar PV power conversion applications. The dc–dc converters used in railway applications demand to have at least 30% to 40% input voltage variation in order to meet the international standard EN50155. However, the main nominal dc input voltages of power converters on railway applications are from 24 to 110 V. To overcome the wide voltage operation problem, a hybrid power converter with the combination of boost and resonant circuits is proposed and investigated to accomplish a wide input voltage operation (10–160 V) on railway vehicle applications. Under low-voltage input case (V_{in} = 10–40 V), both boost and resonant circuits are activated to have high voltage gain. For medium-voltage input case (V_{in} = 40–80 V), only full-bridge resonant circuit is operated to achieve soft switching operation on active devices. Under high-voltage input case (V_{in} = 80–160 V), the half-bridge type resonant circuit is activated to achieve less voltage gain. The circuit performance and effectiveness are provided and evaluated by experiments.

Introduction: Switching power converters with high efficiency have been demanded in commercial and industrial power supplies. The wide input voltage dc–dc converters are required on railway vehicles for the braking system, motor drive control, lighting system and electric door system. The nominal input voltages of the conventional individual dc–dc converter on railway vehicles are 24, 37.5, 48, 72, 96 and 110 V with ± 30% (or ± 40%) variation. Therefore, the universal dc converter with wide input voltage capability is welcomed to be used at all nominal input voltages from 24 to 110 V on the railway vehicle system. The dc converters for remote control system on standalone solar photovoltaic cell (PV) system demand the supply of the independent dc power to control system board. According to the solar intensity (day or night) and geographical location, the solar cell panel has wide output voltage. Thus, the wide voltage operation converters are requested against the solar intensity variation to supply the stable dc voltage to the power unit on the stand-alone PV system. To implement wide voltage dc converters, the cascade dc converters with buck, boost or buck-boost circuit in front-stage and half-bridge or full-bridge circuit in rear-stage were reported in [1]. However, two-stage dc converters will reduce the total circuit efficiency due to the front-stage circuit being always operated. The series or parallel connection circuits have been presented in [2] to overcome wide voltage problem on fuel cell applications. However, the components and cost are increased in these circuit topologies. Pulse-width modulation (PWM) dc–dc converters with phase shift technique were reported in [3] to accomplish wide voltage range operation. Wide voltage range resonant converters were studied and implemented in [4]. However, the voltage variation in [4] is less than 4:1 (V_{max}/V_{min} < 4). The input voltage range of dc converters on railway vehicles or solar power conversion applications is sometimes required to be more than 4:1 to overcome different solar intensity problem during day and night.

A hybrid resonant converter is reported and accomplished to achieve 16:1 ($V_{in,min}$ = 10 V to $V_{in,max}$ = 160 V) wide input voltage operation and zero-voltage switching characteristic. There are three voltage ranges (low voltage: 10–40 V, medium voltage: 40–80 V, and high voltage: 80–160 V) in the presented converter. Two-stage converter (boost circuit + resonant circuit) is adopted in the developed circuit. However, the boost circuit is only activated on the low-voltage input condition. Under the low-voltage input case (V_{in} = $V_{in,min}$~4 $V_{in,min}$), the boost circuit is controlled to realise the voltage boost and the dc-bus voltage is controlled at the constant-voltage value. For medium (V_{in} = 4 $V_{in,min}$~8 $V_{in,min}$) and high (V_{in} = 8 $V_{in,min}$~16 $V_{in,min}$) voltage ranges, boost circuit is off to lessen power loss on boost switch such that the dc-bus voltage is equal to the input voltage. The full-bridge resonant circuit in the rear-stage is operated on low and medium voltage ranges. However, the half-bridge type resonant circuit is operated under the high-voltage input condition to reduce the voltage gain. By using the proposed control approach, the presented hybrid converter can achieve wide voltage operation. Since the series resonant circuit is operated in the rear-stage, the soft-switching characteristic on active devices is realised and the switching loss is reduced. Finally, the effectiveness of the prototype is verified by the experimental tests.

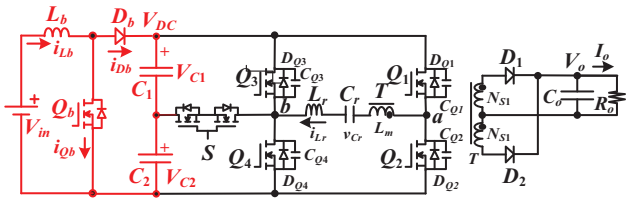


Fig. 1 Hybrid converter with 16:1 wide voltage operation

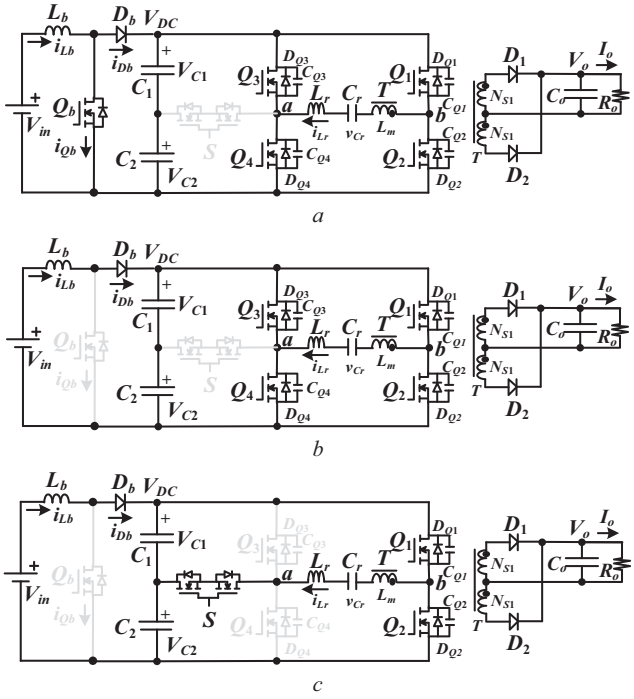


Fig. 2 Equivalent circuits of the presented converter under different voltage ranges. (a) $V_{in,min} < V_{in} < 4\ V_{in,min}$; (b) $4\ V_{in,min} < V_{in} < 8V_{in,min}$; (c) $8\ V_{in,min} < V_{in} < 16\ V_{in,min}$

Structure of proposed converter: The converter configuration is shown in Figure 1. Basically, the converter is a two-stage dc–dc circuit including a boost converter (L_b, Q_b, D_b, C_1 and C_2.) in the front-stage and a hybrid resonant circuit (L_r, T, C_r, Q_1–Q_4, S, D_1, D_2 and C_o) in the rear-stage. According to the input voltage value, the proposed converter can have three operation modes. When the input voltage V_{in} is greater than the minimum voltage $V_{in,min}$ and less than 4 $V_{in,min}$, the ac switch S is turned off as shown in Figure 2a. Boost and resonant circuits are both activated to regulate V_o at the desired value. Active switch Q_b is activated by the PWM to regulate V_{DC} at the desired value. Then, the resonant circuit with constant input voltage V_{DC} is operated and limited at a narrow frequency range (series resonant frequency). If the input voltage 4 $V_{in,min}$ < V_{in} < 8 $V_{in,min}$, then the converter is worked under medium-voltage input condition (Figure 2b). Boost switch Q_b and ac switch S are off and dc-bus voltage $V_{DC} \approx V_{in}$. Only the full-bridge resonant circuit is activated to control load voltage. If 8 $V_{in,min}$ < V_{in} < 16 $V_{in,min}$, then active switch Q_3, Q_4 and Q_b are all turned off and S is turned on (Figure 2c). The half-bridge type dc–dc converter is operated to reduce the dc voltage gain. Based on the proposed control strategy, the 16:1 wide voltage operation capability can be achieved in the presented hybrid converter.

Circuit characteristics and design procedure: The proposed hybrid converter includes two circuits. The boost circuit is only operated in low-voltage input condition ($V_{in,\min} \sim 4\ V_{in,\min}$) and the dc-bus voltage V_{DC} is controlled at $V_{DC,ref}$. If boost switch Q_b is in the on-state, then $V_{Lb} = V_{in}$. Similar, $V_{Lb} = V_{in}$-V_{DC} if S_b is in the off-state. Under continuous conduction mode, the relation between V_{in} and dc-bus voltage V_{DC} is derived in Equation (1):

$$V_{DC} = V_{in}/(1 - d_b) \tag{1}$$

where d_b is the duty cycle of boost switch Q_b. Since $V_{in} = V_{in,\min} \sim 4\ V_{in,\min}$ in low-input voltage range, the corresponding minimum and maximum duty cycles are derived as

$$d_{b,\min} = (V_{DC,ref} - 4V_{in,\min})/V_{DC,ref} \tag{2}$$

$$d_{b,\max} = (V_{DC,ref} - V_{in,\min})/V_{DC,ref} \tag{3}$$

Based on the dc-bus current I_{DC} and duty cycle d_b, the $i_{Qb,rms}$ (root mean square current) is calculated by

$$I_{Qb,rms} = I_{DC}\sqrt{d_b}/(1 - d_b) \tag{4}$$

The average diode current $I_{Db} = I_{DC}$ and the *rms* diode current $I_{Db,rms} = I_{DC}/\sqrt{1 - d_b}$. The *rms* inductor current is given as $I_{Lb,rms} = I_{DC}/(1 - d_b)$. For medium- and high-voltage input cases, boost switch S_b becomes off and the dc-bus voltage $V_{DC} = V_{in}$. Frequency modulation approach is used to obtain the approximately voltage gain of the full- or half-bridge resonant converter. Full-bridge type resonant circuit is operated in low and medium voltage ranges (Figures 2a and b), and half-bridge type resonant circuit is worked in high-voltage input condition (Figure 2c). According to the conducting states of D_1 and D_2, the fundamental magnetising voltage is approximately calculated in Equation (5).

$$V_{Lm,rms} = 2\sqrt{2}nV_o/\pi \tag{5}$$

The fundamental ac resistance R_{ac} on the input-side is related to the dc load resistor R_o and can be calculated as

$$R_{ac} = \frac{v_{Lm,rms}}{i_{Lm}} = \frac{8n^2R_o}{\pi^2} \tag{6}$$

According to the conducting states of Q_1–Q_4, the square voltage waveform with $\pm\ V_{DC}$ or $\pm\ V_{DC}/2$ is generated on the ac leg voltage v_{ab}. The *rms* value of v_{ab} at fundamental switching frequency approximates

$$V_{ab,rms} = \begin{cases} \frac{2\sqrt{2}V_{DC}}{\pi} = \frac{2\sqrt{2}V_{in,L}}{\pi(1-d_b)}, \text{lowvoltagerange} \\ \frac{2\sqrt{2}V_{DC}}{\pi} = \frac{2\sqrt{2}V_{in,M}}{\pi}, \text{mediumvoltagerange} \\ \frac{\sqrt{2}V_{DC}}{\pi} = \frac{\sqrt{2}V_{in,H}}{\pi}, \text{highvoltagerange} \end{cases} \tag{7}$$

Based on the given resonant components (C_r, L_r, R_{ac} and L_m) and input and output voltages $V_{ab,rms}$ and $V_{Lm,rms}$, the transfer function of the full- or half-bridge resonant circuit is determined by

$$\begin{aligned} |G| = \frac{V_{Lm,rms}}{V_{ab,rms}} &= \left|\frac{\frac{R_{ac} \times j\omega_{sw}L_m}{R_{ac} + j\omega_{sw}L_m}}{j\omega_{sw}L_r + \frac{1}{j\omega_{sw}C_r} + \frac{R_{ac} \times j\omega_{sw}L_m}{R_{ac} + j\omega_{sw}L_m}}\right| \\ &= \frac{1}{\sqrt{[1 + \frac{1}{L_B}\frac{F^2 - 1}{F^2}]^2 + Q^2(\frac{F^2 - 1}{F})^2}} \\ &= \begin{cases} \frac{nV_o(1 - d_b)}{V_{in,L}} = \frac{nV_o}{V_{DC,ref}}, \text{lowvoltagerange} \\ \frac{nV_o}{V_{in,M}}, \text{mediumvoltagerange} \\ \frac{2nV_o}{V_{in,H}}, \text{highvoltagerange} \end{cases} \end{aligned} \tag{8}$$

where $V_{in,L}$, $V_{in,H}$ and $V_{in,H}$ are input voltage at low, medium and high voltage range operations, $L_B = L_m/L_r$, $Q = \sqrt{L_r/C_r}/R_{ac}$ and $F = f_{sw}/f_r$. For low voltage range, VDC voltage is controlled at the reference voltage $V_{DC,ref}$ by boost converter. Thus, the resonant circuit is operated at constant dc bus voltage. The switching frequency of the resonant circuit is only related to load current. Since $V_{in,H} = 2V_{in,L}$ in the proposed control strategy, the voltage gains G_M and G_H in Equation (8) are identical.

The proposed prototype is investigated and experiments are provided to show the effectiveness of the studied hybrid converter. The electric specifications of the laboratory prototype are $V_{in} = 160$–10 V (16:1 ratio), $V_o = 12$V and $P_{o,rated} = 200$ W. When $V_{in} = 10$ –40 V (low-input voltage range), boost and full-bridge resonant circuits are both operated to control $V_{DC} = 42$ V and $V_o = 12$ V. When $V_{in} = 40$ –80 V (medium-voltage input case), only full-bridge type resonant converter is worked to control $V_o = 12$ V. In the same manner, if $V_{in} = 80$–160 V (high-voltage input case), half-bridge type resonant converter is activated to let $V_o = 12$ V. In continuous conduction mode of the boost converter, the minimum and maximum duty cycles of Q_b are approximated as

$$d_{b,\min} = \frac{V_{DC,ref} - V_{in,L,\max}}{V_{DC,ref}} = \frac{42 - 40}{42} \approx 0.05 \tag{9}$$

$$d_{b,\max} = \frac{V_{DC,ref} - V_{in,L,\min}}{V_{DC,ref}} = \frac{42 - 10}{42} \approx 0.76 \tag{10}$$

The duty cycle control integrated circuit with 100 kHz switching frequency is adopted to control boost switch Q_b. The ripple current Δi_{Lb} is assumed 20% of the maximum inductor current $I_{LB,\max}$. Thus, L_b can be obtained in Equation (11).

$$L_b = \frac{V_{in,L,\min}d_{b,\max}}{\Delta i_{Lb}f_{sw}} = \frac{10 \times 0.76}{0.2 \times (200/10) \times 10^5} = 19\ \mu\text{H} \tag{11}$$

The *rms* switch current $I_{Qb,rms}$ and the average diode current I_{Db} are expressed in Equations (12) and (13).

$$I_{Qb,rms} = I_{in,\max}\sqrt{d_{b,\max}} \approx (200/10)\sqrt{0.76} \approx 17.44A \tag{12}$$

$$I_{Db} = I_{DC,\max} = 200/42 \approx 4.76A \tag{13}$$

The voltage ratings of Q_b and D_b are equal to the input maximum voltage $V_{in,\max} = 160$ V. Thus, boost switch Q_b is implemented by power Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET) IRFB260N with 200 V/56 A rating and D_b is implemented by fast recovery diode MBR40200PT with 200 V/40A ratings.

For resonant converter, the series resonant frequency by C_r and L_r is designed at 100 kHz. The inductor ratio L_B is designed at 5 to achieve enough high voltage gain in medium and high voltage ranges. Since $V_{in,H} = 2V_{in,L}$ in the proposed control strategy, the voltage gains G_M and G_H in Equation (8) are identical. For medium-input voltage range, device S is off, the full-bridge resonant circuit is operated to control load voltage V_o. In medium voltage range, the lowest and highest input voltages are 40 and 80 V, respectively. The minimum gain $G_{M,\min}$ is designed at unity under 80 V input. Thus, the transformer turn-ratio n is calculated in Equation (14).

$$n = \frac{G_{M,\min}V_{in,M,\max}}{V_o} = \frac{1 \times 80}{12} \approx 6.67 \tag{14}$$

Ferrite core TDK EER42 with $n_p = 13$ turns and $n_s = 2$ turns and actual turn-ratio $n = n_p/n_s = 13/2 = 6.5$. Based on the select turn-ratio, the maximum and minimum voltage gains of the resonant circuit are expressed as

$$G_{M,\max} = \frac{nV_o}{V_{in,M,\min}} = \frac{6.5 \times 12}{40} = 1.95 \tag{15}$$

$$G_{M,\min} = \frac{nV_o}{V_{in,M,\max}} = \frac{6.5 \times 12}{80} = 0.975 \tag{16}$$

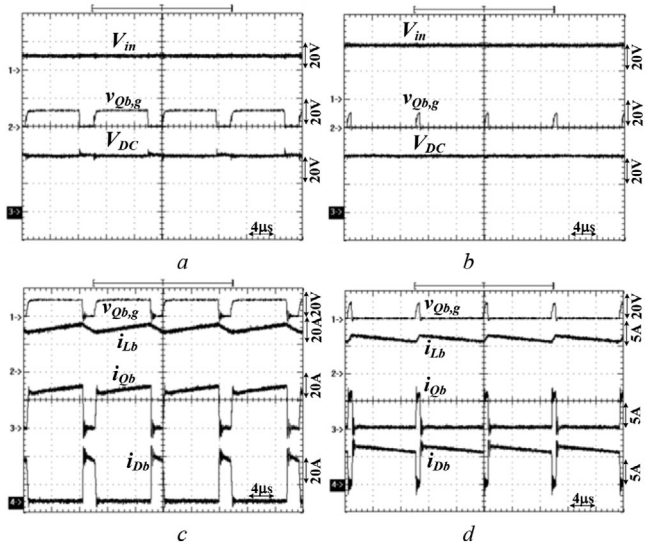


Fig. 3 Experimental results of the boost converter when 10 V < V_{in} < 40 V (low voltage range) and full load (a) V_{in}, $v_{Qb,g}$ and V_{DC} at 10 V input; (b) V_{in}, $v_{Qb,g}$ and V_{DC} at 40 V input; (c) $v_{Qb,g}$, i_{Lb}, i_{Qb} and i_{Db} at 10 V input; (d) $v_{Qb,g}$, i_{Lb}, i_{Qb} and i_{Db} at 40 V input

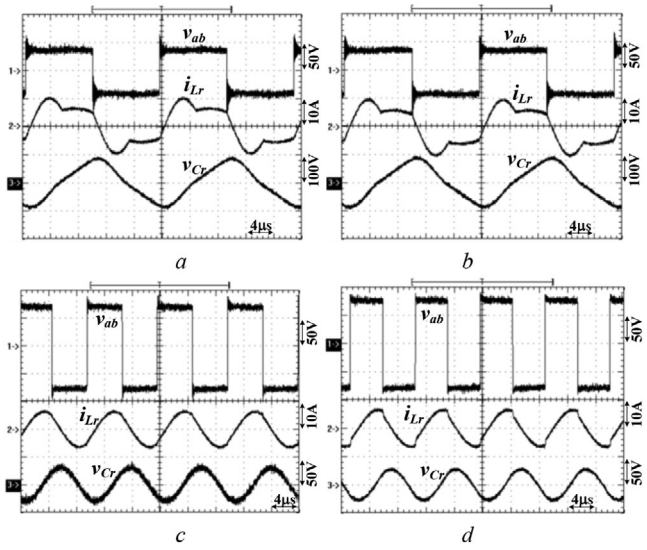


Fig. 4 Measured primary-side waveforms at full load. (a) V_{in} = 10 V, (b) V_{in} = 40 V, (c) V_{in} = 75 V, (d) V_{in} = 160 V

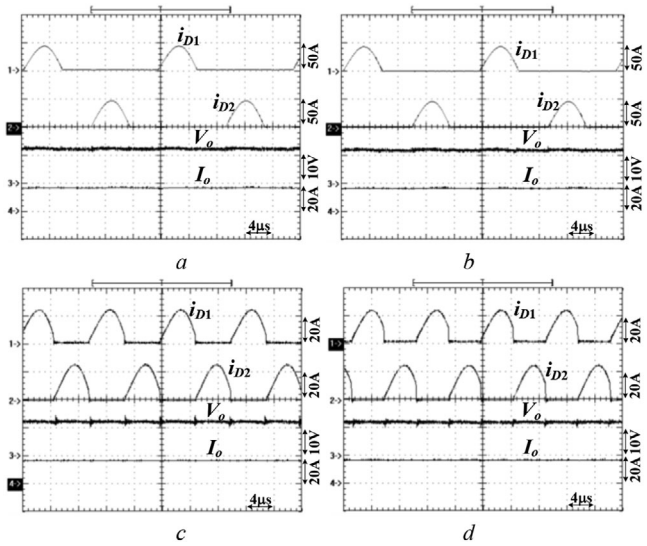


Fig. 5 Measured secondary-side waveforms at full load. (a) V_{in} = 10 V, (b) V_{in} = 40 V, (c) V_{in} = 75 V, (d) V_{in} = 160 V

The fundamental ac resistance at full load can be calculated from Equation (6).

$$R_{ac,full} = \frac{8n^2 R_{o,full}}{\pi^2} = \frac{8 \times 6.5^2 \times (12^2/200)}{\pi^2} \approx 24.66\Omega \quad (17)$$

In order to achieve maximum gain $G_{M,\max}$, the quality Q at full load is designed at 0.2. Thus, L_r can be obtained from the given f_r, Q and $R_{ac,full}$.

$$L_r = \frac{QR_{ac,full}}{2\pi f_r} = \frac{0.2 \times 24.66}{2\pi \times 100000} \approx 7.85\,\mu\text{H} \quad (18)$$

Since $L_B = 5$, the magnetising inductance can be obtained as$L_m = L_r L_B = 7.85 \times 5 = 39.25\ \mu$H. The actual inductances in the prototype are $L_r = 8\ \mu$H and $L_m = 40\ \mu$H. The resonant capacitance C_r is obtained as

$$C_r = \frac{1}{4\pi^2 L_r f_r^2} = \frac{1}{4\pi^2 \times 8 \times 10^{-6} \times (100,000)^2} \approx 316\ \text{nF} \quad (19)$$

The voltage stresses of the power semiconductors are expressed in Equations (20)–(22).

$$V_{Q1,rating} = .. = V_{Q4,stress} = V_{in,\max} = 160\ \text{V} \quad (20)$$

$$V_{S,rating} = V_{in,\max}/2 = 80\ \text{V} \quad (21)$$

$$V_{D1,rating} = V_{D2,rating} = 2V_o = 24\ \text{V} \quad (22)$$

Power MOSFETs n-channel IRFB4229 with 250 V/46 A ratings are selected for switches Q_1–Q_4 and S. MBR40100PT fast recovery diodes with 100 V/40 A are selected for D_1 and D_2. The select capacitances are $C_1 = C_2 = 680\ \mu$F and $C_o = 3000\ \mu$F. The integrated circuit UC3843 is adopted to control boost converter, and UCC25600 is used to control the resonant converter.

Experimental results: Figure 3 provides the test results of the boost converter under 10 V < V_{in} < 40 fV (low voltage range). Figures 3a and b give the input voltage V_{in}, the boost switch signal $v_{Qb,g}$ and dc-bus voltage V_{DC} for 10 and 40 V inputs. The V_{DC} is controlled at 42 V. The experimental waveforms of boost switch signal $v_{Qb,g}$, inductor current i_{Lb}, switch current i_{Qb} and diode current i_{Db} at 10 and 40 V inputs are provided in Figures 3c and d. From Figure 3, the Q_b at 10 V input has more duty cycle than at 40 V input and the boost converter is operated in continuous conduction mode. Figure 4 gives the test results of the resonant circuit at 10, 40, 75 and 160 V input cases. For 10 and 42 V inputs and the full load condition, the V_{DC} is controlled at 42 V and the switching frequency f_{sw} is less than the resonant frequency f_r. Thus, the quasi-sinusoidal voltage and current waveforms are observed in Figures 4a and b. For 75 and 160 V inputs, the voltage gain of the resonant circuit is close to unity so that $f_{sw} \approx f_r$ and the resonant current and voltage are sinusoidal waveforms (Figures 4c and d). Figure 5 provides the experimental results of the secondary-side waveforms under 100% load. It can be observed that the switching frequency for 10 and 40 V inputs is about 50 kHz (< f_r = 100 kHz). Thus, both D_1 and D_2 are turned off without reverse recovery loss. The switching frequencies at 75 and 160 V inputs are close to 100 kHz (series resonant frequency) and the magnetising current loss is reduced.

Conclusion: A new hybrid resonant converter with wide input voltage operation (10–160 V) is presented and experimented for railway vehicle dc power supplies or solar power conversion. To accomplish wide voltage input capability, boost circuit is operated when input voltage is under low voltage range and resonant circuit is operated on the second stage to realise soft-switching operation and achieve the electrical isolation. An alternative current power switch is adopted on the primary side to select the full- or half-bridge converter according to the input

voltage range. The main contributions in this study are (1) the proposed converter has three different equivalent circuits to overcome wide voltage operation, (2) an example of a laboratory prototype is provided, and (3) soft-switching operation for resonant converter is achieved. The performance and effectiveness of the study hybrid converter are examined through the experimental results.

Received: *19 September 2020* Accepted: *28 October 2020*
doi: 10.1049/ell2.12055

References

1 Lin, B.R.: Resonant converter with wide input voltage range and input current ripple-free. *Electron. Lett.* **54**(18), 1086–1088 (2018)
2 Zhang, Y. et al.: A wide input-voltage range quasi-Z-source boost DC–DC converter with high-voltage gain for fuel cell vehicles. *IEEE Trans. Ind. Electron.* **65**(6), 5201–5212 (2018)
3 Lu, J., et al.: Step-down impedance control network resonant DC-DC converter utilizing an enhanced phase-shift control for wide-input-range operation. *IEEE Trans. Ind. Appl.* **54**(5), 4523–4536 (2018)
4 Kim, B., et al.: Hybrid resonant half-bridge DC/DC converter with wide input voltage range. In: Proceedings of IEEE APEC Conference, San Antonio, USA, pp. 1876–1881 (2018)

Comprehensive analysis of a three-phase DC–DC resonant converter with an open delta-wye transformer using variable frequency

Kristian Pessoa dos Santos,[1,✉] Paulo Peixoto Praça,[2] Hermínio Miguel Oliveira Filho,[3] Demercil de Souza Oliveira Jr.,[2] Luís Henrique Silva Colado Barreto,[2] and Gustavo Alves de Lima Henn[3]

[1]*Industry and Control Department, Federal Institute of Education, Science and Technology of Piauí, Parnaíba, Brazil*
[2]*Electrical Engineering Department, Federal University of Ceará, Fortaleza, Brazil (E-mail: paulopp@dee.ufc.br, demercil@dee.ufc.br, lbarreto@dee.ufc.br)*
[3]*Institute for Engineering and Sustainable Development, University for International Integration of the Afro-Brazilian Lusophony, Redenção, Brazil (E-mail: herminio@unilab.edu.br, gustavo@unilab.edu.br)*

✉Email: kristianpessoa@ifpi.edu.br

A three-phase soft-switched DC–DC resonant converter with variable frequency to perform the output voltage control is proposed. The open delta-wye connection provides a double output voltage reducing the primary current stress, turns ratio, weight and volume of high-frequency transformer. To achieve high efficiency, an LLC resonant tank is designed so that all switches can perform zero-voltage switching for the entire power range. In order to verify the operating principle of the converter, a design example is described, with the following converter specifications: input voltage 96 V, output voltage 380 V and a switching frequency up to 120 kHz. The feasibility and practicality of the proposed converter are confirmed by experimental results with a 1.5 kW prototype.

Introduction: Renewable energy sources, such as wind power, solar power and fuel cell, have been researched and developed to reduce environmental issues. Several topologies of DC–DC converters have been widely employed for integration of generation systems, storage or even consumers into a micro-grid or electric vehicles. In addition, control strategies and development of new topologies are one of the main objects of researches to provide converters with high efficiency and power density for the most diverse applications in the industry. In this way, LLC resonant converter has been widely used in these applications due to its capability to operate with zero-voltage switching (ZVS), or zero-current switching (ZCS), decreasing losses in semiconductor switches and electromagnetic interference [1]. Several soft-switching techniques can be applied to perform output voltage control, as phase-shift, Pulse-Width Modulation (PWM), frequency control or combinations of those [2–8]. However, only variable frequency control allows soft switching over the entire load range. In terms of topology, the single-phase and three-phase DC–DC isolated LLC resonant converters, as well as double-active bridge (DAB) topologies, have been widely used in the literature. However, few literatures focus on the inherent mechanisms of multi-phase LLC resonant converters. In this work, a three-phase soft-switched DC–DC resonant converter with an open delta-wye high-frequency transformer using variable frequency is proposed. The aforementioned topology can be used as interface between fuel cells or photovoltaic systems and the AC grid, as shown in Figure 1.

Proposed converter: The circuit structure of the proposed resonant converter is shown in Figure 2. Three single-phase full-bridge circuits are used in parallel in the primary side using S_1–S_{12} power MOSFETs. Components L_{S1}–L_{S3}, L_{p1}–L_{p3} and C_{S1}–C_{S3} work as an LLC resonant tank, and three high-frequency transformers T_1–T_3 provide galvanic isolation. In the secondary side, D_1–D_6 are rectifier diodes, C_o is output filter capacitor and R_o is the load resistance. The switches S_1–S_{12} are driven with variable frequency and present constant duty ratio $D = 50\%$. The use of more switches in the primary side, compared to conventional structures that uses a three-phase inverter, has the advantage of providing an increase in the power capability of the converter, since current stresses are distributed among all switches [3–5]. As disadvantage, the converter needs a more complex control circuit, due to the increased number of switches. Thus, the proposed topology ensures a better behaviour than a conventional three-phase inverter, as it decreases turn ratio due to a double output voltage, provided by the transformer open delta-wye connection. Also, the series capacitor are able to block DC current component, preventing the saturation of transformer, and eliminating the need of a transformer current control [2, 9]. In addition, the magnetising inductance from the high-frequency transformer can be used as the parallel inductor of resonant tank, as so the leakage inductance as part of the series inductance of the resonant tank. Finally, the proposed converter can be extended into n full-bridges in parallel on primary side, and n diode arms in parallel on secondary side, in order to increase the output power. The experimental results of the prototype are provided to verify all design and analyses of the converter.

DC voltage gain characteristic of the proposed converter: The analysis of the proposed converter is performed considering ideal components, fundamental harmonic analysis, which only considers the fundamental component, and DC-bus voltages at the input and output sides are considered constants. The DC voltage gain $M(F)$ versus normalised frequency $F = f_S/f_o$ for different Q and λ values is shown in

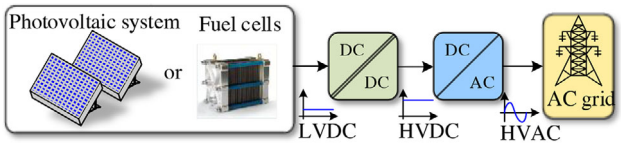


Fig. 1 *Basic circuit blocks of integration with AC grid*

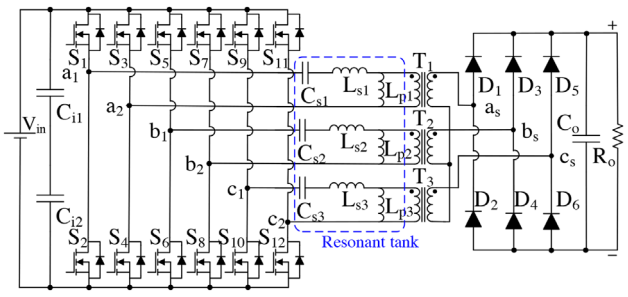


Fig. 2 *Proposed three-phase isolated DC–DC LLC resonant converter*

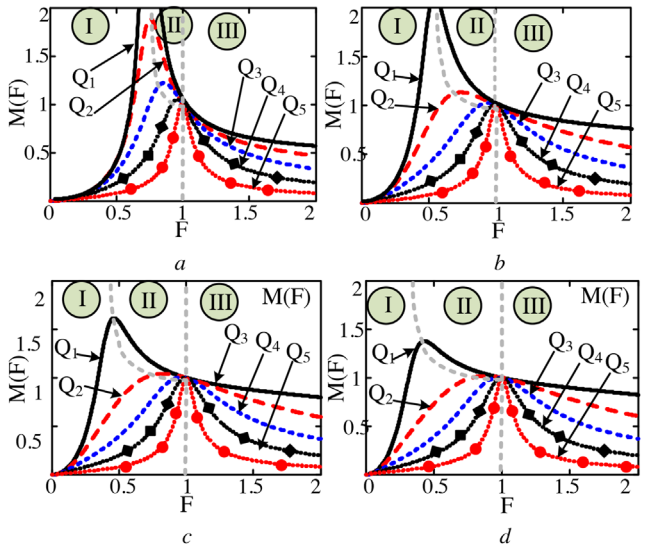


Fig. 3 *DC voltage gain for load variations: (a) λ = 1; (b) λ = 3; (c) λ = 5; (d) λ = 7*

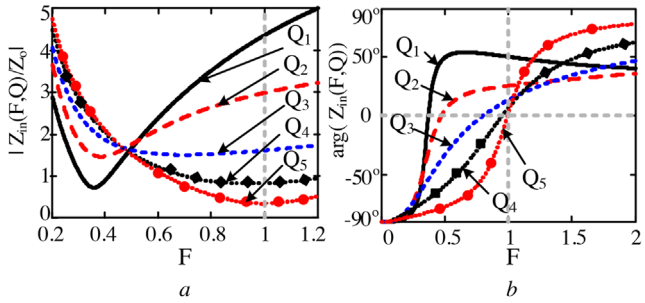


Fig. 4 *Input impedance versus normalised frequency, for λ = 5: (a) $|Z_{in}(F,Q)/Z_o|$; (b) arg($Z_{in}(F,Q)$)*

Figure 3 and represented by (1), where f_S is the switching frequency; $f_o = 1/2\pi\sqrt{(L_S/C_S)}$ is the resonant frequency; $Q = \omega_o L_S/Z'_{ac}$ is the quality factor; $Z'_{ac} = 6R_o/n_t^2\pi^2$ is the equivalent load impedance on the primary side reference; $n = 1/n_t$ is the transformation ratio; $\omega_o = 2\pi f_o$ is the switching angular resonant frequency; and $\lambda = L_P/L_S$ is the inductance ratio of parallel resonant inductor L_P and series resonant inductor L_S. Figure 3 shows waveforms for Q equals $Q_1 = 0.2$, $Q_2 = 0.5$, $Q_3 = 1$, $Q_4 = 2$ and $Q_5 = 5$. For region I, the switches turn off with ZCS and turn on with hard switching, while in regions II and III, the switches turn on with ZVS. However, only in region III this characteristic is independent of load conditions, thus it is a behaviour preferred for a design example:

$$M(F) = \left|\frac{V_o}{2n_t V_{in}}\right| = \left|\frac{\frac{6}{\pi^2}}{\left(\frac{6}{\pi^2}\right)\left(1+\frac{1}{\lambda}\left(1-\frac{1}{F^2}\right)\right)+jQ\left(F-\frac{1}{F}\right)}\right|. \quad (1)$$

Figure 4 shows the normalised input impedance Z_{in}/Z_o versus F for different Q values, where $Z_o = \omega_o L_S$ is base impedance and Z_{in} is represented by (2). For $F < 1$, the converter presents capacitive characteristic, while for $F > 1$, the inductive characteristic. This last characteristic is preferable, as the converter ZVS capability becomes more sensitive to frequency variations with Q changes [10]:

$$Z_{in}(F,Q) = j\left(\left(\frac{2\pi f_o F Q Z_{ac}{}'}{\omega_o}\right) - \frac{\omega_o Q Z_{ac}{}'}{2\pi f_o F}\right) + \frac{j\left(\frac{2\pi f_o F Q \lambda (Z_{ac}{}')^2}{\omega_o}\right)}{Z_{ac'} + j\left(\frac{2\pi f_o F Q \lambda (Z_{ac}{}')}{\omega_o}\right)}. \quad (2)$$

Design guidelines: To ensure ZVS characteristic, phase angle of Z_{in} must be positive and satisfy the condition $L_P \cdot I_{L_{P\_max}^2} > C_{oss} \cdot V_{GS\_MOSFET}{}^2 L_P \cdot I_{L_{P\_max}^2} > C_{oss} \cdot V_{GS\_MOSFET}{}^2$, where $I_{L_{P\_max}}$ is the maximum current of L_P and V_{GS_MOSFET} is the gate-source voltage. This condition ensures that the energy stored in resonant circuit is larger than energy stored in the output capacitor C_{oss} of each switch. Furthermore, the upper and the lower switches of each arm operate complementary with a dead-time t_{dt} that must be larger than the conditions in (3), where I_{DS_MOSFET} is the drain-source current on switch:

$$t_{dt} > \frac{2C_{oss} \cdot V_{in}}{I_{DS\_MOSFET}}. \quad (3)$$

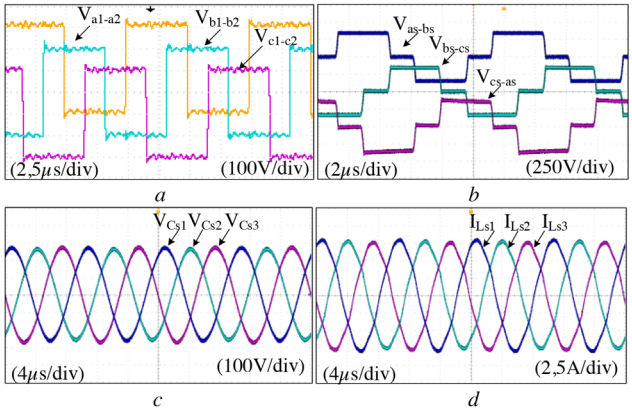


Fig. 5 *Experimental results of the proposed converter at full load input voltages of resonant tank, $V_{a1-a2} - V_{c1-c3}$; (b) three-phase secondary voltage transformer waveforms, $V_{as-bs} - V_{cs-as}$; (c) three-phase series capacitor voltage waveforms, $V_{CS1} - V_{CS3}$; (d) three-phase series inductor current waveforms, $I_{LS1} - I_{LS3}$*

Experimental results: In order to verify the theoretical analysis of the proposed converter, a rated 1.5 kW prototype is designed emulating up to eight solar panels in series, totalising a input voltage V_{in} of 96 V. The electrical specifications of the proposed converter are switching frequency f_S = 95 – 120 kHz; output voltage V_o = 380 V; output current I_o = 3.94 A; and input current I_{in} = 15.63 A. The main circuit components are $C_{i1} - C_{i2}$ = 600 μF; C_o = 470 μF; $L_{S1} - L_{S3}$ = 61.6 μH; $C_{S1} - C_{S3}$ = 47 ηF; $L_{p1} - L_{p3}$ = 360 μH; and n_t = 2.1. Silicon carbide MOSFET CCS020M12CM2 and hyperfast rectifier diodes 30EPH06PbF are used for S_1–S_{12} and D_1–D_6, respectively. A Thornton NEE-55 magnetic core is employed on inductors, and a Magmattec MMT 139T5020 toroidal core to transformers. The variable frequency control is implemented using a Texas Instruments TMS320F28377D digital signal processor board.

Figure 5 shows the main waveforms of the converter at full load and f_S = 96.1 kHz. The switches of each arm of primary side are phase-shifted by 180° and the three full-bridges are phase-shifted by 120° from each other. Figure 5a shows the input voltages of resonant tank, where it can be observed that the maximum value is equal to V_{in}. Figure 5b shows the secondary voltages of T_1–T_3, where the maximum value is equal to V_o.

From Figure 5c and d, it can been observed that three-phase series capacitors voltages, as well as three-phase series inductors currents of the LLC resonant circuit, are phase-shifted by 120° from each other. The peak values of $V_{C_{S1}} - -V_{C_{S3}}$ and $I_{L_{S1}} - -I_{L_{S3}}$ are 280 V and 7.5 A, respectively. The voltage and current waveforms of the resonant tank present a sinusoidal shape, which is an advantage if compared to DAB topologies. Moreover, the series capacitors of the resonant tank block the DC current component, preventing the saturation of transformer. Thus, it is not necessary a current loop control because the energy transfer of three phases are balanced.

Figure 6 presents the waveforms on one of the switches and one diode for different load conditions, considering V_{in} = 96 V. When drain-source voltage V_{DS_MOSFET} is 0 and its anti-parallel diode is conducting, the switch turns on, which guarantees the ZVS condition. At the secondary side, the rectifier diode current I_{diode} drops to 0 naturally when the voltage across the diode V_{AK_diode} is 0, so that rectifier diodes operate with ZCS at full load range.

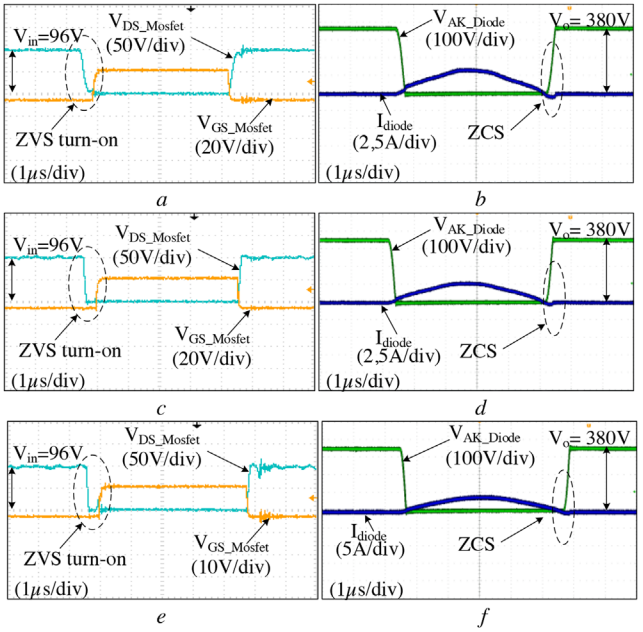


Fig. 6 *Experimental waveforms on switch and diode for different load and switching frequency conditions waveforms switch on, for 30% load and* $f_S = 105.2$ *kHz; (b) waveforms on diode, for 30% load and* $f_S = 105.2$ *kHz; (c) waveforms on switch, for 60% load and* $f_S = 100$ *kHz; (d) waveforms on diode, for 60% load and* $f_S = 100$ *kHz; (e) waveforms on switch, for 100% load and* $f_S = 96.1$ *kHz; (f) waveforms on diode, for 100% load and* $f_S = 96.1$ *kHz*

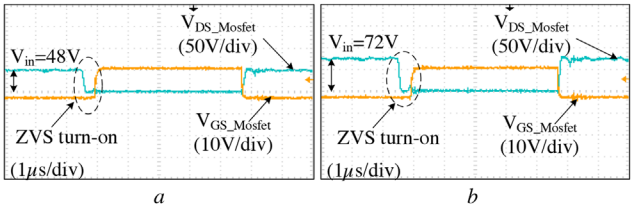


Fig. 7 *Experimental waveforms on a switch for different input voltage level at full load: (a)* $V_{in} = 48$ *V; (b)* $V_{in} = 72$ *V*

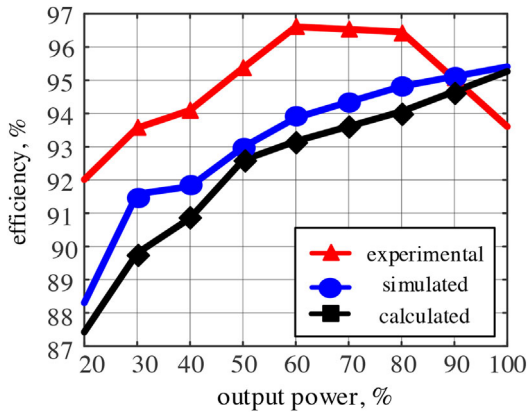


Fig. 8 *Efficiency curve*

Figure 7 shows the waveforms on one of the switches at full load condition, for $f_S = 96.1$ kHz and different input voltage levels. As it can be seen, over a wide input voltage range, from 48 to 96 V, the proposed resonant converter operates with ZVS, ensuring the high efficiency in all input voltage conditions. In these conditions, the measured efficiencies of the proposed converter are 92.2% for $V_{in} = 48$ V; 92.3% for $V_{in} = 60$ V; 93.31% for $V_{in} = 72$ V; 93.34% for $V_{in} = 84$ V; and 93.6% for $V_{in} = 96$ V.

Figure 8 shows calculated, simulated and measured efficiency curves of the prototype, estimating losses at $V_{in} = 96$ V. As it can be observed,

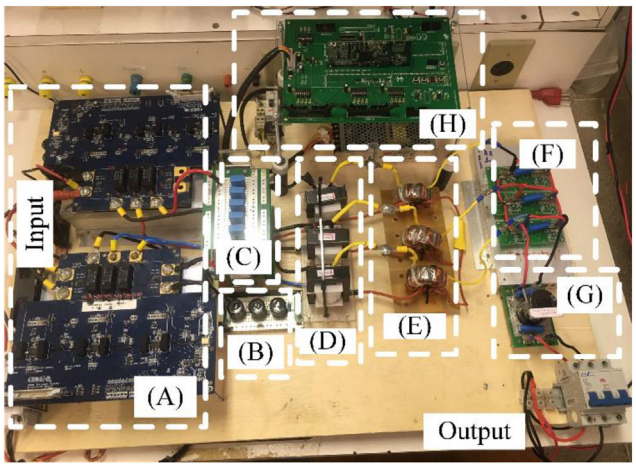


Fig. 9 *Laboratory prototype: (a) primary side circuit; (b) input capacitors; (c) series resonant capacitors; (d) series resonant inductors; (e) high-frequency transformers; (f) secondary side circuit; (g) output capacitor; (h) digital signal processor*

the converter achieves its maximum efficiency (96.6%) at 60% of the full load, whereas, at full load, the efficiency was around 94%. This behaviour can be explained due to the fact that all switches can realise ZVS, and all rectifier diodes realise ZCS at full load range. The laboratory prototype of the proposed converter has been built, as shown in Figure 9.

Conclusion: This work proposes a three-phase soft-switched DC–DC resonant converter with an open delta-wye transformer using variable switching frequency. Compared to conventional structures, it presents improved performance due to its capacity to realise ZVS for the entire load range and different input voltage levels. In addition, high frequency provides an improvement at the power density. The converter is suitable to be used as an the interface between low-level input voltages systems, as fuel cells or photovoltaic systems, and an inverter DC–AC to integrate with AC grid. Theoretical analysis and experimental results show that the proposed converter benefits from wide ZVS range and narrow variation of switching frequency. Experimental results of a 1.5 kW prototype confirm the performance of the proposed resonant converter.

Acknowledgement: The authors would like to thank Cearense Foundation to Support Scientific and Technological Development (FUNCAP) and National Council for Scientific and Technological Development (CNPQ) through the project no. 432744/2016-9 for financial support.

Received: *31 August 2020* Accepted: *5 October 2020*
doi: 10.1049/ell2.12043

References

1 Yaqoob, M., Loo, K.H., Lai, Y.M.: Fully soft-switched dual-active-bridge series-resonant converter with switched-impedance-based power control. *IEEE Trans. Power Electron.* **33**(11), 9267–9281 (2018)
2 Liu, F., Chen, Y., Chen, X.: Comprehensive analysis of three-level LC-type resonant DC/DC converter with variable frequency control – series resonant converter. *IEEE Trans. Power Electron.* **32**(7), 5122–5131 (2017)
3 Liu, C., Johnson, A., Lai, J.: A novel three-phase high-power soft-switched DC/DC converter for low-voltage fuel cell applications. *IEEE Trans. Ind. Appl.* **41**(6), 1691–1697 (2005)
4 Liu, F., Chen, Y., Chen, X.: Comprehensive analysis of three-phase three-level LC-type resonant DC/DC converter with variable frequency

control – series resonant converter. *IEEE Trans. Power Electron.* **32**(7), 5122–5131 (2017)
5 Oliveira Filho, H.M., Oliveira Junior, D.S., Praça, P.P.: Steady-state analysis of a ZVS bidirectional isolated three-phase DC–DC converter using dual phase-shift control with variable duty-cycle. *IEEE Trans. Power Electron.* **31**(3), 1663–1872 (2016)
6 Jauch, F., Biela, J.: Combined phase-shift and frequency modulation of a dual-active-bridge AC-DC converter with PFC. *IEEE Trans. Power Electron.* **31**(12), 8387–8397 (2016)
7 Showybul Islam Shakib, S.M., Mekhilef, S.: A frequency adaptive phase shift modulation control based LLC series resonant converter for wide input voltage applications. *IEEE Trans. Power Electron.* **32**(11), 8360–8370 (2017)
8 Dung, N.A. et al.: Novel modulation of isolated bidirectional DC-DC converter for energy storage systems. *IEEE Trans. Power Electron.* **34**(2), 1266–1275 (2019)
9 Khan, A.Z., Loo, K.H., Lai, Y.M.: Design, analysis and performance charactherization of dual-active-bridge DC-DC converter utilizing three-phase resonant immitance network. *IEEE Trans. Power Electronics.* **34**(2), 1159–1180 (2019)
10 Almardy, M.S., Bhat, A.K.S.: Three-phase (LC)(L)-type series-resonant converter with capacitive output filter. *IEEE Trans. Power Electron.* **26**(4), 1172–1183 (2011)

Hough transform based ego-velocity estimation in automotive radar system

Sohee Lim[1] and Seongwook Lee[2,✉]
[1]*Seoul National University, Seoul, Republic of Korea*
[2]*Korea Aerospace University, Gyeonggi-do, Republic of Korea*
✉Email: swl90@kau.ac.kr

In general, the velocity estimated through the automotive radar is the relative velocity between the ego-vehicle and the target. To estimate the absolute velocity of the ego-vehicle (i.e. ego-velocity), it is necessary to get help from other automotive sensors, such as cameras and lidars. In this letter, therefore, a method for estimating the ego-velocity is proposed using only radar sensor data. To estimate the ego-velocity, stationary targets in the detection result have to be identified because the velocity relative to the stationary target can be directly converted to the ego-velocity. In the proposed method, the radar data is first processed to extract the relative distance from the ego-vehicle to the target, relative velocity between the ego-vehicle and the target, and angle between the ego-vehicle and the target. Then, the detected targets are represented in the angle-velocity domain because moving and stationary targets in this domain exhibit different characteristics. In other words, the moving targets are randomly placed, but the stationary targets exist around a specific line. Finally, the Hough transformation is applied to automatically determine the line, and the ego-velocity is calculated from the transform result. The proposed method shows a maximum error of 2.4% in the velocity range of 30–90 km/h.

Introduction: To realize autonomous driving, data obtained from automotive sensors such as cameras, lidars, and radars must be used together. The main purpose of using these sensors is to effectively detect targets located around the ego-vehicle. For accurate target detection, it is important to identify the absolute movement of the ego-vehicle relative to the surrounding environments, which is called ego-motion estimation. For example, the authors in [1] estimated the ego-motion using deep neural networks in a single monocular camera system. In addition, in [2], laser range data is used to estimate the ego-motion of the vehicle in real time. To obtain complete information on ego-motion, the position and velocity of the ego-vehicle must be identified. In general, using radar sensor data, it is possible to estimate the relative distance to the target and the relative velocity of the target. When considering the position of the ego-vehicle as the origin, the relative distance can be directly converted into the absolute distance. However, to convert the relative velocity into the absolute velocity, we need to know the velocity of the ego-vehicle. Therefore, this letter proposes a method for estimating the velocity of the ego-vehicle using only radar sensor data without the help of other vehicle sensors.

The most important thing in estimating the velocity of the ego-vehicle is to identify stationary targets among the detected targets because they can be used as reference targets. In other words, the velocity relative to the stationary target can be directly converted to the velocity of the ego-vehicle. In this work, we use the detection result in the angle-velocity domain to discriminate the stationary targets. In this domain, the moving and the stationary targets exhibit different patterns. In other words, the detection points corresponding to stationary targets exist on a specific curve, but the points corresponding to moving targets are randomly scattered on a two-dimensional plane. Therefore, if the curve formulated by points corresponding to the stationary targets is estimated, we can calculate the absolute velocity of the ego-vehicle.

When several points are distributed in a plane, methods such as the linear least squares (LLS) [3] and random sampling consensus (RANSAC) [4], [5] are widely used to find the tendency of these points. In this work, we use the Hough transform [6] to extract only points corresponding to the stationary targets in the angle-velocity domain. This transformation can be effectively used to find inliers when points are scattered on the two-dimensional plane. Some studies have been conducted using Hough transforms on radar sensor data [7], [8]. For example, a vehicle moving at a constant velocity in the time-distance domain was extracted through the Hough transformation in [7]. In addition, patterns of moving and stationary targets were classified by applying the Hough transform to the time-frequency spectrum in the automotive radar system [8]. When this transformation is applied to detection points located in the angle-velocity domain, we can find the trend line corresponding to the stationary targets. Then, from the estimated trend line, the velocity of the ego-vehicle is calculated directly. To evaluate the estimation performance of the proposed method, radar sensor data obtained from actual road environments are used. In our experiment, a frequency-modulated continuous wave (FMCW) radar is mounted on the vehicle.

The remainder of this letter is organized as follows: First, we introduce the automotive radar system used in the experiment. Then, the target detection result in a actual road environment is described. Next, we explain the Hough transform and how to apply it to estimate the absolute velocity of the ego-vehicle. Finally, we conclude this letter.

Automotive radar system used in experiments: The automotive radar we used in this study consists of the transmit antenna (Tx), receiving antenna array (Rx), waveform generator (WG), voltage-controlled oscillator (VCO), frequency mixer, low-pass filter (LPF), analog-to-digital converter (ADC), and digital signal processor (DSP), as shown in Figure 1. The WG generates an FMCW with a center frequency of f_c, a bandwidth of δf, and a sweep time of δt. As shown in Figure 1, the waveform is composed of an up-chirp and a down-chirp whose frequency linearly increase and decreases with time, respectively. By extracting beat frequencies from each chirp, we can estimate the relative distance to the target and the relative velocity of the target [9]. In addition, the antenna array is used to estimate the angle of the target, which consists of 4 antenna elements. The distance between the antenna elements is $1.8\frac{c}{f_c}$, where c denotes the speed of light. In our radar system, the Bartlett algorithm [10] is applied to the signals received at Rx to estimate the angle of the target.

Target detection result through radar signal processing: We mounted the automotive radar sensor described in the previous section on the bumper of the vehicle and conducted experiments on the road. In our measurements, we set the values of f_c, δf, and δt to 76.5 GHz, 500 MHz,

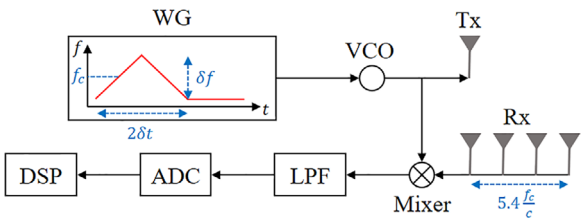


Fig. 1 *Configuration of the 77 GHz automotive radar system*

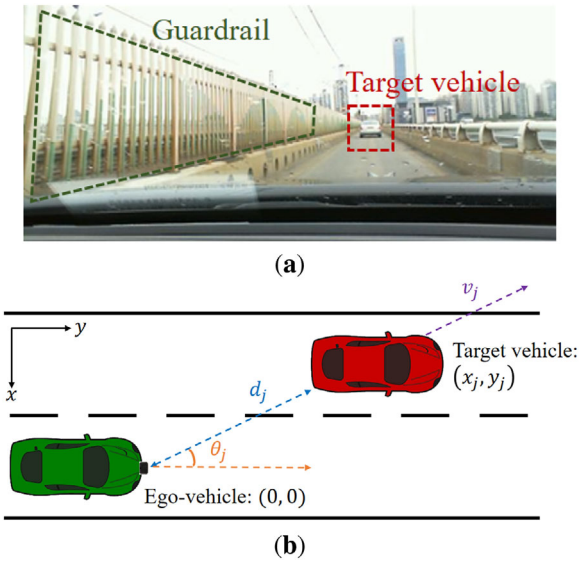


Fig. 2 *Target detection using the automotive radar: (a) radar signal measurement on the bridge (b) coordinate system conversion of the radar detection result*

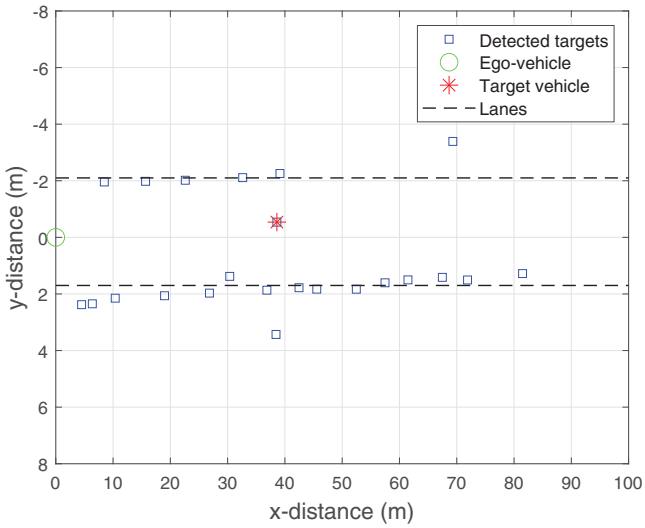


Fig. 3 *Target detection result on the bridge*

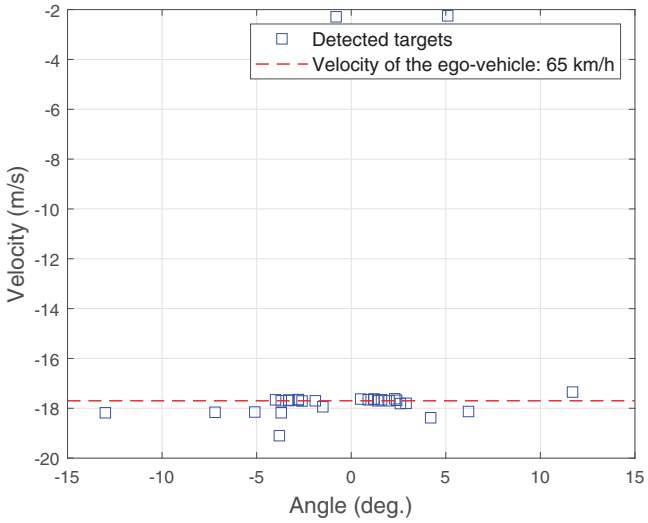


Fig. 4 *Target detection result in the angle-velocity domain*

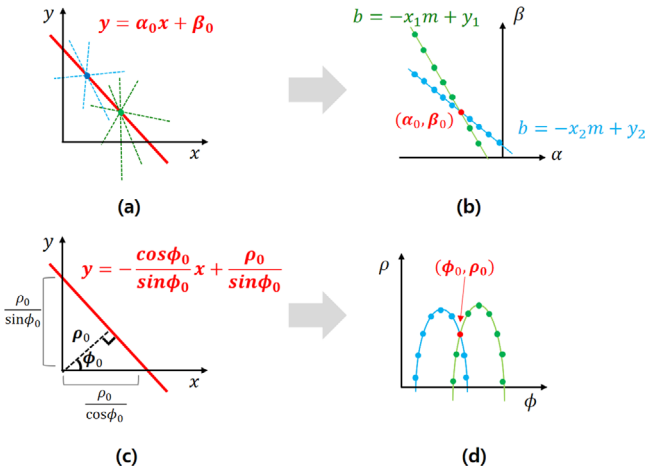


Fig. 5 *Basic concept of the Hough transform: (a) in the x–y domain; (b) in the α-β domain; (c) in the x-y domain; (**d**) in the ϕ-ρ domain*

and 5 ms, respectively. Figure 2a shows one of the road environments where we acquired radar sensor data. In this case, the velocity of the vehicle was maintained at 65 km/h. By processing the radar signal acquired in this environment, we can estimate the relative distance d_j, the relative velocity v_j in the radial direction, and the angle θ_j information of the target, where j is the index of each detected target. Then, using the target information represented by the polar coordinate system (i.e. $(d_j,\ \theta_j)$), the positions of the targets can be converted into a two-dimensional Cartesian coordinate system (i.e. $(x_j,\ y_j) = (d_j \sin\theta_j,\ d_j \cos\theta_j)$, as shown in Figure 2b. The target detection result corresponding to Figure 2a is shown in Figure 3. In this figure, including a target vehicle located in front, road structures such as guardrails around the ego-vehicle are also detected by the radar sensor.

Stationary and moving targets in the angle-velocity domain: To estimate the ego-velocity, it is important to select only fixed targets among the detected targets. This is because the relative velocity to a stationary target means the absolute velocity of the ego-vehicle. Referring to Figure 2b, the velocity of the target vehicle in the y-axis $v_{j,y}$ can be expressed as $v_j \sin\theta_j$. We redraw the detection result in Figure 3 in the angle-velocity (i.e. θ_j and $v_{j,y}$) domain, as shown in Figure 4. In this domain, moving and stationary targets exhibit different characteristics. The moving targets are randomly placed, but the stationary targets exist around a specific line. Therefore, if the line formed by the stationary targets is estimated, we can discriminate the stationary targets, and it is possible to estimate the absolute velocity of the ego-vehicle.

Hough transform based ego-velocity estimation: In this work, we use the Hough transform to find the tendency from the scattered points in the angle-velocity domain. The Hough transform is widely used to extract linear components from the image. For example, as shown in Figure 5a, if there are two points in the image space, a straight line with the two points can be expressed in the slope-intercept form as $y = \alpha_0 x + \beta_0$, where α_0 is the slope and β_0 is the y-intercept of the straight line. This line can also be expressed as a single point $(\alpha_0,\ \beta_0)$ in the parameter space, as shown in Figure 5b. In the parameter space, however, a vertical line parallel to the y-axis cannot be represented. Thus, in the Hough transform, this straight line is represented using the distance from the origin ρ_0 and the angle between the x-axis and ρ_0, as shown in Figure 5c, which means each straight line in the x-y domain can be expressed as a point in the ϕ-ρ domain. The straight lines passing through each point in Figure 5a is expressed as points on each curve in Figure 5d. Finally, the red point $(\rho_0,\ \phi_0)$ in Figure 5d indicates the point corresponding to a straight line passing through two points. Thus, from the point of intersection, we can estimate the straight line formed by the points through the Hough transform.

Figure 6 shows the result of applying the Hough transform to the detection result in the angle-velocity domain of Figure 4. In the Hough transform domain, several curves meet at one point $p_0 = (\rho_0,\ \phi_0) = (-17.7,\ 1.56)$. Thus, the straight line corresponding to p_0 is the most obvious straight line, which can be expressed as $x\cos\theta_0 + y\sin\theta_0 = \rho_0$ (i.e. $y = -0.0175x - 17.7$) in the original image space. We plot the

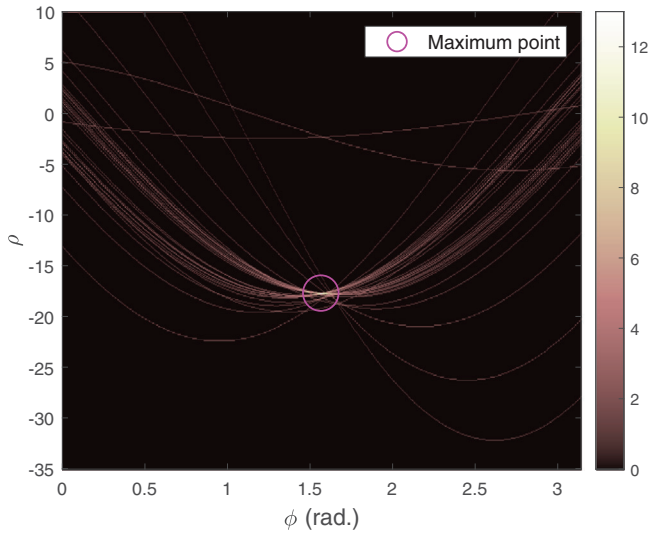


Fig. 6 *Result of applying the Hough transform to the detection result in the angle-velocity domain*

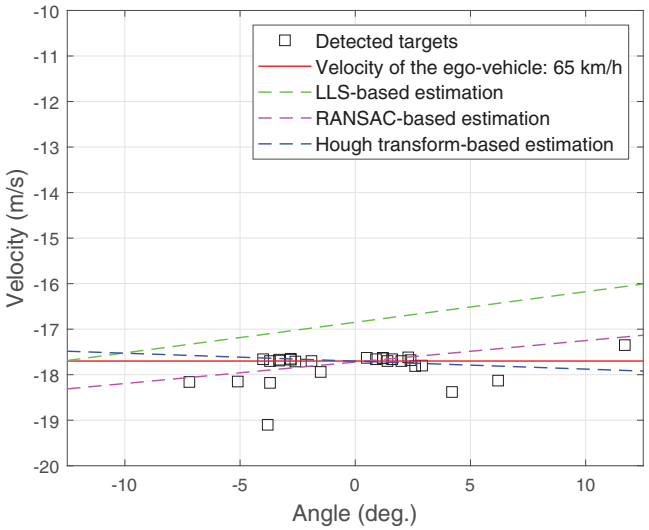


Fig. 7 *Ego-velocity estimation results in the angle-velocity domain*

estimated line back on the detection result in the angle-velocity domain, as shown in Figure 7. Because the slope of this line is close to zero, the ego-velocity can be estimated using the *y*-intercept. Thus, the absolute velocity of the ego-vehicle is estimated as 63.7 km/h, which is very close to the actual velocity of 65 km/h. Considering the concept of relative velocity, the absolute value of the *y*-intercept becomes the ego-velocity.

We also estimated the absolute velocity of the ego-vehicle by applying the LLS method and the RANSAC algorithm proposed in [11] to the detection result in angle-velocity domain, as shown in Figure 7. In the case of the LLS, the estimation result is not accurate compared to other methods because this method calculates a trend line using both detection results corresponding to stationary and moving targets. The RANSAC algorithm finds the trend line quite accurately based on inliers. However, in RANSAC-based methods, parameter values, such as the minimum sample size and maximum distance, must be set empirically, which is not required in the Hough transform. Finally, we verified the performance of our proposed method by changing the velocity of the ego-vehicle, as shown in Table 1. The estimation accuracy decreased slightly as the velocity increased, but the maximum error was 2.4% of the actual velocity. Through our proposed method, it is possible to accurately estimate the ego-velocity by automatically selecting points corresponding to stationary targets among several detection points.

Table 1. Ego-velocity estimation results according to the velocity of the ego-vehicle

Actual velocity (km/h)	35.0	45.0	55.0	65.0	75.0	85.0
Estimated velocity (km/h)	34.7	44.3	54.0	63.7	73.5	83.0

Conclusion: In this letter, we proposed a method for estimating the absolute velocity of the ego-vehicle using only the radar detection result. To estimate the velocity of the ego-vehicle, it is important to identify stationary targets from the radar detection results. Thus, we used the distribution of targets in the angle-velocity domain. In this domain, stationary targets existed on a specific line, and the Hough transformation was used to estimate the line. We verified the performance of the proposed method with the radar data measured on the actual road. Through the Hough transform, inliers corresponding to the stationary targets were effectively extracted among all the detection points. The proposed method has a maximum error of 2.4% in the velocity range of 30–90 km/h.

Acknowledgement: This work was supported by the 2020 Korea Aerospace University Faculty Research Grant and also by the Technology Innovation Program (or Industrial Strategic Technology Development Program, No. 10080086, Development of 77/79 GHz Dual Band Radar) funded By the Ministry of Trade, Industry and Energy (MOTIE, Korea).

Received: *10 August 2020* Accepted: *26 October 2020*
doi: 10.1049/ell2.12048

References

1 Hayakawa, J., Dariush, B.: Ego-motion and surrounding vehicle state estimation using a monocular camera. In: *IEEE Intelligent Vehicles Symposium (IV)*, Paris, France, 2019, pp. 2550–2556
2 Almeida, J., Santos, V. M.: Real time egomotion of a nonholonomic vehicle using LIDAR measurements. *J. Field Rob.* **30**(1), 129–141 (2013)
3 Boyd, S., Vandenberghe, L.: *Convex Optimization*, 1st edn. Cambridge University Press, UK (2004)
4 Fischler, M. A., Bolles, R. C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **24**(6), 381–395 (1981)
5 Hartley, R., Zisserman, A.: *Multiple View Geometry in Computer Vision*, 2nd edn. Cambridge University Press, UK (2004)
6 Duda, R. O., Hart, P. E.: Use of the Hough transformation to detect lines and curves in pictures. *Commun. ACM* **15**(1), 11–15 (1972)
7 Okamoto, Y., Matsunami, I., Kajiwara, A.: Moving vehicle discrimination using Hough transformation. In: *IEEE Radio and Wireless Symposium (RWS)*, Phoenix, AZ, 2011, pp. 367–370
8 Song, H., Shin, H.-C.: Classification and spectral mapping of stationary and moving objects in road environments using FMCW radar. *IEEE Access* **8**, 22955–22963 (2020)
9 Mahafza, B. R.: *Radar Systems Analysis and Design Using MATLAB*, 3rd edn. CRC Press, Florida (2013)
10 Gross, F.: *Smart Antennas for Wireless Communications with MATLAB*, 1st edn. McGraw-Hill, New York (2005)
11 Kellner, D., et al.: Instantaneous ego-motion estimation using Doppler radar. In: *The 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, The Hague, Netherlands, 2013, pp. 869–874

A novel hybrid termination structure for vertical gallium nitride Schottky barrier diode by using technology computer aided design simulation

Jian Li, Feng Lin, Yong Chen, Wei He, and Xinke Liu✉
College of Materials Science and Engineering, College of Electronics and Information Engineering, Shenzhen University, Shenzhen, P.R. China

✉**Correspondence**
Xinke Liu, College of Materials Science and Engineering, College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, P.R. China.
Email: xkliu@szu.edu.cn

Gallium nitride based high-power electronic devices are now in full swing. However, the phenomenon that the gallium nitride Schottky diodes break down prematurely without reaching the gallium nitride material limit is unsolved. This paper proposes a novel hybrid termination structure for vertical gallium nitride Schottky diodes to improve breakdown voltage. This work is carried out to simulate the breakdown voltage and reverse characteristics of the vertical gallium nitride Schottky diode by using technology computer aided design (TCAD) simulation. Under the same testing conditions, we demonstrate that compared with the control vertical Schottky diode, the breakdown voltage of the proposed Schottky diode can be significantly advanced, which has increased by 350 V and reached 850 V.

Introduction: Gallium nitride (GaN), the wide bandgap semiconductor, has superior properties in chemical stability, high breakdown voltage, low on-resistance and higher operating temperature compared to silicon (Si) and gallium arsenide (GaAs) [1]. GaN is mostly utilised in high-temperature, high-power and high-frequency environment, such as electric vehicle, information communication, aerospace and so on [2]. Therefore, the researches on GaN materials and devices have great strategic value as well as significance of development. As one of the main branches of GaN power electronic devices study, GaN-based Schottky barrier diodes (SBDs) have significant strengths in high-frequency and high-power applications for characteristics such as metal-semiconductor rectification characteristics, fast-switching speed and low-switching power loss [3]. Compared to horizontal GaN SBDs, the vertical GaN SBDs escape from the dilemma in lattice mismatch, thermal expansion mismatch and high surface defect density.

In vertical GaN SBDs, the effect of peak electric field prevails between the Schottky contact metal edge and the n-GaN drift layer [4]. As a result, the devices break down prematurely at the edge of anode contact area without reaching GaN material limit under reverse bias. Termination technologies are needed to improve the breakdown voltage. In recent years, many foreign and domestic scholars have proposed solutions for improving breakdown voltage. In 2017, Zheng et al. [5] gained a 2250 V reverse breakdown SBD with a 30 μm thick drift layer by using simulation software TCAD. Liu et al. [6] employed a CMOS process compatible TiN electrode to achieve a 1200 V breakdown characteristic with a 13 μm drift layer thickness. Yang et al. [7] took advantage of nitrogen treatment to form termination structure and successfully obtained a SBD with reverse 995 V breakdown voltage. In 2019, Gu et al. [3, 8] acquired a self-supporting germanium-doped GaN SBD with 802 V breakdown voltage and 15.8 ns reverse recovery time (Trr). To sum up, vertical GaN SBDs are still in the development stage.

This paper puts forward a novel hybrid termination structure (HTS) for vertical free-standing GaN SBD to enhance the breakdown voltage. The aims are to simulate and analyse the device structure as well as its electrical characteristics by utilising simulation software TCAD. Finally, the breakdown voltage of the proposed device is almost 350 V higher than the GaN SBD with control structure (CS), and its Trrs are 1.6 and 1.8 ns at 300 and 500 K, respectively.

Structures and characteristics: The HTS and the CS of GaN SBD are shown in Figure 1(a) and (b), respectively. In the device simulation of HTS, we first construct 390 μm thickness germanium-doped GaN substrate with a doping concentration of 1×10^{18} cm^{-3}, and deposit

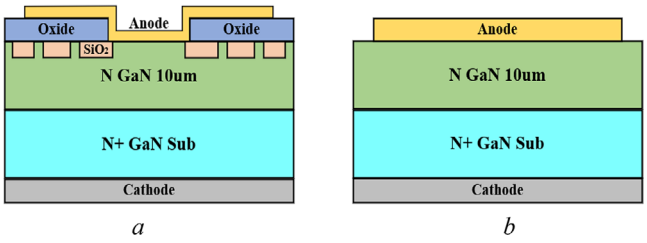


Fig. 1 *GaN SBD structure: (a) the hybrid termination structure (HTS) GaN SBD and (b) the control structure (CS) GaN SBD*

10 μm thickness n-GaN drift layer with a doping concentration of 8×10^{15} cm^{-3}. Later, we etch numbers of 0.5 μm depth grooves with variable length in the drift layer, deposit silicon dioxide (SiO_2) and then etch the excess SiO_2 to form guard oxide ring (GOR) structure. A 0.4 μm thick oxide dielectric layer is deposited soon, and a middle groove is etched for anode. Ultimately, the cathode contact electrode Al is evaporated at the bottom of the device, and the anode Schottky contact electrode Au is evaporated at the top to form field plate (FP) structure. The CS GaN SBD is fabricated by an identical process for comparison.

The physical simulation models are the parallel electric field dependent mobility model, concentration-dependent mobility model, Shockley–Read–Hall recombination model, impact selb model, energy bandgap narrowing model and so on.

In Figure 2(a), the peak electric field at the edge of anode contact area is compared between HTS and CS GaN SBD at 500 V reverse voltage. The peak electric field of HTS is smaller, which indicates that HTS can impair the crowding effect of electric field at the interface contact between Schottky and GaN. In Figure 2(b), the leakage current of HTS is greater because HTS introduces defects of Oxide/GaN interface and damage caused by etching. Under growing reverse voltage, the breakdown voltage of HTS is about 350 V higher than CS, which proves the validity of the proposed termination structure.

Electric field formula, capacitance formula and conductance formula are as follows:

$$E = U/d = 4\pi kQ/\varepsilon S \tag{1}$$

$$C = \varepsilon S/4\pi kd \tag{2}$$

$$G = 1/R = S/L\rho. \tag{3}$$

The relationship between them are as follows:

$$C = Q/Ed \tag{4}$$

$$G = 4\pi kdC/\varepsilon\rho \tag{5}$$

In the above formulas (1)–(5), where Q is the quantity of electricity, k denotes the electrostatic force constant, ε represents the dielectric constant, S measures the contact area, d is the distance, G is the conductance, R implies the resistance, L is the length of resistor and ρ means the resistivity.

The capacitance-voltage (CV) characteristic curve and conductance comparison diagram at 300 K temperature and 1 MHz frequency are shown in Figure 3(a) and (b).

For CS GaN SBD, electrons accumulate in the semiconductor below anode will result small capacitance under large negative voltage. When the voltage becomes larger and reaches about −3.8 V, the accumulated electrons are quickly depleted to form space charge region, leading to rapid increasement of capacitance. As voltage advances further, electrons are completely depleted, which slows down the expansion of space charge region and flats the trend of CV curve. Similarly, the conductance diagram shows the same trend.

However, considering HTS GaN SBD, the anode contact area S is reduced due to the introduction of FP and GOR structure. The capacitance is smaller than CS GaN SBD under large negative voltage. The advancing maximum amplitude of capacitance is also less than the other for the introduction of defects and damage with enlarging voltage. Furthermore, the charging mechanism consumes electrons contributing to less Q in formula (4) because of the parasitic capacitances in HTS.

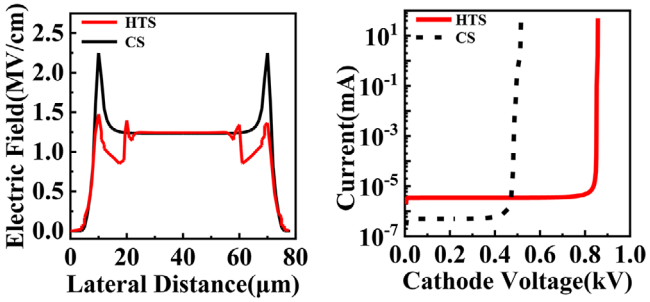


Fig. 2 *At 500 V reverse voltage (a) electric field value and (b) reverse breakdown voltage*

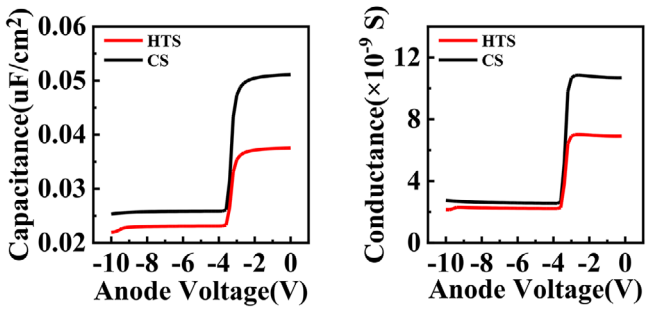


Fig. 3 *At 300 K temperature and 1 MHz frequency: (a) capacitance-voltage characteristic curve and (b) conductance comparison diagram*

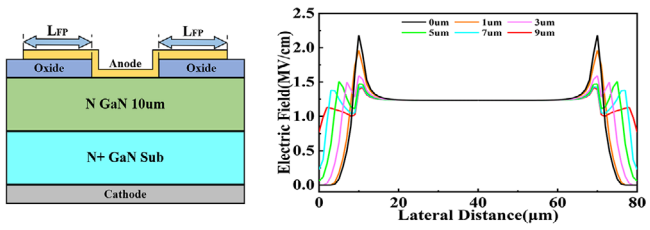


Fig. 4 *Under 500 V reverse voltage: (a) the GaN SBD structure with field plate (FP) length* L_{FP} *and (b) the electric field comparison chart of various* L_{FP}

Mechanism research and analysis:

Study the effect on the device electric field among various FP length: Various FP lengths on oxide dielectric layer have great difference in the electrical performance of SBD. Figure 4(a) shows the GaN SBD structure with FP length (L_{FP}). The corresponding L_{FP} are 0, 1, 3, 5, 7 and 9 μm.

The electric field of GaN SBD with different L_{FP} under 500 V reverse voltage is shown in Figure 4(b). There are two peak electric fields. One is generated by GaN. The peak electric field at the edge of anode contact area is defined as peak junction electric field, which is located at 10 and 70 μm on the abscissa in the figure. The other is produced by FP structure. With L_{FP} increasing, the variable ending position of FP makes the peak electric field shift, which is recognised as offset electric field.

For 1 μm L_{FP}, the electric field is larger than devices with other L_{FP} due to the superposition of junction electric field and offset electric field. It is considered that the FP structure is out of operation, but the electric field is still smaller than GaN SBD with 0 μm L_{FP}. From 3 to 9 μm L_{FP}, the offset electric field first increases and then decreases, while the turning point occurs at 5 μm L_{FP}. When L_{FP} is 9 μm, the two peak electric field are the smallest.

According to formula (1), it can be known that under the same reverse voltage, the longer L_{FP} leads to greater depletion layer distance d, more uniform distribution of electric field and smaller peak electric field.

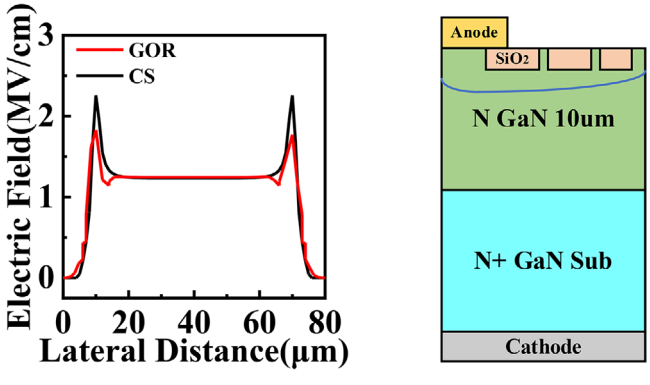


Fig. 5 *Under 500 V reverse voltage: (a) peak electric field comparison diagram and (b) MOS structure diagram*

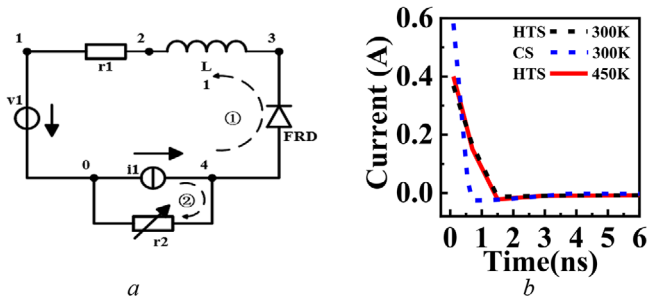


Fig. 6 *Reverse recovery property: (a) schematic diagram of GaN SBD reverse recovery circuit and (b) reverse recovery diagram under variable temperature*

Under advancing reverse voltage, it can withstand larger reverse voltage without breakdown.

Study the influence of GOR structure on the device electric field: CS GaN SBDs generally have peak junction electric field. The employment of GOR structures can extend the depletion layer at the peak junction electric field to the GOR and connect with it. Hence, the increased reverse voltage is affordable by GOR, which is also known as voltage divider ring. The comparison chart at their peak electric field is shown in Figure 5(a), which demonstrates that the utilisation of GOR structure can weaken the peak electric field and improve the withstand voltage.

In addition, the employment of GOR is equivalent to forming a high-resistance structure, which is the same as metal oxide semiconductor (MOS) structure [9]. In Figure 5(b), the Schottky junction is reverse biased and the MOS structure is depleted under reverse voltage. As the voltage develops, the adjacent oxidised structure gradually contacts with each other, which restrains the barrier lowering and reduces the leakage current. Thus, the current will not enhance sharply at reverse bias.

Reverse recovery property: The Trr means that when the diode suddenly switches from stable forward voltage to reverse voltage, the time required for the minority carriers decreases from the accumulation state to the depletion state. The ideal SBD only has majority carriers, so its Trr is theoretically zero. However, considering the capacitances in contact electrode and oxide layer, they spend time fully charging and discharging. In total, the required time is shorter than positive negative (PN) junction diode. Therefore, the SBD has not only smaller reverse current but also faster switching speed [10]. Figure 6(a) shows a simplified reverse recovery circuit of vertical HTS GaN SBD. The loops ① and ② are the circuit under the on-state and the reverse recovery, respectively.

Via the MixedMode of TCAD, as shown in Figure 6(b), the Trr obtained at 300 and 450 K are 1.6 and 1.8 ns, respectively. The carriers generated by its intrinsic excitation at room temperature are negligible for the large bandgap (about 3.4 eV) of GaN material. Instead, the carriers mainly come from impurity ionisation. When the temperature rises, although it will stimulate the intrinsic excitation, the generated

carriers are limited. The main carriers are still from impurity ionisation. So the Trr at 500 K is only 0.2 ns longer than the Trr at 300 K. In contrast, for Si-based diodes, when the temperature rises from 25°C to 175°C, the Trr greatly increases from 26.6 to 116 ns [11]. This illustrates that GaN is suitable for high-temperature applications owing to its good thermal stability.

Conclusion: This paper realises a vertical GaN SBD with a new HTS composed of GOR and FP through TCAD simulation, which obviously improves the breakdown voltage. The mechanism to weaken the crowding peak electric field by introducing FP structure and GOR structure are fully demonstrated. This simulation on the electrical performance and reverse characteristics of the new HTS GaN SBD has certain significant reference for future work.

Acknowledgements: This work was supported by National Key Research and Development Program of China (2017YFB0404100), National Natural Science Foundation of China (61974144), Key-Area Research and Development Program of Guangdong Province (2020B010169001), the Science and Technology Foundation of Shenzhen (JSGG20191129114216474) and the National Taipei University of Technology-Shenzhen University Joint Research Program (2020009). This paper includes all data generated or analysed during this study. The authors declare no competing financial interest.

Author contributions: Jian Li: methodology, project administration, validation, visualisation, writing original draft and review and editing; Feng Lin: investigation, resources and software; Yong Chen: data curation and resources; Wei He: funding acquisition and supervision; Xinke Liu: funding acquisition, methodology, supervision and review and editing.

Received: *30 October 2020* Accepted: *5 October 2020*
doi: 10.1049/ell2.12067

References

1 Kachi, T.: Current status of GaN power devices. *IEICE Electron. Express* **10**(21), 20132005–20132005 (2013).
2 Vurgaftman, I., Meyer, J.R., Ram-Mohan, L.R.: Band parameters for III–V compound semiconductors and their alloys. *J. Appl. Phys.* **89**(11), 5815–5875 (2001).
3 Gu, H. et al.: Vertical GaN Schottky barrier diodes on Ge-doped free-standing GaN substrates. *J. Alloys Compd.* **780**, 476–481 (2019).
4 Sun, Y. et al.: Review of the recent progress on GaN-based vertical power Schottky barrier diodes (SBDs). *Electronics* **8**(5), 575 (2019).
5 Zheng, L. et al.: Vertical power diodes based on bulk GaN substrates. 2017 IEEE 12th International Conference on ASIC (ASICON), pp. 163–166. IEEE, Piscataway, NJ (2017).
6 Liu, X. et al.: 1.2 kV GaN Schottky barrier diodes on free-standing GaN wafer using a CMOS-compatible contact material. *Jpn. J. Appl. Phys.* **56**(2), 026501 (2017).
7 Yang, S. et al.: 1 kV/1.3 mΩ·cm^2 vertical GaN-on-GaN Schottky barrier diodes with high switching performance. 2018 IEEE 30th International Symposium on Power Semiconductor Devices and ICs (ISPSD), pp. 272–275. IEEE, Piscataway, NJ (2018).
8 Gu, H. et al.: Recovery performance of Ge-doped vertical GaN Schottky barrier diodes. *Nanoscale Res. Lett.* **14**(1), 40 (2019).
9 Azam, F. et al.: Engineering a unified dielectric solution for AlGaN/GaN MOS-HFET gate and access regions. *IEEE Trans. Electron Devices* **67**(3), 881–887, (2020).
10 Han, S. et al.: Current-collapse-free and fast reverse recovery performance in vertical GaN-on-GaN Schottky barrier diode. *IEEE Trans. Power Electron.* **34**(6), 5012–5018 (2018).
11 Lee, J.-H. et al.: AlGaN/GaN-based lateral-type Schottky barrier diode with very low reverse recovery charge at high temperature. *IEEE Trans. Electron Devices* **60**(10), 3032–3039 (2013).

On the steady-state performance of bias-compensated LMS algorithm

Rodrigo Pimenta,[1,✉] Leonardo Resende,[2] Mariane R. Petraglia,[3] and Diego B. Haddad[1]
[1]*CEFET/RJ, Rio de Janeiro, Brazil*
[2]*IFRJ, Rio de Janeiro, Brazil*
[3]*UFRJ, Rio de Janeiro, Brazil*
✉E-mail: rodrigo.pimenta@cefet-rj.br

Adaptive filtering algorithms are widespread today owing to their flexibility and simplicity. Due to environments in which they are normally immersed, their robustness against noise has been a topic of interest. Traditionally, in the literature it is assumed that noise is mainly active in the reference signal. Since this hypothesis is often violated in practice, recently some papers have advanced strategies to compensate the bias introduced by noisy excitation data. The contributions of this paper are twofold. The first one establishes that, in some conditions, the bias-compensated least mean square algorithm implements an optimum estimator, in the sense that it presents the smallest variance of the set of unbiased estimators. Since the asymptotic mean-square performance of this algorithm has not yet been investigated in detail, the second contribution adopts an energy conservation relationship to derive its theoretical steady-state mean squared distortion. The final result is presented in a closed form, is consistent with simulations and is able to provide important guidelines to the designer.

Introduction: Adaptive filtering is a versatile tool that solves a wide range of real-world problems where input and output data are available. This paper focuses on system identification tasks, in which an adaptive filter aims at emulating an ideal (and unknown) transfer function, whose coefficients are collected in vector[1] $\boldsymbol{w}^\star \in \mathbb{R}^N$.

In supervised contexts, a sequence of noisy scalar measurements $\{d(k) \in \mathbb{R}\}_{k=0}^{\infty}$ is supposed to be accessible (sometimes through a creative trick). A linear-in-the-parameters model is considered, such that the reference signal $d(k)$ satisfies

$$d(k) = \left[\boldsymbol{w}^\star\right]^T \boldsymbol{u}(k) + v(k), \tag{1}$$

where $v(k)$ accounts for measurement additive noise and

$$\boldsymbol{u}(k) \triangleq \left[u(k)\ u(k-1)\ \ldots\ u(k-N+1)\right]^T \tag{2}$$

concatenates N successive excitation data samples. An adaptive algorithm updates the filter weight vector $\boldsymbol{w}(k) \in \mathbb{R}^N$ in a sample-by-sample basis using the error signal

$$e(k) \triangleq d(k) - y(k), \tag{3}$$

where $y(k) \triangleq \boldsymbol{w}^T(k)\boldsymbol{u}(k)$ denotes the output of the adaptive filter at the kth iteration.

By adopting the mean square error $\xi(k) \triangleq \frac{1}{2}\mathbb{E}[e^2(k)]$ as the cost function, the standard least mean squares (LMS) algorithm can be derived using the stochastic gradient method, that is,

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) - \beta\nabla_{\boldsymbol{w}(k)}[\xi(k)] = \boldsymbol{w}(k) + \beta\boldsymbol{u}(k)e(k), \tag{4}$$

where $\beta \in \mathbb{R}_+$ denotes the step size, whose choice creates a trade-off between convergence rate and asymptotic performance. It is widely known that the LMS algorithm performs an asymptotically unbiased estimate of the coefficient vector [1].

[1]In this paper, all vectors are of column-type.

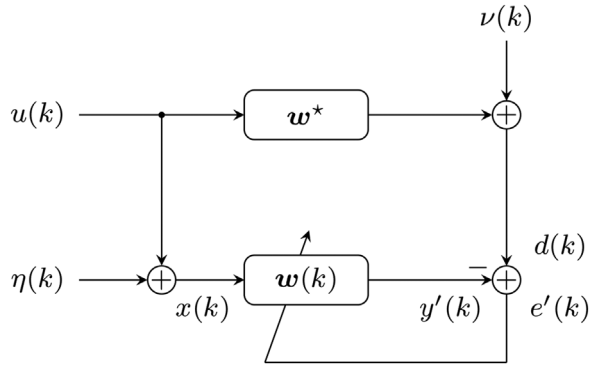


Fig. 1 *Block diagram of an adaptive system identification problem, where the adaptive filter input is corrupted by noise $\eta(k)$*

Unfortunately, the classic algorithm in (4) assumes that the input data is measured with infinite precision. The noisy characteristic of the excitation data can be taken into account assuming that the measured input vector $\boldsymbol{x}(k) \in \mathbb{R}^N$ can be written as

$$\boldsymbol{x}(k) = \boldsymbol{u}(k) + \boldsymbol{\eta}(k), \tag{5}$$

where $\boldsymbol{\eta}(k) \triangleq \big[\eta(k)\ \eta(k-1)\ \ldots\ \eta(k-N+1)\big]^T$ contains N samples of the input noise component $\eta(k)$, which is supposed to be stationary, zero-mean and with variance σ_η^2. Using $\boldsymbol{x}(k)$ as the input signal, one leads to the *noisy* LMS (see Figure 1):

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta \boldsymbol{x}(k) e'(k), \tag{6}$$

where

$$e'(k) \triangleq d(k) - y'(k) = d(k) - \boldsymbol{w}^T(k)\boldsymbol{x}(k) \tag{7}$$

is the error signal that employs the (observable) noisy input $\boldsymbol{x}(k)$, instead of the ideal one (i.e. $u(k)$).

Remark. Measurement noise at the input of the LMS algorithm induces a bias in the resulting steady-state solution [2].

Classical derivation of the bias-constrained LMS: The asymptotic bias resulting from (6) can be compensated by inserting a random perturbation $\boldsymbol{\psi}(k)$ in the update term [2, 3], that is,

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta \boldsymbol{x}(k) e'(k) + \boldsymbol{\psi}(k). \tag{8}$$

In order to derive an appropriate perturbation term $\boldsymbol{\psi}(k)$, the deviation vector $\tilde{\boldsymbol{w}}(k) \triangleq \boldsymbol{w}^\star - \boldsymbol{w}(k)$ is considered. Using (5), (7) and (8), the following recursion for $\tilde{\boldsymbol{w}}(k)$ is obtained:

$$\begin{aligned}\tilde{\boldsymbol{w}}(k+1) = {} & \tilde{\boldsymbol{w}}(k) - \beta \boldsymbol{x}(k)\boldsymbol{u}^T(k)\tilde{\boldsymbol{w}}(k) - \beta \boldsymbol{x}(k)\nu(k) \\ & + \beta \boldsymbol{x}(k)\boldsymbol{\eta}^T(k)\boldsymbol{w}(k) - \boldsymbol{\psi}(k).\end{aligned} \tag{9}$$

A choice for $\boldsymbol{\psi}(k)$ which guarantees steady-state unbiasedness can be achieved by using the following usual statistical hypotheses:

H1. The input data is statistically independent of the adaptive weight vector $\boldsymbol{w}(k)$.
H2. The zero-mean stochastic sequences $\{u(k), \eta(k), \nu(k)\}$ are statistically independent.
H3. The algorithm converges, so that $\lim_{k\to\infty} \mathbb{E}[\tilde{\boldsymbol{w}}(k+1)] = \mathbb{E}[\tilde{\boldsymbol{w}}(k)]$ or $\lim_{k\to\infty} \mathbb{E}[\|\tilde{\boldsymbol{w}}(k+1)\|^2] = \mathbb{E}[\|\tilde{\boldsymbol{w}}(k)\|^2]$.

Remark. Commonly referred to as *independence heuristic* [4], H1 is very popular in the analyses of adaptive filtering algorithms. Unfortunately, its application results in non-accurate prediction when large step sizes are adopted [5].

Taking expected value of (9) and using H1–H3, the steady-state bias of the solution of (8) can be written as

$$\lim_{k\to\infty} \mathbb{E}[\tilde{\boldsymbol{w}}(k)] = \frac{1}{\beta}\boldsymbol{R}_u^{-1}\big[\beta\sigma_\eta^2 \boldsymbol{w}(k) - \boldsymbol{\psi}(k)\big], \tag{10}$$

where $\boldsymbol{R}_u \triangleq \mathbb{E}[\boldsymbol{u}(k)\boldsymbol{u}^T(k)]$. This steady-state bias can be zeroed by applying $\boldsymbol{\psi}(k) = \beta\sigma_\eta^2 \boldsymbol{w}(k)$, which provides the update equation of the BC-LMS algorithm

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta \boldsymbol{x}(k) e'(k) + \beta\sigma_\eta^2 \boldsymbol{w}(k), \tag{11}$$

where a perfect estimation of σ_η^2 is assumed. When an inaccurate estimate $\hat{\sigma}_\eta^2$ of such variance is employed, the update equation is described by

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta \boldsymbol{x}(k) e'(k) + \beta\hat{\sigma}_\eta^2 \boldsymbol{w}(k), \tag{12}$$

resulting in an asymptotic mean deviation vector given by

$$\lim_{k\to\infty} \mathbb{E}[\tilde{\boldsymbol{w}}(k)] = \big(\boldsymbol{R}_x - \hat{\sigma}_\eta^2 \boldsymbol{I}\big)^{-1}\big(\sigma_\eta^2 - \hat{\sigma}_\eta^2\big)\boldsymbol{w}^\star, \tag{13}$$

where $\boldsymbol{R}_x \triangleq \mathbb{E}[\boldsymbol{x}(k)\boldsymbol{x}^T(k)]$.

Remarks. The deviation vector converges in the mean to a value that does not depend on the step size. In the following, an ideal estimate of the variance of the input noise is assumed.

Mean-square optimality of the BC-LMS: The asymptotic unbiasedness of the BC-LMS algorithm is related to the first-order statistical constraint $\lim_{k\to\infty} \mathbb{E}[\tilde{\boldsymbol{w}}(k)] = \boldsymbol{0}$. In the following, it will be demonstrated that, under some statistical assumptions, such an estimator is the one (among all possible unbiased estimators) that minimises the mean square distortion (MSD) in steady state. This fact is important to establish the optimality of the BC-LMS estimator in a more general sense.

From (9), one obtains

$$\begin{aligned}\|\tilde{\boldsymbol{w}}(k+1)\|^2 = {} & -2\boldsymbol{\psi}^T(k)\tilde{\boldsymbol{w}}(k) + 2\beta\boldsymbol{\psi}^T(k)\boldsymbol{x}(k)\boldsymbol{u}^T(k)\tilde{\boldsymbol{w}}(k) \\ & -2\beta\boldsymbol{\psi}^T(k)\boldsymbol{x}(k)\boldsymbol{\eta}^T(k)\boldsymbol{w}(k) + 2\beta\boldsymbol{\psi}^T(k)\boldsymbol{x}(k)\nu(k) \\ & +\boldsymbol{\psi}^T(k)\boldsymbol{\psi}(k) + \mathcal{O}(k),\end{aligned} \tag{14}$$

where $\mathcal{O}(k)$ contains terms that do not depend on $\boldsymbol{\psi}(k)$. It is possible to minimise $\|\tilde{\boldsymbol{w}}(k+1)\|^2$ by zeroing its gradient with respect to $\boldsymbol{\psi}(k)$, which leads to

$$\boldsymbol{\psi}(k) = -\tilde{\boldsymbol{w}}(k) - \beta\boldsymbol{x}(k)\boldsymbol{u}^T(k)\tilde{\boldsymbol{w}}(k) + \beta\boldsymbol{x}(k)\boldsymbol{\eta}^T(k)\boldsymbol{w}(k) - \boldsymbol{x}(k)\nu(k). \tag{15}$$

Such solution is unfeasible in practice, since $\tilde{\boldsymbol{w}}(k)$, $\boldsymbol{u}(k)$, $\boldsymbol{\eta}(k)$ and $\nu(k)$ are not available. This problem can be circumvented by imposing that (15) is true *on average*, that is, applying the expectation operator to (15) and using H1-H3, one obtains

$$\mathbb{E}[\boldsymbol{\psi}(k)] = \beta\sigma_\eta^2 \mathbb{E}[\boldsymbol{w}(k)]. \tag{16}$$

Such constraint can be enforced by the feasible choice $\boldsymbol{\psi}(k) = \beta\sigma_\eta^2 \boldsymbol{w}(k)$, which results in the BC-LMS algorithm (11).

Steady-state MSD: The MSD, given by

$$\mathrm{MSD}(k) = \varsigma(k) \triangleq \mathbb{E}\big[\|\boldsymbol{w}^\star - \boldsymbol{w}(k)\|^2\big], \tag{17}$$

is a popular performance metric in the adaptive filtering field. Motivated by the fact that it is important to predict the filter performance after reaching steady state, the second contribution of this paper employs the energy flow method in order to derive the asymptotic MSD [6]. The following definitions will be employed henceforth:

$$e_a(k) \triangleq d(k) - \boldsymbol{w}^T(k)\boldsymbol{x}(k), \tag{18}$$

$$e_p(k) \triangleq d(k) - \boldsymbol{w}^T(k+1)\boldsymbol{x}(k), \tag{19}$$

$$\tilde{e}_a(k) \triangleq \tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k), \tag{20}$$

$$\tilde{e}_p(k) \triangleq \tilde{\boldsymbol{w}}^T(k+1)\boldsymbol{x}(k). \tag{21}$$

$$\bar{e}_a(k) \triangleq \tilde{\boldsymbol{w}}^T(k)\boldsymbol{u}(k), \tag{22}$$

$$\bar{e}_p(k) \triangleq \tilde{\boldsymbol{w}}^T(k+1)\boldsymbol{u}(k). \tag{23}$$

Therefore, from (11) one obtains

$$\tilde{\boldsymbol{w}}(k+1) = (1+\beta\sigma_\eta^2)\tilde{\boldsymbol{w}}(k) - \beta e_a(k)\boldsymbol{x}(k) - \beta\boldsymbol{w}^\star\sigma_\eta^2. \tag{24}$$

Multiplying both sides of (24) on the left by $\boldsymbol{x}^T(k)$ and performing some substitutions lead to

$$\beta e_a(k) = -\frac{\tilde{e}_p(k)}{\|\boldsymbol{x}(k)\|^2} + \frac{(1+\beta\sigma_\eta^2)\tilde{e}_a(k)}{\|\boldsymbol{x}(k)\|^2} - \frac{\beta\boldsymbol{x}^T(k)\boldsymbol{w}^\star\sigma_\eta^2}{\|\boldsymbol{x}(k)\|^2}. \tag{25}$$

Combining (25) and (24) yields

$$\begin{aligned}\tilde{\boldsymbol{w}}(k+1) &+ \frac{(1+\beta\sigma_\eta^2)\tilde{e}_a(k)\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2} + \beta\boldsymbol{w}^\star\sigma_\eta^2 \\ &= (1+\beta\sigma_\eta^2)\tilde{\boldsymbol{w}}(k) + \frac{\tilde{e}_p(k)\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2} + \frac{\beta\boldsymbol{x}^T(k)\boldsymbol{w}^\star\sigma_\eta^2\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2}.\end{aligned} \tag{26}$$

Considering $\|\boldsymbol{x}\|^2$ as the energy of vector $\boldsymbol{x}$, the energy calculation of the vectors of both sides of (26) results in

$$\begin{aligned}&\|\tilde{\boldsymbol{w}}(k+1)\|^2 + 2\frac{(1+\beta\sigma_\eta^2)\tilde{e}_a(k)\tilde{\boldsymbol{w}}^T(k+1)\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2} \\ &+ 2\beta\sigma_\eta^2\tilde{\boldsymbol{w}}^T(k+1)\boldsymbol{w}^\star + \frac{(1+\beta\sigma_\eta^2)^2\tilde{e}_a^2(k)}{\|\boldsymbol{x}(k)\|^2} \\ &+ 2\frac{\beta\sigma_\eta^2(1+\beta\sigma_\eta^2)\tilde{e}_a(k)[\boldsymbol{x}^T(k)\boldsymbol{w}^\star]}{\|\boldsymbol{x}(k)\|^2} + \beta^2\sigma_\eta^4(\boldsymbol{w}^\star)^T\boldsymbol{w}^\star \\ &= (1+\beta\sigma_\eta^2)^2\|\tilde{\boldsymbol{w}}(k)\|^2 + 2\frac{(1+\beta\sigma_\eta^2)\tilde{e}_p(k)\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2} \\ &+ 2\frac{\beta\sigma_\eta^2(1+\beta\sigma_\eta^2)[\boldsymbol{x}^T(k)\boldsymbol{w}^\star]\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2} + \frac{\tilde{e}_p^2(k)}{\|\boldsymbol{x}(k)\|^2} \\ &+ 2\frac{\beta\sigma_\eta^2\tilde{e}_p(k)[\boldsymbol{x}^T(k)\boldsymbol{w}^\star]}{\|\boldsymbol{x}(k)\|^2} + \frac{\beta^2\sigma_\eta^4[\boldsymbol{x}^T(k)\boldsymbol{w}^\star]^2}{\|\boldsymbol{x}(k)\|^2}.\end{aligned} \tag{27}$$

By defining

$$A \triangleq 1 + \beta\sigma_\eta^2 - \beta\|\boldsymbol{x}(k)\|^2, \tag{28}$$

$$C \triangleq [(\boldsymbol{w}^\star)^T\boldsymbol{\eta}(k) - \nu(k)]\|\boldsymbol{x}(k)\|^2, \tag{29}$$

$$D \triangleq \boldsymbol{x}^T(k)\boldsymbol{w}^\star\sigma_\eta^2, \tag{30}$$

so that

$$\tilde{e}_p(k) = A\tilde{e}_a(k) + \beta[C - D], \tag{31}$$

and using (20) and (21), then (27) can be rewritten as

$$\begin{aligned}&\|\tilde{\boldsymbol{w}}(k+1)\|^2 + 2\beta\sigma_\eta^2\tilde{\boldsymbol{w}}^T(k+1)\boldsymbol{w}^\star + \frac{(1+\beta\sigma_\eta^2)^2\tilde{e}_a^2(k)}{\|\boldsymbol{x}(k)\|^2} + \beta^2\sigma_\eta^4\|\boldsymbol{w}^\star\|^2 \\ &= (1+\beta\sigma_\eta^2)^2\|\tilde{\boldsymbol{w}}(k)\|^2 + \frac{A^2\tilde{e}_a^2(k)}{\|\boldsymbol{x}(k)\|^2} + \frac{2A\tilde{e}_a(k)\beta(C-D)}{\|\boldsymbol{x}(k)\|^2} \\ &+ \frac{\beta^2(C-D)^2}{\|\boldsymbol{x}(k)\|^2} + 2\frac{\beta\sigma_\eta^2A\tilde{e}_a(k)[\boldsymbol{x}^T(k)\boldsymbol{w}^\star]}{\|\boldsymbol{x}(k)\|^2} \\ &+ 2\frac{\beta^2\sigma_\eta^2(C-D)(\boldsymbol{x}^T(k)\boldsymbol{w}^\star)}{\|\boldsymbol{x}(k)\|^2} + \frac{\beta^2\sigma_\eta^4[\boldsymbol{x}^T(k)\boldsymbol{w}^\star]^2}{\|\boldsymbol{x}(k)\|^2}.\end{aligned} \tag{32}$$

After several algebraic manipulations and simplifications, (32) results in

$$\begin{aligned}&\|\tilde{\boldsymbol{w}}(k+1)\|^2 + 2\beta\sigma_\eta^2\tilde{\boldsymbol{w}}^T(k+1)\boldsymbol{w}^\star + \beta^2\sigma_\eta^4\|\boldsymbol{w}^\star\|^2 \\ &= (1+\beta\sigma_\eta^2)^2\|\tilde{\boldsymbol{w}}(k)\|^2 - 2\beta\tilde{e}_a^2(k) - 2\beta^2\sigma_\eta^2\tilde{e}_a^2(k) \\ &+ \beta^2\|\boldsymbol{x}(k)\|^2\tilde{e}_a^2(k) + 2\beta(\boldsymbol{w}^\star)^T\boldsymbol{\eta}(k)\tilde{e}_a(k) - 2\beta\nu(k)\tilde{e}_a(k) \\ &+ 2\beta^2\sigma_\eta^2(\boldsymbol{w}^\star)^T\boldsymbol{\eta}(k)\tilde{e}_a(k) - 2\beta^2\sigma_\eta^2\nu(k)\tilde{e}_a(k) \\ &- 2\beta^2(\boldsymbol{w}^\star)^T\boldsymbol{\eta}(k)\|\boldsymbol{x}(k)\|^2\tilde{e}_a(k) + 2\beta^2\nu(k)\|\boldsymbol{x}(k)\|^2\tilde{e}_a(k) \\ &+ \beta^2[(\boldsymbol{w}^\star)^T\boldsymbol{\eta}(k)]^2\|\boldsymbol{x}(k)\|^2 - 2\beta^2(\boldsymbol{w}^\star)^T\boldsymbol{\eta}(k)\|\boldsymbol{x}(k)\|^2\nu(k) \\ &+ \beta^2\nu^2(k)\|\boldsymbol{x}(k)\|^2,\end{aligned} \tag{33}$$

which is an *exact* energy-conservation relationship that is strictly valid for every iteration (i.e. it is not restricted to the steady-state regime).

By applying the hypotheses H1 and H2 to the expectation operator in (33), one may establish the following relationship:

$$\begin{aligned}&\mathbb{E}\big[\|\tilde{\boldsymbol{w}}(k+1)\|^2\big] + 2\beta\sigma_\eta^2\mathbb{E}\big[\tilde{\boldsymbol{w}}^T(k+1)\boldsymbol{w}^\star\big] + \beta^2\sigma_\eta^4\mathbb{E}\big[\|\boldsymbol{w}^\star\|^2\big] \\ &= (1+\beta\sigma_\eta^2)^2\mathbb{E}\big[\|\tilde{\boldsymbol{w}}(k)\|^2\big] - 2\beta\mathbb{E}\big[[\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)]^2\big] \\ &- 2\beta^2\sigma_\eta^2\mathbb{E}\big[[\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)]^2\big] + \beta^2\mathbb{E}\big[\|\boldsymbol{x}(k)\|^2[\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)]^2\big] \\ &+ \beta^2\sigma_\eta^2\mathbb{E}\big[\|\boldsymbol{w}^\star\|^2\|\boldsymbol{x}(k)\|^2\big] + \beta^2\sigma_v^2\mathbb{E}\big[\|\boldsymbol{x}(k)\|^2\big].\end{aligned} \tag{34}$$

In order to simplify (34), the following additional stochastic hypothesis is assumed [7]:

H4. The input signal is white.

From H4 and the fact that the BC-LMS algorithm is asymptotically unbiased (see (10)), it can be demonstrated that the following identity is valid in steady state:

$$\begin{aligned}&\mathbb{E}\big[\|\tilde{\boldsymbol{w}}(k+1)\|^2\big] + \beta^2\sigma_\eta^4\mathbb{E}\big[\|\boldsymbol{w}^\star\|^2\big] \\ &= (1+\beta\sigma_\eta^2)^2\mathbb{E}\big[\|\tilde{\boldsymbol{w}}(k)\|^2\big] - 2\beta\mathbb{E}\big[[\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)]^2\big] \\ &- 2\beta^2\sigma_\eta^2\mathbb{E}\big[[\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)]^2\big] + \beta^2\mathbb{E}\big[\|\boldsymbol{x}(k)\|^2[\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)]^2\big] \\ &+ \beta^2\sigma_\eta^2\mathbb{E}\big[\|\boldsymbol{w}^\star\|^2\|\boldsymbol{x}(k)\|^2\big] + \beta^2\sigma_v^2\mathbb{E}\big[\|\boldsymbol{x}(k)\|^2\big].\end{aligned} \tag{35}$$

By applying H3 in (35), one obtains

$$\mathbb{E}\big[\|\tilde{\boldsymbol{w}}(k))\|^2\big] = \frac{B\mathbb{E}\big[\|\boldsymbol{w}^\star\|^2\big] - \beta\sigma_v^2N(\sigma_u^2+\sigma_\eta^2)}{2\sigma_\eta^2 + \beta\sigma_\eta^4 - 2F(1+\beta\sigma_\eta^2) + \beta F^2}, \tag{36}$$

where

$$B \triangleq \beta\sigma_\eta^2\big[\sigma_\eta^2 - N(\sigma_u^2+\sigma_\eta^2)\big], \tag{37}$$

$$F \triangleq N(\sigma_u^2+\sigma_\eta^2). \tag{38}$$

Note that (36) predicts the asymptotic MSD of the BC-LMS algorithm in a closed form. It should be emphasised that there are two characteristics of the estimation of (36) that are not typical in standard analysis results of adaptive filtering algorithms:

(i) it depends on the energy of the unknown ideal plant ($\mathbb{E}[\|\boldsymbol{w}^\star\|^2]$);
(ii) it has some dependence on the *square* of the variance of $\eta(k)$ (σ_η^4).

Simulations: In all simulations, the coefficients of the ideal transfer function are given by

$$w_i^\star = -10\sin[4.5(i+1)]\exp\left(-\frac{i+1}{100}\right), \text{ for } i \in \{0, 1, \ldots, N-1\}. \tag{39}$$

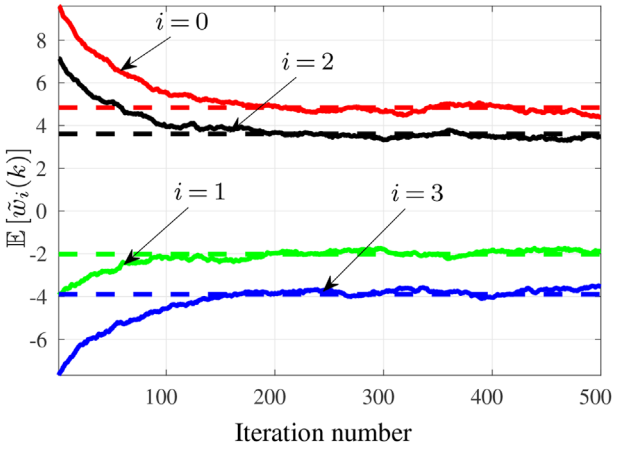


Fig. 2 *Average deviation $\mathbb{E}[\tilde{w}_i(k)]$, for $i \in \{0, 1, 2, 3\}$. The dashed horizontal lines indicate the theoretical asymptotic values derived from (13)*

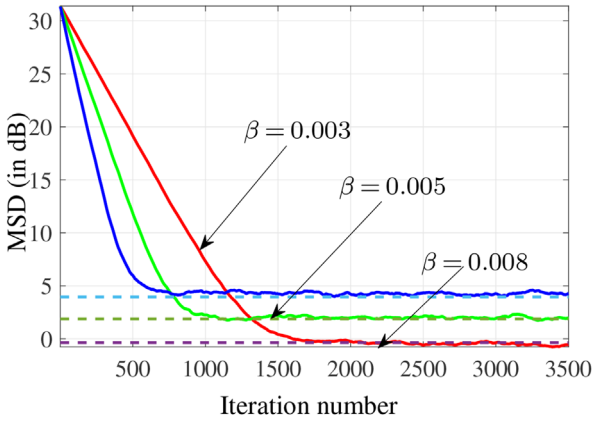


Fig. 3 *MSD evolution as a function of the number of iterations for three different step sizes. The dashed horizontal lines indicate the theoretical asymptotic values obtained from (36)*

In the first simulation, the ideal transfer function has length $N = 20$, both input signal and noise signals are zero-mean white Gaussian processes with variances $\sigma_u^2 = 1$, $\sigma_\eta^2 = 1$ and $\hat{\sigma}_\eta^2 = 10^{-4}$, and the step-size is $\beta = 10^{-2}$. Figure 2 shows the empirical steady-state mean coefficient deviations computed with 100 independent Monte Carlo trials and the corresponding theoretical predictions obtained from (13).

In the second simulation, the noise signals are zero-mean white Gaussian processes having variances $\sigma_\eta^2 = 10^{-2}$ and $\sigma_v^2 = 10^{-4}$, whereas the input $u(k)$ is a coloured signal obtained by passing through the filter $H(z) = 1 - 0.8z^{-1}$ a white Gaussian process with unitary variance. The averaged results are obtained from 500 independent Monte Carlo trials and the ideal transfer function of (39) with $N = 40$. Figure 3 shows the algorithm performance for three different values of β and the theoretical asymptotic values obtained from (36).

In the third simulation, the theoretical and simulated steady-state MSD values are compared for different noise variances σ_v^2 and step size values β. The input and noise signals are zero-mean white Gaussian processes with variances $\sigma_u^2 = 1$ and $\sigma_\eta^2 = 10^{-6}$, respectively. The averaged results are obtained from 100 independent Monte Carlo trials and the ideal transfer function of (39) with $N = 40$. Figure 4 shows very good agreement between the estimates obtained from the closed-form model and the simulated results.

Conclusions: This paper demonstrates that the bias-compensated LMS algorithm implements an optimal estimator, not only in the mean sense, but also when mean square statistics are taken into account. A closed form expression that predicts steady-state performance of the BC-LMS algorithm is also advanced, the energy conservation method. This stochastic model sheds light on the factors that drive the performance of the algorithm. Theoretical predictions were confirmed by simulations.

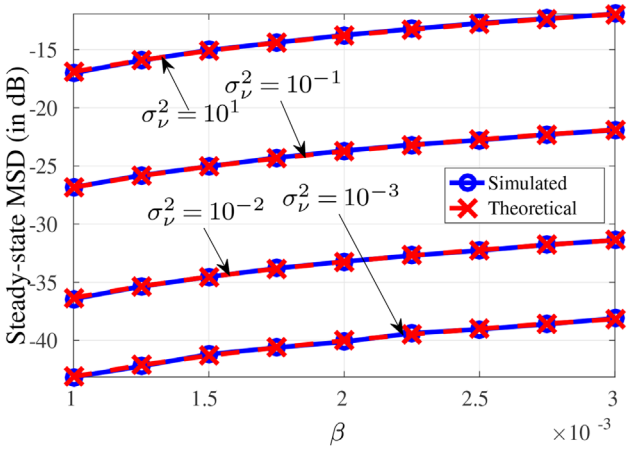


Fig. 4 *Steady-state MSD for different values of σ_v^2*

Acknowledgments: This work has been supported by CNPq, FAPERJ and CAPES, Brazil.

Received: *17 September 2020* Accepted: *14 October 2020*
doi: 10.1049/ell2.12044

References

1 Haykin, S.: *Adaptive filter theory*. Prentice-Hall, Inc., Upper Saddle River, NJ (1996)
2 Kang, B., Yoo, J., Park, P.: Bias-compensated normalised LMS algorithm with noisy input. *Electron. Lett.* **49**(8), 538–539 (2013)
3 Zheng, Z., Liu, Z., Lu, L.: Bias-compensated robust set-membership NLMS algorithm against impulsive noises and noisy inputs. *Electron. Lett.* **53**(16), 1100–1102 (2017)
4 Solo, V.: LMS: Past, Present and Future. *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, Brighton, pp. 7740–7744. IEEE, Piscataway, NJ (2019)
5 Lara, P., et al.: Exact expectation evaluation and design of variable step-size adaptive algorithms. *IEEE Signal Process. Lett.* **26**(1), 74–78 (2019)
6 Yousef, N.R., Sayed, A.H.: A unified approach to the steady-state and tracking analyses of adaptive filters. *IEEE Trans. Signal Process.* **49**(2), 314–324 (2001)
7 Shi, K, Shi, P.: Convergence analysis of sparse LMS algorithms with ℓ_1-norm penalty based on white input signal. *Signal Process.* **90**(12), 3289–3293 (2010)

A BCS microwave imaging algorithm for object detection and shape reconstruction tested with experimental data

Nicolás Zilberstein,[1] Juan Augusto Maya,[2,✉] and Andrés Altieri[2]
[1]*Universidad de Buenos Aires, Facultad de Ingeniería, Buenos Aires, Argentina*
[2]*CSC-CONICET and Universidad de Buenos Aires, Facultad de Ingeniería, Buenos Aires, Argentina*
✉Email: jmaya@fi.uba.ar

An approach based on the Green function and the Born approximation is used for impulsive radio ultra-wideband microwave imaging, in which a permittivity map of the illuminated scenario is estimated us-

ing the scattered fields measured at several positions. Two algorithms are applied to this model and compared: the first one solves the inversion problem using a linear operator. The second one is based on the Bayesian compressive sensing technique, where the sparseness of the contrast function is introduced as a priori knowledge in order to improve the inverse mapping. In order to compare both methods, measurements in real scenarios are taken using an ultra-wideband radar prototype. The results with real measurements illustrate that, for the considered scenarios, the Bayesian compressive sensing imaging algorithm has a better performance in terms of range and cross-range resolution allowing object detection and shape reconstruction, with a reduced computational burden, and fewer space and frequency measurements, as compared to the linear operator.

Introduction: Microwave imaging techniques using ultra-wideband (UWB) signals have been widely studied in recent years due to the various applications in different fields, such as breast cancer detection [1], synthetic aperture radar (SAR) image classification [2], ground penetrating radars systems [3], and material sample characterization [4]. However, inverse microwave imaging problems are intrinsically challenging due to their non-linear nature and ill-posedness [5], and algorithms are continuously developing. Different approaches have been used for building qualitative and quantitative microwave images. In the former, the objective is to detect changes in the permittivity map without trying to obtain the true value of permittivity of the target. For these cases, linear problems can be formulated and solved using several techniques as diffraction summation and F-K migration. Furthermore, linearization methods of the non-linear equations resulting from the scattering theory using the Born approximation and the Green function are well-established first principles models [6].

In the case of quantitative microwave images, it is expected to estimate the true value of the permittivity map and typically, the non-linear problem has to be solved. For example, this is the case of the contrast-source formulation, which is based on an iterative method. Here we are interested in detecting the location and shape of metallic and dielectric objects, which finds several applications [7–9]. Therefore, we follow the first approach to estimate a qualitative microwave image using a model based on the Born approximation (BA).

Additionally, to improve the conditioning of the problem, the use of schemes which take advantage of a priori information helps to increase the robustness of the solutions and the overall performance of the estimators. Also, typical inversion schemes require the acquisition and storage of a large number of measurements over the region of interest. This is costly and involved, so the development of schemes which achieve a good performance with a small set of measurements is of interest in most practical scenarios. In this context, compressive sensing (CS) techniques have been applied to the inverse problem at hand where the signal to be estimated is sparse in some basis. An interesting approach is the Bayesian compressive sensing (BCS) framework [10, 11], which has been applied to inverse problems showing promising results. However, in most works in the literature, its performance is evaluated using synthetic data in 2D scenarios [12, 13].

Here, we employ a BCS algorithm for object detection and shape reconstruction. We evaluate its performance in 3D scenarios with experimental data, where conditions are more adverse due to non-linearity, noise and jitter in the electronic devices, interference from sources outside of the experiment, and non-desired echoes produced by scatterers other than the targets. The physical model that we use is strictly valid for scenarios with cylindrical symmetry. Despite that, in practical scenarios, this assumption is clearly not satisfied, we show that the model allows to obtain good performance tomographic images of the permittivity map. We compare the performance of the BCS algorithm with a reconstruction technique, frequently used in the literature [9, 14], which uses the adjoint operator (AO) that maps the contrast function into the measured scatter fields (under the Born approximation). Different targets are tested in order to show the capacity of the BCS imaging algorithm to detect the presence of the object and estimate its geometric shape.

This work is organized as follows. First we present the model and the AO algorithm. Second, we apply the BCS algorithm to obtain the microwave image. Then, we present the results, and finally, draw the main conclusions.

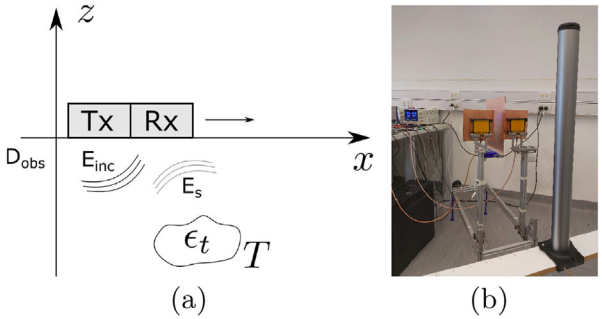


Fig. 1 *(a) Schematic representation of the setup and (b) experimental setup for Scenario 1*

Problem formulation: Let us consider the scenario sketched in Figure 1, where a single-input single-output configuration is used to obtain a tomographic image (over the xz-plane for a fixed y_0) of the observation domain $D_{\text{obs}} = \{\boldsymbol{r} = (x, z) : z < 0\}$. Assume that the transmitting and receiving antennas are located at N_a positions $\boldsymbol{r}_{\text{rn}} = (x_{\text{rn}}, z_0)$ and $\boldsymbol{r}_{\text{tn}} = (x_{\text{tn}}, z_0)$, respectively, where the center position of the Tx–Rx system $x_{\text{cn}} = \frac{x_{\text{tn}}+x_{\text{rn}}}{2}$ is moved in the interval $[x_{\min}, x_{\max}]$ and $z_0 = 0$. We consider the common offset acquisition method, where the distance between the Tx and Rx antennas is fixed. The system has -10 dB bandwidth $B = [f_{\min}, f_{\max}]$, which is divided in M_f bins. We consider scenarios without interface, although it is straightforward to extend the results to scenarios with multiple layers given that we only have to recompute the Green function [15]. The observation domain is illuminated by the incident electric field $\boldsymbol{E}_{\text{inc}}$, generated by the transmitting antenna Tx. The interaction of $\boldsymbol{E}_{\text{inc}}$ with the scatterers produces the scattered electric field $\boldsymbol{E}_{\text{s}}$. The total electric field is $\boldsymbol{E}_{\text{tot}} = \boldsymbol{E}_{\text{inc}} + \boldsymbol{E}_{\text{s}}$. Under the assumption that $\boldsymbol{E}_{\text{inc}}$ has a linear polarization in the y component, that is, $\boldsymbol{E}_{\text{inc}} = E_{\text{inc}}\,\hat{\boldsymbol{y}}$, and that the observation domain has cylindrical symmetry (with an arbitrary section) also through the y component, the Green function, which is the impulse response of the system, is a scalar function (instead of a tensor). In fact, this is the case if we consider as an excitation a 3D line source directed in the y direction. Under these hypotheses and using the BA for *weak* scatterers [5], where the total electric field is approximated by the incident field, the scattered electric field measured at $\boldsymbol{r}_{\text{rn}}$ for a fixed wave number k can be written as

$$E_{\text{s}}(\boldsymbol{r}_{\text{rn}}, k) = k^2 \int_{D_{\text{obs}}} \Delta\epsilon_t(\boldsymbol{r}) E_{\text{inc}}(\boldsymbol{r}, \boldsymbol{r}_{\text{tn}}, k) g(\boldsymbol{r}_{\text{rn}}, \boldsymbol{r}, k) d\boldsymbol{r}, \tag{1}$$

where $\boldsymbol{r}$ represents the different positions of the scenario under investigation, $E_{\text{inc}}(\boldsymbol{r}, \boldsymbol{r}_{\text{tn}}, k) = \imath\eta_0 k P(k) g(\boldsymbol{r}, \boldsymbol{r}_{\text{tn}}, k)$ is the incident field produced by the source at position $\boldsymbol{r}_{\text{tn}}$ and measured at $\boldsymbol{r}$, η_0 is the intrinsic impedance of free space ($\approx 377\Omega$), and $P(k)$ is the Fourier transform of the pulse transmitted by the Tx. The wave number k is related to the frequency f in Hertz through the equation $k = 2\pi f/c$, where c is the speed of light in free space. For the homogeneous case under consideration (without considering the target), $g(\boldsymbol{r}_{\text{rn}}, \boldsymbol{r}, k) = \frac{\imath}{4} H_0^{(2)}(k\|\boldsymbol{r} - \boldsymbol{r}_{\text{rn}}\|)$ is the scalar Green function, where $H_0^{(2)}$ is the Hankel function of second kind and order 0[1]. Finally, $\Delta\epsilon_t = \epsilon_{\text{rt}}(\boldsymbol{r}) - \epsilon_{\text{rb}}$ is the contrast function, where in this case the background relative permittivity is $\epsilon_{\text{rb}} = 1$ and ϵ_{rt} is the relative permittivity of the target. Note that $\Delta\epsilon_t$ is non-zero only in the target position.

Equation (1) can be expressed by a linear operator applied to the contrast function. Then the microwave image can be obtained using its adjoint operator applied to the measured scattered electric field [14, 15] as:

$$\hat{\Delta}\epsilon_r(\boldsymbol{r}) = -\imath\eta_0 \int_{k_{\min}}^{k_{\max}} dk P(k) k^3 \int_{x_{\min}}^{x_{\max}} dx_{\text{cn}}\, g^*(\boldsymbol{r}_{\text{rn}}, \boldsymbol{r}, k) \times g^*(\boldsymbol{r}_{\text{rn}}, \boldsymbol{r}, k) E_s(\boldsymbol{r}_{\text{rn}}, k), \tag{2}$$

where the limits of the outer integral are $k_{\min/\max} = \frac{2\pi}{c} f_{\min/\max}$, and * means complex conjugate.

[1]The time harmonic convention $\exp(\imath 2\pi f t)$ is used here.

Discrete model for BCS: In order to obtain a solution for (1), the domain is discretized using a basis of indicator functions, that is,

$$\theta^n(\mathbf{r}) = \begin{cases} 1 & \text{si } \mathbf{r} \in D^n, \\ 0 & \text{si } \mathbf{r} \notin D^n, \end{cases} \quad n = 1, \ldots, N \tag{3}$$

where D_{obs} is divided into N pixels and D^n is the nth pixel. Then, the contrast function is approximated by a piecewise 2D constant function

$$\Delta\epsilon_r(\mathbf{r}) \simeq \sum_{n=1}^{N} w_n \, \theta^n(\mathbf{r}), \tag{4}$$

and we obtain a matrix formulation of the problem,

$$\mathbf{E_s} = \mathbb{G}\mathbf{w},$$

where the elements of $\mathbf{w} \in \mathbb{C}^N$, $\mathbb{G} \in \mathbb{C}^{M_f N_a \times N}$, and $\mathbf{E_s} \in \mathbb{C}^{M_f N_a}$ are, respectively, $(\mathbf{w})_j = w_j$, $(\mathbf{E_s})_l = E_s(\boldsymbol{r}_{rj}, k_q)$, and

$$(\mathbb{G})_{l,n} = \imath\eta_0 k_q^3 P(k_q) \int_{D^n} g(\mathbf{r}, \mathbf{r}_{tj}, k_q) g(\mathbf{r}_{rj}, \mathbf{r}, k_q) \theta^n(\mathbf{r}) \mathrm{d}\mathbf{r}, \tag{5}$$

with $l = q + (j-1)M_f$, $1 \le j \le N_a$, $1 \le q \le M_f$, $1 \le n \le N$. This means that for the position of each Tx–Rx pair there are M_f wave number (frequency) samples, where k_q is the wave number at the center of the qth bin.

Bayesian compressive sensing: We consider the case where the measurements of the electric field are noisy, and the noise is zero-mean white normally distributed with variance σ^2, that is,

$$\mathbf{E_s} = \mathbb{G}\mathbf{w} + \mathbf{N}, \quad \mathbf{N} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}), \tag{6}$$

where $\mathbf{I}$ is the identity matrix of size $M_f N_a \times M_f N_a$. This equation defines a linear problem for estimating $\mathbf{w}$ from $\mathbf{E_s}$. In scenarios where the size of the target is much smaller than the observation domain[2], the contrast function defined in the previous section is a *sparse* signal, meaning that only a few components of $\mathbf{w}$ are different from zero. Then, a compressive sensing approach can be used. In particular, we used the BCS approach, which is a framework where a compressive sampling approach is used. In other words, the measurements are obtained as a linear combination of the projection of the signal $\mathbb{G}\mathbf{w}$ onto a random basis. We define the sensing matrix $\mathbf{\Psi} \in \{0, 1\}^{N_s \times M_f N_a}$ as a selection matrix,that is, a matrix with orthogonal rows and a unique non-zero component in each row with value 1, which selects N_s components of $\mathbf{E_s}$:

$$\mathbf{t} = \mathbf{\Psi}\mathbf{E_s} = \mathbf{\Phi}\mathbf{w} + \mathbf{N}, \ \mathbf{\Phi} = \mathbf{\Psi}\mathbb{G}. \tag{7}$$

An advantage of using a probabilistic framework is that it is not required to satisfy the restricted isometry property (RIP) on the $\mathbf{\Phi}$ matrix. The image problem using BCS is formulated as follows: given the measurements $\mathbf{t}$, obtain

$$\hat{\mathbf{w}} = \arg\max_{\mathbf{w}} p(\mathbf{w}|\mathbf{t}) \tag{8}$$

where $p(\mathbf{w}|\mathbf{t})$ is the a posteriori density function. Following [16], a closed-form solution to this problem can be obtained if $p(\mathbf{w}|\mathbf{t}) = \int p(\mathbf{w}|\mathbf{t}, \boldsymbol{\alpha}, \sigma^2) p(\boldsymbol{\alpha}, \sigma^2|\mathbf{t}) d\boldsymbol{\alpha} d\sigma^2 \approx p(\mathbf{w}|\mathbf{t}, \boldsymbol{\alpha}_{\text{MP}}, \sigma^2_{\text{MP}}) = \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$, where $\boldsymbol{\mu} = \sigma^{-2}\mathbf{\Sigma}\mathbf{\Phi}^T\mathbf{t}$, $\mathbf{\Sigma} = (\sigma^{-2}\mathbf{\Phi}^T\mathbf{\Phi} + \mathbf{A})^{-1}$, and $\mathbf{A} = \text{diag}(\alpha_1, \ldots, \alpha_N)$ is a diagonal matrix. Thus, problem (8) has solution $\hat{\mathbf{w}} = \boldsymbol{\mu}$, and the so-called hyper-parameters $\boldsymbol{\alpha}_{\text{MP}}$ and σ^2_{MP} are computed as in [16], and the image is constructed using (4).

Results: In order to evaluate the advantages of the BCS algorithm compared to the direct inversion technique in Equation (2), we measured different targets in air: a metallic cylinder and a tank with water. We use an impulsive UWB radar to acquire the time-domain measurements of the electric field at N_a positions in the x direction. The antennas are only moved in a linear motion in front of the target to mimic scenarios in which the target is embedded in a material which is only accessible from one side.

The scattered electric field is needed as an input to the algorithms (Equations 2 and 7). In order to obtain it, the incident electric field E_{inc} is measured without the target and then it is subtracted from each measurement. As we need the signal in the frequency domain, we use the FFT to transform each time trace.

The size of the pixels D^n is 2cm × 1cm. This value was selected as a trade-off between computation time of the algorithm and good resolution of the reconstructed targets. The observation domain for the first experiment is $x_{\max} - x_{\min} = 2.2$ m, with measurement steps of 5 cm for the position of Tx and Rx resulting $N_a = 41$. For the others scenarios, the length of the domain is $x_{\max} - x_{\min} = 1.1$ m, with the same step and $N_a = 21$.

A custom test-bed hardware was built using off-the-shelf components. A field programmable gate array (FPGA) is used to generate a train of narrow pulses which is then upconverted to the carrier frequency $f_c = 1.9$ GHz, amplified, and then transmitted. The receiver architecture is composed of a low-noise amplifier, a bandpass filter and a single-stage zero-frequency I/Q demodulator. After the conversion to baseband, the I and Q signals are amplified via a high linearity differential driver amplifier which feeds a high speed dual channel analog-to-digital converter (ADC), operating with a sampling frequency $f_s = 1.8$ GS/s. A real-time sampling strategy was used for this prototype, mainly for reasons of availability, but an equivalent-time sampling strategy could also be used to achieve a substantial reduction in hardware costs. The ADC is controlled by the same FPGA which is used to generate the pulses, which allows for an easy synchronization between the transmitted pulses and the receiver. The FPGA will generate a pulse train composed of $N_p = 91$ pulses, with a pulse period of $T_s = 100$ns. At the receiver, the scattered pulses will be coherently combined to obtain an increase in the signal to interference and noise ratio (SINR) of the received signal.

The antennas are printed elliptical monopole antennas [17], which have the advantage of being low cost, simple to design, compact, and planar. One of their disadvantages is that their radiation pattern is not very directional. In order to improve their directionality, a square reflector of side 160 mm (the wavelength at the central frequency of 1.9 GHz) was placed behind each antenna at a distance of 39 mm (a quarter wavelength at the antenna central frequency). The two antennas were placed side-by-side, and, in order to reduce their mutual coupling, a conductive reflector was placed between the two. The horizontal separation between the antennas was chosen to be 200 mm. A single antenna with its reflector has an azimuth 3 dB beam width of 80°, and an elevation beam width of approximately 110°. The antennas are linearly polarized in the vertical direction and their boresight gain is approximately 10 dB. The antenna setup can be seen in Figure 1a.

The resolution of the image is limited by the hardware (bandwidth), the antenna characteristics (central frequency, fundamentally), and the setup configuration, and can be analyzed through the point spread function (PSF). The resolutions in the x and z directions are $\delta x \approx \frac{c}{f_c} = 15.8$ cm and $\delta z = \frac{c}{B} = 30$cm, respectively.

Scenario 1: Metallic cylinder. The first experiment was done using a metallic cylinder with a diameter of 6 cm and a height of 70 cm as a target. Its position in the (x, z) plane (using the coordinate system of Figure 1a) was $\boldsymbol{r}_t = (1.1 \text{ m}, -42.7 \text{ cm})$. The experimental setup is shown in Figure 1b. For the BCS algorithm, a random sampling both in space and frequency was used: 10 pairs of Tx/Rx positions and 10 frequency bins were selected from a total number of 41 positions and 100 frequency bins. So, for BCS only a 25% of spatial measurements were used and 10% of the frequencies. The results obtained using both AO and BCS algorithms are shown in Figures 2a and 2b, respectively. In both cases, the black circle represents the true target. The scale is normalized with respect to the absolute maximum. The position and size of the cylinder are successfully estimated using both algorithms. Although the weak scattering hypothesis in the BA fails due to a high contrast target, the model is still valid for estimating its geometry. This can be explained by noting that the support of the contrast source that generates the scattered electric field is the same for both the incident electric field and the total electric field. Note also that the BCS algorithm achieves much better resolution than the OA algorithm, especially in the range (z) direction,

[2]In other scenarios one typically can find an adequate basis to explain $\mathbf{w}$ through few coefficients.

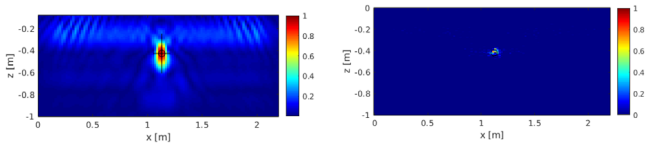


Fig. 2 *Scenario 1: Metallic cylinder. (a) AO and (b) BCS*

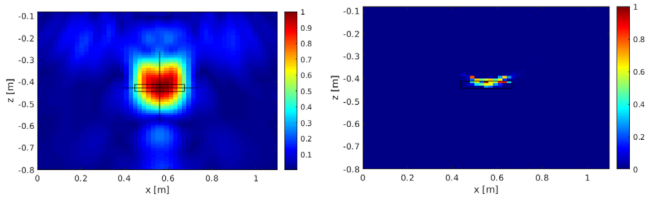


Fig. 3 *Scenario 2: Water tank. (a) AO and (b) BCS*

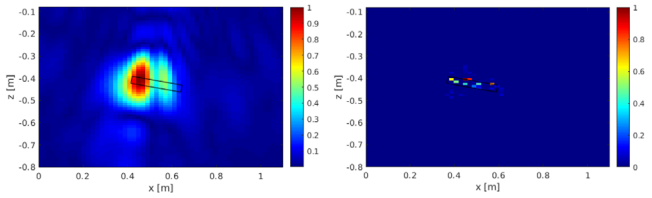


Fig. 4 *Scenario 3: Rotated water tank. (a) AO and (b) BCS*

improving the resolution predicted by the PSF analysis. Additionally, this experiment allows us to validate the use of the 3D model with cylindrical symmetry in this scenario, in spite of the extension of the metallic cylinder finite and the transmitting antenna is not a line source but a linear polarized finite-size directional antenna, with the target centered respect to it in the y direction.

Scenario 2: Water tank. In this case, the target was a plastic tank filled with water ($\epsilon_{rt} \approx 80$), with the shape of a rectangular prism; its width, height and depth are 22.8 cm × 32.5 cm × 3.2 cm, respectively. The position was $\boldsymbol{r} = (0.55\,\text{m}, -42.7\,\text{m})$. In Figure 3a, the reconstruction with the AO algorithm is shown while Figure 3b corresponds to the BCS algorithm. For BCS, 10 Tx/Rx (50%) were used and the number of frequencies bins is 10 (10% of the total).

In this case, the target has a rectangular shape whose perimeter is shown in the figure with a black line. Similar to the previous experiment, the advantage that BCS presents with respect to the AO is the increased spatial resolution, allowing in this case to estimate the frontal shape of the target with more precision.

Scenario 3: Rotated water tank. In this scenario the same tank of Scenario 2 was used, but rotated approximately 10° with respect to the antennas, as shown in Figure 4a. The position was $\boldsymbol{r} = (0.55\,\text{m}, -42.7\,\text{cm})$. For the BCS algorithm, 15 random positions for Tx/Rx and 50 frequencies bins were taken.

From Figure 4a it is clear that the AO algorithm detects the position but does not manage to estimate accurately neither the rotation nor the shape. On the other hand, the performance of the BCS algorithm is better, as it estimates well the position, allows to see the rotation of the target and permits to reconstruct, at least partially, the front of its shape (see Figure 4b).

Conclusions: Here we evaluated two microwave imaging algorithms with experimental data obtained with a impulsive UWB radar test bed in different scenarios. From the experiments we draw the following conclusions: (i) For point targets, like Scenario 1, or moderately complex scenarios such as the rotated water tank of Scenario 3, both algorithms detect the position of the targets but BCS is able to detect the shape of the target better. (ii) The number of measurements when using BCS can be substantially reduced, but not with the adjoint operator. This allows to reduce the complexity of measurement system, the time taken to perform the measurements, and the processing time of the algorithm. (iii) The spatial resolution is improved with the BCS algorithm given that better discrimination is obtained. (iv) The 2D physical model is a good model for a practical 3D scenario when a tomographic microwave image is required. Although in real scenarios, as examined here, the hypothesis of cylindrical symmetry of the target does not hold, the algorithms perform well as they can detect the position and shape of the targets.

Acknowledgements: This work was partially supported by the projects UBACyT 20020170200283BA and PICT-2017-4533.

Received: *14 September 2020* Accepted: *5 November 2020*
doi: 10.1049/ell2.12059

References

1 Naghsh, N.Z., Ghorbani, A., Amindavar, H.: Compressive sensing for microwave breast cancer imaging. *IET Sig. Proc.* **12**(2), 242–246 (2017)
2 Xia, S., Yu, Z., Yu, J.D.: Adaptive convolutional network for SAR image classification. *J. Eng.* **2019**(20), 6868–6872 (2019)
3 Xu, X., et al.: Integrated synchronous data acquisition subsystem for high-speed GPR system. *IET Circuits Devices Syst.* **13**(7), 1049–1055 (2019)
4 Mextorf, H., et al.: New UWB free-space method for the classification and characterization of dielectric objects. In: *2011 IEEE Int. Conf. on Ultra-Wideband (ICUWB)*, pp. 380–384 (2011)
5 Chew, W.: *Waves and Fields in Inhomogeneous Media*. IEEE Press, New York (1995)
6 Leone, G., Soldovieri, F.: Analysis of the distorted born approximation for subsurface reconstruction: truncation and uncertainties effects. *IEEE Trans. Geos. Rem. Sens.* **41**(1), 66–74 (2003)
7 Tantong, S., et al.: Near-field microwave imaging techniques for object detection and shape reconstruction. In: *2007 IEEE Region 5 Technical Conference*, pp. 299–302 (2007)
8 Irishina, N., et al.: Detecting and imaging dielectric objects from real data: A shape-based approach. *Math. Comput. Modell.* **50**(5), 743–749 (2009)
9 Dorn, O., Miller, E.L., Rappaport, C.M.: A shape reconstruction method for electromagnetic tomography using adjoint fields and level sets. *Inverse Prob.* **16**(5), 1119 (2000)
10 Ji, S., Xue, Y., Carin, L.: Bayesian compressive sensing. *IEEE Trans. Sig. Proc.* **56**(6), 2346–2356 (2008)
11 Tipping, M.: Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* **1**, 211–244 (2001)
12 Fouda, A.E., Teixeira, F.L.: Ultra-wideband microwave imaging of breast cancer tumors via Bayesian inverse scattering. *J. Appl. Phys.* **115**(6), 064701 (2014)
13 Poli, L., et al.: Microwave imaging within the first-order born approximation by means of the contrast-field Bayesian compressive sensing. *IEEE Trans. Ant. Prop.* **60**(6), 2865–2879 (2012)
14 Liu, Y.L., Li, L.L., Li, F.: Imaging of two-dimensional targets buried in a lossy earth with unknown characteristics from multi-frequency and multi-monostatic data. *IET Microwaves Antennas Propag.* **4**(10), 1647–1653 (2010)
15 Zhang, W.: Two-dimensional microwave tomographic algorithm for radar imaging through multi layered media. *Progress Electr. Res.* **144**, 261–270 (2014)
16 Tipping, M.E., Faul, A.C.: Fast marginal likelihood maximisation for sparse Bayesian models. AISTATS (2003)
17 Liang, J., et al.: Study of a printed circular disc monopole antenna for UWB systems. *IEEE Trans. Ant. Prop.* **53**(11), 3500–3504 (2005)

PAPR analysis of non-contiguous duplex multicarrier signals

V. Savaux[1,✉] and Y. Louët[2]
[1] *Advanced Connectivity Lab, b-com, Rennes, France*
[2] *CentraleSupélec, Rennes, France*
✉E-mail: vincent.savaux@b-com.com

This paper deals with the peak to average power ratio of duplex multicarrier signals. A simple but accurate expression of the complementary cumulative distribution of the peak to average power ratio is suggested, based on both empirical and theoretical analyses of the mean peak to average power ratio of such a continuous duplex signal. To this end, it is shown that the mean peak to average power ratio can be written in function of the frequency distance between the two signals composing the duplex one. The relevance and the accuracy of the suggested analysis are shown through simulations, and a discussion paves the way to further analyses involving general multiplex signals with different powers and subcarriers numbers.

Introduction: Multicarrier signals (MC) are well known to be prone to high power fluctuations due to the inherent summation of independent information carried on different tones. These power fluctuations can be characterised through the ratio of the maximum instantaneous power and the mean power of the transmitted signal, which leads to the definition of the peak to average power ratio (PAPR). PAPR is a metric that is directly related to the instantaneous power fluctuations of a signal. When considering a power amplifier, the power budget is the ratio of the output power of the amplified signal and the supply power. As the output power is a function of the input power, the efficiency of the power amplifier is generally inversely proportional to the PAPR of the input signal. This is why PAPR reduction methods aim to mitigate the PAPR so as to drive the signal as close as possible to the saturation point of the power amplifier where the efficiency is the highest [1].

Although PAPR derivation is well known considering a single-band multicarrier signal [2–4], very few works have been suggested for non-contiguous multiplex of signals [5–7]. Nevertheless, this situation rises in contexts of heterogeneous signals transmission such as cognitive radio or 5G, when several non-contiguous signals, each associated to different users, have to be transmitted and amplified all together. In this context, the resulting PAPR should not only depend on the signals themselves (respective number of subcarriers, respective mean powers, number of users) but on the frequency distance between the signals as well. Given these features, any opportunistic spectrum access or frequency resource allocation has to be evaluated on the PAPR criteria.

In [5, 6], the PAPR is analysed only through simulations [5], and through upper bounds of the complementary cumulative distribution functions (CCDFs) [6]. More recently, the authors in [7] derived a thorough study of the PAPR of multiplex signals with different numerology for an application to 5G technology. Although the theoretical expression of the PAPR generally well matches the simulations results, it is not easily tractable in practice. Furthermore, the authors in [7] did not specifically analyse the effect of the frequency gap between signals on the PAPR behaviour.

In this paper, we suggest handling this PAPR issue by considering the combination of two non-contiguous orthogonal frequency division multiplexing (OFDM) signals with the same number of carriers and power. A simple and tractable expression of the CCDF of PAPR is proposed, based on both theoretical and empirical original analyses of the mean PAPR of duplex multicarrier signals. The resulting PAPR is evaluated with the frequency distance between the two spectra, and we show the relevance of the derived results through simulations. In particular, we show that the PAPR of a duplex signal increases when the frequency distance between signal increases, but it reaches an upper bound. Moreover, we discuss the results to provide leads for further studies including more general multiplex of multicarrier signals as in [7]. This PAPR analysis can be useful to any engineer or researcher seeking to rapidly evaluate

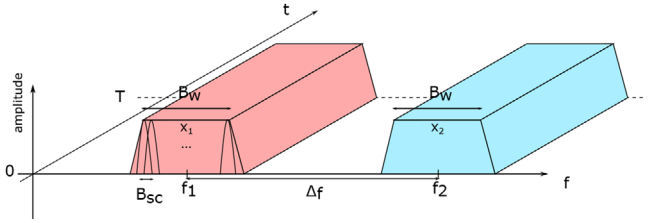


Fig. 1 *Time-frequency representation of duplex of multicarrier signals x_1 and x_2 transmitted at frequencies f_1 and f_2*

the PAPR of duplex multicarrier signals without the need of sophisticated functions.

The remainder of the paper is organised as follows. The PAPR of the duplex signal is derived in Section 3 after the statement of the problem given in Section 2. Simulations are provided in Section 4 and a conclusion ends the paper in Section 6.

Problem statement: We consider a duplex multicarrier signal $x(t)$ composed of two signals $x_1(t)$ and $x_2(t)$ as

$$x(t) = x_1(t) + x_2(t), \tag{1}$$

where t belongs to an observation window of duration T, that is, $t \in [0, T]$. Typically, T corresponds to the duration of an OFDM symbol. Moreover, we assume that x_1 and x_2 are non-contiguous in the frequency domain, that is, x_1 is transmitted at frequency f_1 and x_2 at frequency f_2 such that we define the frequency distance $\Delta_f = f_2 - f_1$. Without loss of generality, we arbitrarily consider that $f_2 \geq f_1$. Furthermore, we assume that x_1 and x_2 are composed of the same number of subcarriers denoted by N with a common subcarrier spacing B_{sc}. The total bandwidth of each signal $x_1(t)$ or $x_2(t)$ is then $B_w = NB_{sc}$. The time-frequency representation of x_1 and x_2 is illustrated in Figure 1.

The sampled version of $x_i(t)$, $i \in \{1, 2\}$, at Nyquist rate t_s, is noted $x_{i,n}$ with $n = 0, 1, \ldots, N-1$. In this case, the observation window can be rewritten as $T = Nt_s$. If N is larger than about 10 subcarriers, it can be reasonably assumed that the samples $x_{i,n}$ are independent and identically distributed and $x_{i,n}$ obey a complex Gaussian distribution, that is, $x_{i,n} \sim \mathbb{C}\mathcal{N}(0, \sigma^2)$, where $\sigma^2 = \mathbb{E}\{|x_{i,n}|^2\}$ [8], with $\mathbb{E}\{.\}$ the mathematical expectation. Under this assumption, Van Nee *et al*. suggested an expression of the CCDF of the PAPR of $x_i(t)$, $i = 1, 2$, expressed as [2–4]:

$$CCDF(\lambda) = \mathbb{P}\left(\max_{t\in[\![0,T]\!]} \frac{|x_i(t)|^2}{\sigma^2} \geq \lambda\right) \tag{2}$$

$$\approx 1 - (1 - e^{-\lambda})^{\alpha N}, \tag{3}$$

where $\mathbb{P}(E)$ means 'probability of the event E', and α is an adjusting parameter that has been empirically set to $\alpha = 2.8$ in [2]. The popularity of the CCDF expression in (3) lies in its simplicity and its good accuracy. In fact, the PAPR of a continuous multicarrier signal can be properly approximated by knowing its subcarriers number only. Therefore, it is relevant to suggest a simple expression of the CCDF of the PAPR of the duplex multicarrier signal $x(t)$ in a Van Nee's fashion. To this end, we empirically set different α values in function of Δ_f due to the mean PAPR of the duplex multicarrier signal. The developments are hereby derived. Nevertheless, limitations of Van Nee's model have been stated in [4, 9], such as the lack of theoretical justification. We hereby use this approach for its simplicity, but we will, in turn, fairly discuss the suggested adaptation of (3) to duplex signals in Section 5.

PAPR analysis: We assume that the CCDF of the PAPR of a multicarrier duplex signal $x(t)$ in (1) can be expressed as in (3), where the coefficient α needs to be defined. This assumption will be verified through simulations. To do so, on one hand we theoretically derive the mean of the PAPR as a function of α (it must be emphasised that the mean of PAPR has no practical significance, but in this paper, it is used as a trick of arithmetic to derive α). On the other hand, we empirically express the

mean of PAPR as a function of the frequency distance Δ_f. From these two developments, we obtain a simple expression of the CCDF of PAPR with a coefficient α as a function of Δ_f.

Proposition 1. *We define μ as the mean PAPR of the signal $x(t)$ in (1). For large N value ($N \geq 128$), the mean PAPR can be simply expressed as*

$$\mu = \gamma + \ln(\alpha N), \tag{4}$$

where γ is the Euler–Mascheroni constant, $\gamma \approx 0.577$, and $\ln(.)$ is the natural logarithm.

Proof. In order to develop the mean of PAPR, we rewrite the CCDF in (3) by using the binomial series expansion of $(1 - e^{-\lambda})^{\alpha N}$ as

$$CCDF(\lambda) = 1 - \sum_{k=0}^{+\infty} \binom{\alpha N}{k} (-1)^k \mathrm{e}^{-k\lambda} \tag{5}$$

$$= - \sum_{k=1}^{+\infty} \binom{\alpha N}{k} (-1)^k \mathrm{e}^{-k\lambda}, \tag{6}$$

where $\binom{\alpha N}{k}$ is the generalised binomial coefficient. The distribution of the PAPR, defined as $f(\lambda) = \frac{\partial(1-CCDF(\lambda))}{\partial\lambda}$, is then obtained as

$$f(\lambda) = - \sum_{k=1}^{+\infty} \binom{\alpha N}{k} (-1)^k k \mathrm{e}^{-k\lambda}, \tag{7}$$

hence the mean is derived as

$$\mu = \int_0^{+\infty} \lambda f(\lambda) d\lambda = - \sum_{k=1}^{+\infty} \frac{\binom{\alpha N}{k}(-1)^k}{k}. \tag{8}$$

For large αN value, we have $\binom{\alpha N}{k} \sim \frac{(\alpha N)^k}{k!}$, then (8) yields

$$\mu = - \sum_{k=1}^{+\infty} \frac{(\alpha N)^k (-1)^k}{k k!}. \tag{9}$$

From [10], we know that the series expansion of the exponential integral defined for any $z \in \mathbb{C}\backslash\mathbb{R}^-$ as $E_1(z) = \int_z^{+\infty} \frac{e^{-t}}{t} dt$ leads to

$$E_1(z) = -\gamma - \ln(z) - \sum_{k=1}^{+\infty} \frac{(-1)^k z^k}{k k!}. \tag{10}$$

Since $\lim_{z \to +\infty} E_1(z) = 0$, and reminding that we assume large αN, then the substitution of $z = \alpha N$ into (10) leads to (4), which concludes the proof. □

It can be noticed that this result is very similar to those in [11, 12] where Nyquist-sampled signals are considered. Thus, μ in (4) can be seen as a generalised expression of the mean PAPR. From (4), we note that the α coefficient can in turn be expressed in function of the mean μ as

$$\alpha = \frac{\exp(\mu - \gamma)}{N}. \tag{11}$$

In order to obtain α relatively to the frequency distance Δ_f, we empirically express μ as a function g of Δ_f such as suggested in Proposition 2. For convenience, and in order to make the expression g independent of the bandwidth B_w, we set $\theta = \frac{\Delta_f}{B_w}$.

Proposition 2. *Consider a non-contiguous duplex multicarrier signal $x(t)$ such as defined in (1). Let $\theta \in \mathbb{R}_+$, then the mean PAPR μ of $x(t)$ can be expressed as a function of θ as*

$$\mu = g(\theta) \tag{12}$$

$$= \beta - \nu \exp(-q.\theta^c), \tag{13}$$

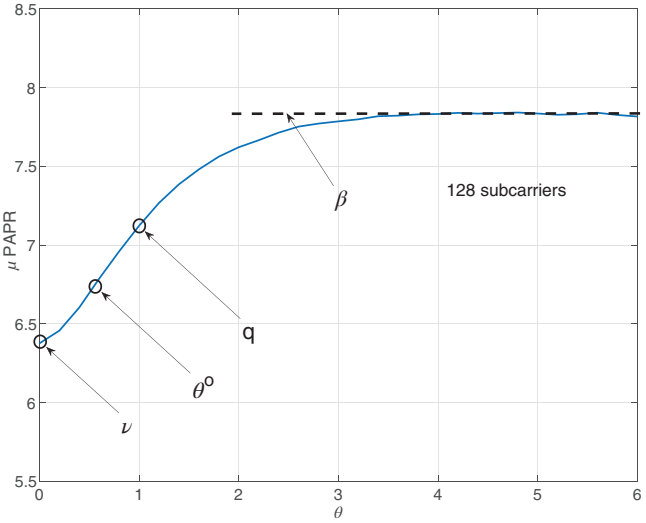


Fig. 2 *Mean PAPR μ as a function of θ, for 128 subcarriers per signal. Markers indicate the values from which the parameters β, ν, q, and θ^o (leading to c) are deduced*

where β, ν, q, and c are positive parameters that need to be determined. It can be noted that μ in (13) is a strictly increasing function of θ, which is upper bounded by β.

The four parameters can be assessed as follows:

- $\beta = \lim_{\theta \to +\infty} \mu$, even though in practice θ should feature low values due to the limitations of the transmitter in term of bandwidth.
- Once β is obtained, we deduce ν by means of the equality $g(0) = \beta - \nu$, which corresponds to the mean PAPR of a continuous (at least largely oversampled) signal featuring N subcarriers. In practice, this case should not happen since it means that x_1 and x_2 overlap since $f_1 = f_2$, but for the sake of the analysis, it allows us to deduce ν.
- Similarly, q can be deduced due to $g(1) = \beta - \nu e^{-q}$, which corresponds to the mean PAPR of a continuous signal featuring $2N$ contiguous subcarriers.
- Let us note $\theta^o = \sup_{\theta \in \mathbb{R}_+} \frac{\partial g(\theta)}{\partial \theta}$, then c can be deduced by solving $\frac{\partial^2 g(\theta^o)}{\partial \theta^2} = 0$. In fact, a unique solution θ^o exists, leading to a unique solution to $\frac{\partial^2 g(\theta^o)}{\partial \theta^2} = 0$ as well. Alternatively, the parameter c can be empirically set to fit the behaviour of μ in function of θ obtained through simulations.

To summarise our approach, we suggest estimating the CCDF of PAPR of duplex multicarrier signal $x(t)$ in (1) by using a simple Van Nee-like expression as in (3) where the coefficient α must be defined. To this end, we theoretically express α as a function of the mean PAPR μ in (11) on one hand. On the other hand, we empirically derive an expression of μ in (13) as a function of the frequency distance Δ_f (or equivalently θ) between the two multicarrier signals that compose the duplex. Then, substituting the empirical mean μ into (11) finally leads to the simple expression of CCDF of PAPR (3).

Simulations: This section aims to evaluate the relevance of the previous developments through simulations. Both signals x_1 and x_2 are composed of N subcarriers with $N \in \{128, 256, 512\}$, carrying elements that are randomly taken from a 16-QAM constellation. We assume that x_1 and x_2 are synchronised in time, even though time synchronisation should not have influence on the PAPR value, according to 'remark 2' in [7]. A large oversampling rate of $R = 32$ is considered in order to approximate continuous signals, and at least to respect the condition $\Delta_f << \frac{R}{t_s}$ for reasonable Δ_f range. When this condition does not hold, some simulations artefacts appears when $\Delta_f > \frac{R}{4t_s}$ (i.e. when Δ_f is too close to the sampling frequency), which are not further described in this paper. All the simulations results have been averaged over 10 000 independent runs.

Figure 2 shows the mean PAPR μ (linear scale) versus θ for a duplex signal $x(t)$ featuring 128 subcarriers per signal $x_i(t)$, $i = 1, 2$. The

Table 1. Parameters β, ν, q, and c obtained from simulations

Parameter	N	Value
β	{128, 256, 512}	{7.83, 8.6, 9.36}
ν	{128, 256, 512}	{1.45, 1.49, 1.5}
q	{128, 256, 512}	0.7
c	{128, 256, 512}	1.5

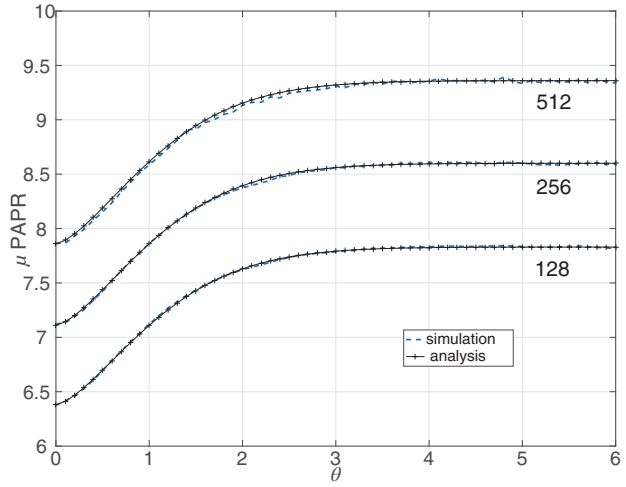


Fig. 3 *Mean PAPR μ as a function of θ, for 128, 256, and 512 subcarriers per signal. Comparison of simulations and analysis (13)*

behaviour of μ has been obtained by simulation, and allows us to deduce the four parameters β, ν, q, and θ^o (leading to c), such as highlighted by the different markers. The corresponding values have been reported in Table 1, for $N = 128$, 256, and 512. We observe from Figure 2 that μ is strictly increasing with θ, and reaches an upper bound for $\theta \geq 4$.

It is worth noticing from Table 1, that β and ν depend on θ, whereas q and c are constant. Furthermore, ν tends to 1.5 when N increases. We deduce that, for N large (typically $N \geq 512$) it is sufficient to assess β for $\theta \geq 4$ to estimate μ for any θ, which further simplifies the PAPR estimation. Otherwise, further simulations have shown that both β and ν must be set when $N < 512$.

In Figure 3, we compare the mean PAPR μ versus θ for 128, 256, and 512 subcarriers per signal, obtained through simulations and from (13) using the parameters in Table 1. It can be observed that the empirical μ values in function of θ perfectly match the trajectories obtained through simulations, therefore validating (13). For applications, we deduce from results in Figure 3 that it is preferable to transmit the signals x_1 and x_2 in contiguous bands (i.e. $\theta = 1$), when possible, as it minimises the PAPR. However, in practice, other parameters should be considered to properly design the system, such as the synchronisation between signals (synchronisation mismatch between x_1 and x_2 would lead to interference), or the possible intermodulation distortions that may cause disruptions in both x_1 and x_2, but this is not dealt with in this paper.

Figure 4 compares the trajectories of $CCDF(\lambda)$ versus λ (dB) obtained through simulations with that obtained with (3). Note that $N = 256$ subcarriers are considered, and $\theta = 1, 3$. The corresponding α values are 5.68 and 11.45, respectively. It can be seen that the curve obtained from analysis for $\theta = 1$ matches that obtained through simulations. Moreover, it should be noticed that $\alpha/2 = 2.82$, which almost corresponds to the empirical value α =2.8 defined in [2] by Van Nee. Note that we find $\alpha/2 = 2.7$ and $\alpha = 3.02$ for $N = 128$, and $N = 512$, respectively. We then deduce that the suggested approach allows to adapt α value according to N in Van Nee model (i.e. $\theta = 1$). It can be seen for $\theta = 3$ that the results obtained from analysis slightly underestimate the CCDF of PAPR obtained through simulations (of about 0.05 dB). The suggested analysis is nevertheless an accurate approximation of the CCDF of PAPR regarding the simplicity of the expression.

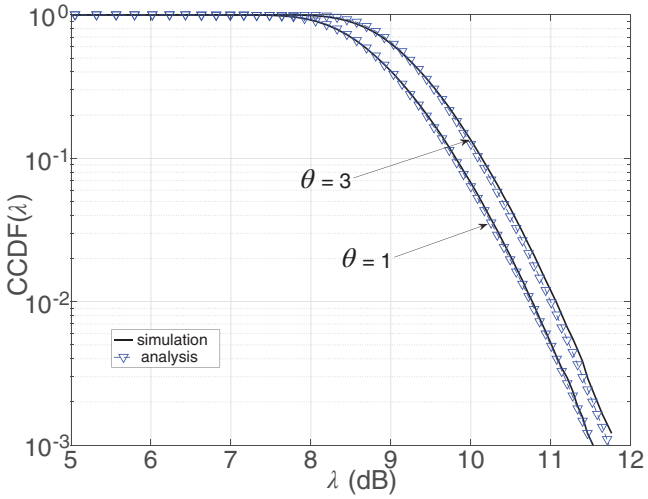


Fig. 4 *$CCDF(\lambda)$ versus λ (dB) for $N = 256$, $\theta = 1, 3$. Comparison between simulations and analysis*

Discussion: Developments and simulations results have shown that the PAPR of duplex multicarrier signals can be simply but accurately approximated in a Van Nee's fashion where the α coefficient is a function of the frequency distance Δ_f between the signals x_1 and x_2. However, we voluntary limited the analysis to the case where x_1 and x_2 are equipowered and share the same number of subcarriers. A similar but more general study case could then be carried out in case the two signals do not share the same parameters. In this case, the four parameters β, ν, q, and c should be tuned with respect to the different powers and subcarriers numbers of x_1 and x_2. To go further, the present analysis could even be extended to derive simple expressions of the CCDF of PAPR of multiplex multicarrier signals such as considered in [7, 9]. In that case, a more general mean PAPR expression $\mu = g(\theta_1, \theta_2, \ldots, \theta_{M-1})$ should be considered, where M is number of signals composing the multiplex one, and θ_i are the corresponding frequency distances.

In addition, it must be emphasised that the function g in (13) is empirically derived. Although Figure 3 shows that g accurately matches simulations, an analytical derivation of g should be investigated as well. Other studies could be undertaken from the present paper, such as the analysis of the instantaneous signal envelope $|x(t)|$ instead of $|x(t)|^2$, and its adaptation to duplex signals featuring low subcarriers numbers (typically $N \leq 64$) by using results in [12].

Conclusion: This paper highlights the dependency between the frequency distance between signals composing a non-contiguous duplex signal and the resulting PAPR. By deriving the mean PAPR of such kind of signal, we came to the conclusions that (i) the lowest PAPR value of the duplex is given when the two signals are contiguous in frequency, corollarilly (ii) the larger the frequency gap, the larger the PAPR, but (iii) the mean PAPR is upper bounded for large frequency distance between the signals that compose the duplex. To go further, this study could be enriched by considering more than two users of unequal powers and different subcarriers numbers.

Received: *2 September 2020* Accepted: *27 October 2020*
doi: 10.1049/ell2.12049

References

1 Louët, Y., et al.: Global power amplifier efficiency evaluation with PAPR reduction method for post-OFDM waveforms. *2018 15th International Symposium on Wireless Communication Systems (ISWCS)*, August 2018, Lisbon, Portugal. IEEE, Piscataway, NJ (2018)

2 Van.Nee, R., De.Wild, A.: Reducing the peak-to-average power ratio of OFDM. VTC '98. *48th IEEE Vehicular Technology Conference. Pathway to Global Wireless Revolution (Cat. No.98CH36151)*, May 1998, Ottawa, ON. IEEE, Piscataway, NJ (1998)
3 Van.Nee, R., Prasad, R.: *OFDM for Wireless Multimedia Communications*. Artech House, Norwood, MA (2000)
4 Ochiai, H., Imai, H.: On the distribution of the peak to average power ratio in OFDM signals. *IEEE Trans. Commun.* **49**(2), 282–289 (2001)
5 Rajbanshi, R., Alexander, M.W., Minden, G.J.: Peak-to-average power ratio analysis for NC-OFDM transmissions. *2007 IEEE 66th Vehicular Technology Conference*, October 2007, Baltimore, MD, pp. 1351–1355. IEEE, Piscataway, NJ (2007)
6 Kryszkiewicz, P., Kliks, A., Louët, Y.: PAPR analysis in noncontiguous OFDM systems. *Trans. Emerging Telecommun. Technol.* **28**(6), e3133 (2016)
7 Liu, X., et al.: Peak-to-average power ratio analysis for OFDM-based mixed-numerology transmissions. *IEEE Trans. Veh. Technol.* **69**, 1802–1812 (2020)
8 Rudziński, A.: Normalized Gaussian approach to statistical modeling of OFDM signals. *J. Telecommun. Info. Technol.* **1**(1), 54–61 (2014)
9 Jiang, T., et al.: Derivation of PAPR distribution for OFDM Wireless systems based on extreme value theory. *IEEE Trans. Wireless Commun.* **7**(4), 1298–1305 (2008)
10 Abramowitz, M., Stegun, I.: 5: Exponential Integral and Related Functions. In: Abramowitz, M., Stegun, I.A. (eds.) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, pp. 227–237. Dover Publications, Mineola, NY (1970)
11 Ochiai, H., Imai, H.: Peak-power reduction schemes in OFDM systems: A review. *Proceedings of the WPMC'98*, pp. 247–252. IEEE, Piscataway, NJ (1998)
12 Savaux, V., Louët, Y.: PAPR Analysis as a ratio of two random variables: Application to multicarrier systems with low subcarriers number. *IEEE Trans. Commun.* **66**(11), 5732–5739 (2018)