
# CENICS 2016

The Ninth International Conference on Advances in Circuits, Electronics and Micro-electronics

ISBN: 978-1-61208-496-1

July 24 - 28, 2016

Nice, France

**CENICS 2016 Editors**

Steffen G. Scholz, Karlsruhe Institute of Technology, Germany

Adrien Brunet, Karlsruhe Institute of Technology, Germany

Leo Schranzhofer, Profactor GmbH, Austria

# CENICS 2016

# Forward

The Ninth International Conference on Advances in Circuits, Electronics and Micro-electronics (CENICS 2016), held between July 24-28, 2016 in Nice, France, continued a series of events initiated in 2008, capturing the advances on special circuits, electronics, and micro-electronics on both theory and practice, from fabrication to applications using these special circuits and systems. The topics covered fundamentals of design and implementation, techniques for deployment in various applications, and advances in signal processing.

Innovations in special circuits, electronics and micro-electronics are the key support for a large spectrum of applications. The conference is focusing on several complementary aspects and targets the advances in each on it: signal processing and electronics for high speed processing, micro- and nano-electronics, special electronics for implantable and wearable devices, sensor related electronics focusing on low energy consumption, and special applications domains of telemedicine and ehealth, bio-systems, navigation systems, automotive systems, home-oriented electronics, bio-systems, etc. These applications led to special design and implementation techniques, reconfigurable and self-reconfigurable devices, and require particular methodologies to be integrated on already existing Internet-based communications and applications. Special care is required for particular devices intended to work directly with human body (implantable, wearable, ehealth), or in a human-close environment (telemedicine, house-oriented, navigation, automotive). The mini-size required by such devices confronted the scientists with special signal processing requirements

The conference had the following tracks:

- Design, models and language
- Application-driven Advances in Additive Manufacturing Technologies
- Special circuits

We take here the opportunity to warmly thank all the members of the CENICS 2016 technical program committee, as well as all the reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and effort to contribute to CENICS 2016. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations and sponsors. We also gratefully thank the members of the CENICS 2016 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope CENICS 2016 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in the field of circuits, electronics and microelectronics. We also hope that Nice, France provided a pleasant

environment during the conference and everyone saved some time enjoy the beautiful French Riviera.

**CENICS 2016 Advisory Committee**

Vladimir Privman, Clarkson University - Potsdam, USA
Sergey Y. Yurish, Technical University of Catalonia (UPC-Barcelona), Spain
Martin Horauer, University of Applied Sciences Technikum Wien, Austria
Adrian Muscat, University of Malta, Malta

**CENICS 2016 Research/Industry Chairs**
Ravi M. Yadahalli, S G Balekundri Institute of Technology, Belagavi, Karnataka, India

**CENICS 2016 Industry Liaison Chairs**
Falk Salewski, Muenster University of Applied Sciences, Germany

**CENICS 2016 Special Area Chairs**

**Application-oriented**
Josu Etxaniz Marañon, University of the Basque Country / Universidad del País Vasco / Euskal Herriko Unibertsitatea - Bilbao, Spain
**Sensors**
Yulong Zhao, Xi'an Jiaotong University, China
**Consumer electronics**
Sandra Sendra Compte, University of Granada, Spain

# CENICS 2016

# Committee

**CENICS Advisory Committee**

Vladimir Privman, Clarkson University - Potsdam, USA
Sergey Y. Yurish, Technical University of Catalonia (UPC-Barcelona), Spain
Martin Horauer, University of Applied Sciences Technikum Wien, Austria
Adrian Muscat, University of Malta, Malta

**CENICS 2016 Research/Industry Chairs**

Ravi M. Yadahalli, S G Balekundri Institute of Technology, Belagavi, Karnataka, India

**CENICS 2016 Industry Liaison Chairs**

Falk Salewski, Muenster University of Applied Sciences, Germany

**CENICS 2016 Special Area Chairs**

**Application-oriented**
Josu Etxaniz Marañon, University of the Basque Country / Universidad del País Vasco / Euskal Herriko Unibertsitatea - Bilbao, Spain

**Sensors**
Yulong Zhao, Xi'an Jiaotong University, China

**Consumer electronics**
Sandra Sendra Compte, University of Granada, Spain

**CENICS 2016 Technical Program Committee**

Amr Abdel-Dayem, Laurentian University, Canada
Amir Shah Abdul Aziz, TM Research & Development, Malaysia
Adel Al-Jumaily, University of Technology, Sydney
Said Al-Sarawi, The University of Adelaide, Australia
Mohammad Amin Amiri, Iran University of Science and Technology, Iran
Jose Hugo Barron Zambrano, Autonomous University of Tamaulipas, Mexico
Henri Basson, University of Lille North of France (Littoral), France
Lotfi Bendaouia, ETIS-ENSEA, France
Manuel José Cabral dos Santos Reis, University of Trás-os-Montes e Alto Douro, Portugal
Javier Calpe, University of Valencia, Spain
James M. Conrad, University of North Carolina at Charlotte, USA
Jose Carlos Meireles Monteiro Metrolho, Polytechnic Institute of Castelo Branco, Portugal

**Copyright Information**

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

**Table of Contents**

# Color Invariant Study for Background Subtraction

Lorena Guachi, Giuseppe Cocorullo, Pasquale Corsonello, Fabio Frustaci, Stefania Perri
Department of Electronics, Computer Sciences and Systems
DIMES - University of Calabria
Arcavacata di Rende, Italy
e-mail: loreanggeles@hotmail.com, g.cocorullo@unical.it, p.corsonello@unical.it, ffrustaci@deis.unical.it, perri@dimes.unical.it

***Abstract*—Effectiveness detection to extract objects of interest is a fundamental step in many computer vision systems. In real solutions, the accurate Background Subtraction (BS) is a challenge due to diverse and complex background types. Being the color widely used as descriptor to improve accuracy in several BS algorithms, in this paper we analyze four Color Invariants (CIs) based on the Kubelka-Munk theory combined with Gray scale. The capability of several CIs combinations in segmenting foreground is evaluated referring to five video sequences. This experimental study provides a point-of-view to choose the best color combination considering accuracy and the channel numbers which can be applied for image segmentation. The results demonstrate that the combination of the color invariant H with Gray scale achieves higher performance for foreground segmentation for both indoor and outdoor video sequences. Furthermore, it uses the minimum number of color channels.**

***Keywords-image processing; background subtraction; color invariant.***

## I. INTRODUCTION

In the recent past, Background Subtraction (BS) has gained an extensive application as a fundamental pre-processing task of video systems especially to detect objects of interest (vehicles, people, animals and so on) for security, traffic monitoring, surveillance systems among others, which include people counting, intrusion detection and tracking [1][2].

The BS algorithms typically use five features as descriptor: color, edge, stereo, motion and texture features [2], each one having specific characteristics to handle different environments and critical situations as motion changes, viewing direction, structure background and illumination changes. In order to be more robust in presence of critical situations, some algorithms tend to combine different features. The multi-scale region BS algorithm [21] performs the Gaussian Mixture modeling in conjunction with color histograms, texture information, and consecutive division of image regions to detect efficiently edges of the moving objects. Also in [22], the use of color and edge information is applied to handle slow illumination changes and camera noise, being able to run on standard platform for real time applications.

However, the capability of segmenting moving objects from video sequences is even a challenge in vision systems, where many algorithms work for specific environments in very controlled situations.

The rest of this paper is organized as follows. Section II describes most relevant related works. The color invariant descriptors are introduced in Section III. Section IV describes the algorithm used for the evaluation. The experimental results are presented in Section V. Finally, in Section VI we discuss the conclusion.

## II. RELATED WORKS

Color features are widely used in many algorithms and the ability to be very discriminative is mainly related to the way of representing colors in the image. Different color spaces provide different accuracies [3], due to several limitations in the presence of shadows, illumination changes, and camouflage.

Several works have been proposed in order to determine which color space is best for shadow detection and BS. In [4], an experimental study is presented to show how the "RGB", "XYZ", "YCrCb", "HSV" and the normalized "rgb" formats differently affect the moving objects classification and the shadow detection. For identifying shadow edges, the color opponent space is exploited in sunlit scenes. Furthermore, it is shown that the different uniform color opponent space is the most suitable for indoor environments, and the LαB is an appropriate selection for both indoor and outdoor environments [5].

The influence of several color spaces on the shadow detection has been evaluated in [6]. That work demonstrated that CIE L*u*v color model allows moving objects to be extracted efficiently also in the presence of moving shadows. As demonstrated in [7] the color space also affects tracking methods. As an example, in tracking applications, YCbCr and HSV color models are more suitable than RGB and Grayscale color models [7].

With the main objective of improving the achieved performances and avoiding time consuming color transformation, in [9], YUV color model is exploited for shadow detection in video conference applications. On the other hand, the statistical BS algorithm proposed in [10], separates the brightness from the chromaticity component in a pixel to exploit a computational color space.

Photometric color invariants as normalized rgb, hue (H), saturation (S), l1l2l3 and c1c2c3 are functions that describe the color configuration discounting shadows, highlight and shading. These functions are invariant to surface orientation, viewing direction and illumination conditions [8]. C1c2c3 model is adopted in [11] to exploit the spectral and geometrical characteristics for automatically shadow detection for static images and video sequences. This

 ISBN: 978-1-61208-496-1

approach firstly hypothesizes the presence of shadows, considering some initial evidence based on the fact that shadows darken the surface which they are cast upon.

Although many works presented in the literature have demonstrated how the color features interfere in the achieved accuracy, typical descriptors are based on specifically spectral information. On the contrary, the CIs are derived from a physical model and can take into account color spectral information and color spatial structure. Therefore, focused on CIs of the Kubelka-Munk theory, in [12] the CIs H, Wx, and Wy are introduced as descriptors in a novel BS algorithm to segment several video sequences in color similar situations. A first approach to combine the CIs with different color models is introduced in [13], where particularly CI H is used in conjunction with Gray scale to build robust descriptors with the aim of reducing post-processing task. Hence, this paper evaluates the possibility of combining the complete set of CIs (H, N, C and W) with Gray scale information. Several combinations are referenced to demonstrate that the efficiency in extraction of moving objects depends on the descriptors selected and combined through different logic operators. Overall results are presented for both indoor and outdoor experimental environments.

## III. COLOR INVARIANT DESCRIPTOR

This section presents the fundaments of CI and Gray color space. Since the color features are often very discriminative, many BS approaches use the color as descriptor, but in certain environments it has several limitations in the presence of camouflage, shadows and illumination changes. However, the combination with other features allows achieving more robust solutions for the BS [2].

Any method for describing CI model relies on assumptions about the physical variables involved and on photometric configuration [14]. Photometric CI are characterized as a function of surface reflectance, illumination spectrum and the sensing device, which consider the spatial configuration of color, and also the color spectral energy distribution coding color information [12]. Color spaces with properties independent of illumination intensity, reflectance property, viewing direction, and object surface orientation are defined as the color invariants [8]. These properties characterize the image color configuration discounting highlights, shadows, noise and shading. As an example, the Gaussian color model with spectral and spatial parameters is exploited in [12] to define a framework for the robust measurement of colored object reflectance. The CIs are derived from a physical reflectance model based on the Kubelka-Munk theory for colorant layers [14], where, illumination and geometrical invariant properties depend of the use of reflectance model.

The invariants are useful for materials as dyed paper and textiles, paint films, opaque plastics, dental silicate cements and up to enamel. A set of CIs derived from Kubelka-Munk theory is listed in TABLE 1. The latter shows how computing the CIs named H, N, C, and W, with E, Eλ and Eλλ being

TABLE 1. SET OF COLOR INVARIANTS

| *CI* | *Definition* |
|---|---|
| H | $E\lambda \,/\, E\lambda\lambda$ |
| N | $(E\lambda\chi \times E - E\lambda \times E\chi) \,/\, (E \times E)$ |
| C | $E\lambda \,/\, E$ |
| W | $E\chi \,/\, E$ |

the spectral differential quotients based on the scale-space theory [15].

The above defined CIs can be combined incrementally to achieve an alternative to invariant features extraction [14]. The Gray color space model is based on the brightness information and uses the measurement of amount of light (intensity). It is applied for object tracking often on a blob or a specific region [7]. However, taking into account that the color furnishes more information on the objects in a scene, it would be expected that this model can be used in conjunction with other models to achieve more robust solutions and higher accuracy than the basic separated models. For this reason, the Gray color space is included in the study here presented with the additional advantage of using a color space that does not require complex color transformations.

## IV. BACKGROUND SUBTRACTION ALGORITHM

The main computational steps required to classify the foreground pixels by using CIs can be summarized as follows: 1) RGB input frames are processed to obtain the CIs; 2) the background model is initialized by collecting, as the historical frames, the CIs obtained for the first *Nf* frames and the current background is computed; 3) as soon as the (*Nf+1*)-th frame is acquired, the foreground detection initiates and it is executed pixel-by-pixel by comparing the CIs of the current pixel the CIs of historical frames; 4) the current background model is updated taking into account the obtained classification.

To study the performance of CIs defined in Section III, we reference the algorithm schematized in Figure 1, which uses only ten historical frames. Some evaluated combinations of the features selected include a channel with Gray scale information whereas others are compounded only by CIs. Each channel is analyzed separately computing the percentage variation between the current frame and the historical mean. To classify the pixels within the generic frame of a video sequence into the background and the foreground sets, a thresholding is performed for each adopted descriptor. In our study, we refer to H, W, N, C and Gray scale components with the threshold values Th=55, Tw=90, Tn=90, Tc=90, Tg=60 that have been set experimentally to the values for which the number of wrong classified pixels is minimized for typical benchmark video sequences [17-20]. Several tests have demonstrated that higher threshold values reduce the accuracy in detecting foreground pixels, whereas smaller values increase the noise

 ISBN: 978-1-61208-496-1 

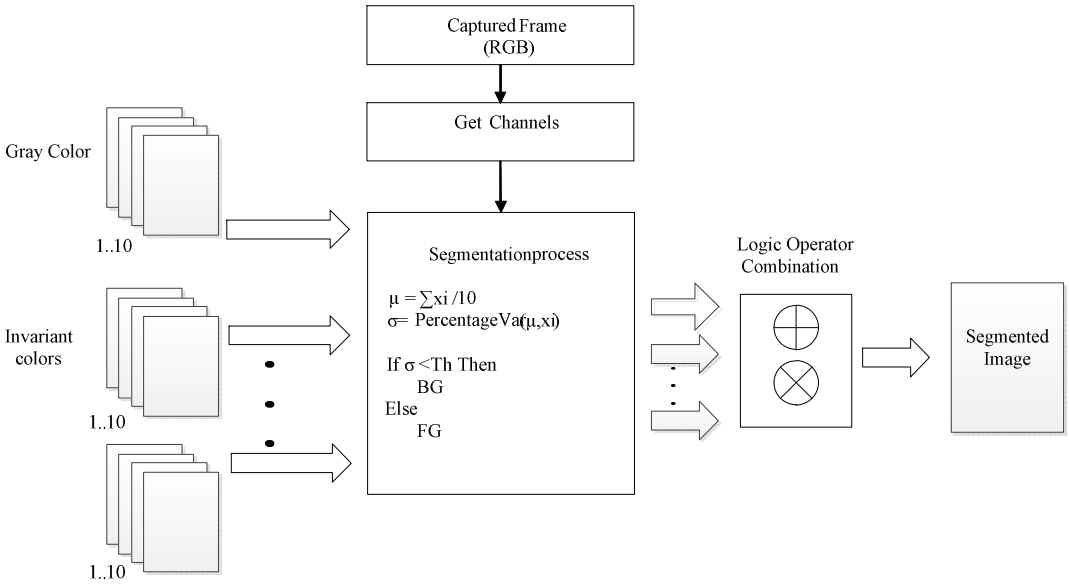



Figure 1. Background Subtraction diagram overview.

sensitivity. As suggested in [3], *Nf*=10 is initially used to model the background with a single Gaussian distribution. Several experiments demonstrated that increasing *Nf* does not significantly improve the accuracy but increases the memory requirement. Each component of the generic pixel of the current frame is compared to the mean value computed for the correspondent channel of historical frames. When the difference between the current examined channel and the historical mean overcomes the relative threshold, the current component is classified as belonging to a foreground pixel. Otherwise it is recognized as the component of a background pixel. Partial results obtained separately from the examined channels are then combined through appropriate logic operators to obtain the final segmented images. Background model is updated introducing a new frame at position zero, and discarding the oldest frame of position nine, all frames are sorted after each analysis.

## V. EXPERIMENTAL RESULTS

C++ software routines have been on purpose implemented to evaluate twenty three color combinations. Experimental tests have been done on different video sequences, related to both indoor and outdoor environments and the achieved performances are measured in terms of recall (Rec), specificity (Sp), precision (Pr), and percentage of correctly classified pixels (PCC). Rec measures the accuracy of the approach at the pixel level with a low False Negative Rate;

TABLE 2. COMPARISON RESULTS

| *Combination* | *Rec* | *Sp* | *Pr* | *PCC* |
|---|---|---|---|---|
| H AND GRAY | 11.13 | 99.77 | 81.82 | 93.88 |
| H OR GRAY | 52.65 | 89.87 | 27.61 | 87.50 |
| H AND N | 13.58 | 98.31 | 33.87 | 92.68 |
| H OR N | 54.60 | 82.18 | 18.08 | 81.74 |
| (H OR N) AND GRAY | 13.19 | 99.72 | 81.13 | 93.98 |
| (H OR N) OR GRAY | 59.34 | 82.10 | 19.27 | 80.68 |
| H AND C | 19.95 | 96.17 | 29.22 | 91.07 |
| H OR C | 50.09 | 85.77 | 26.24 | 83.40 |
| (H OR C) AND GRAY | 15.08 | 99.07 | 79.21 | 93.44 |
| (H OR C) OR GRAY | 56.93 | 89.11 | 27.61 | 87.13 |
| H AND W | 9.27 | 98.67 | 31.63 | 92.79 |
| H OR W | 59.79 | 76.14 | 15.44 | 76.41 |
| (H OR W) AND GRAY | 13.70 | 99.72 | 81.33 | 94.03 |
| (H OR W) OR GRAY | 64.02 | 76.06 | 16.26 | 75.30 |
| H OR N OR C | 64.08 | 75.50 | 15.97 | 74.77 |
| (H OR N OR C ) OR GRAY | 68.49 | 70.14 | 14.34 | 70.05 |
| (H OR N OR C) AND GRAY | 15.09 | 99.70 | 81.60 | 94.13 |
| H OR N OR W | 65.31 | 70.20 | 13.78 | 69.84 |
| (H OR N OR W) AND GRAY | 14.76 | 99.70 | 81.15 | 94.09 |
| (H OR N OR W) OR GRAY | 66.92 | 75.44 | 16.47 | 74.93 |
| H OR N OR C OR W | 68.76 | 69.63 | 14.19 | 69.59 |
| (H OR N OR C OR W) AND GRAY | 15.74 | 99.68 | 81.33 | 94.16 |
| (H OR N OR C OR W) OR GRAY | 70.95 | 69.59 | 14.54 | 69.72 |

Sp stimulates combinations with a low False Positive Rate; Pr favors combinations with a low False Positive Rate, and PCC measures the percentage of correct classifications [17]. The overall results are summarized in Table 2. The first column shows the logic operation applied to classify foreground pixels. As an example, the combination (H OR W) AND GRAY detects the generic pixel as foreground only if either its component H or its component W belongs to a foreground pixel, and also its Gray scale data is associated to a foreground pixel.

The results presented in Table 2 show that, as expected, differently combining CIs with Gray scale data very different accuracy can be achieved in detecting foreground objects. It is worth pointing out that the number of channels used to achieve a given accuracy significantly affects the computational complexity. In Figure 2. , the average accuracy obtained with each combination is directly related to the number of channels involved. Based on numeric analysis we can see that the combination (*H* OR *N* OR *C* OR *W*) AND *GRAY* achieves the best accuracy for indoor and outdoor experimental environments, and focused on the number of channels, the set of *H AND GRAY* reaches good performance on average with the minimum number of color channels. Figure 3. shows some of the segmented images obtained with these two combinations.

The results depicted in Figure 4 show the benefits achieved by introducing Gray scale in the set of CI combination to reduce the noise and improve the accuracy.

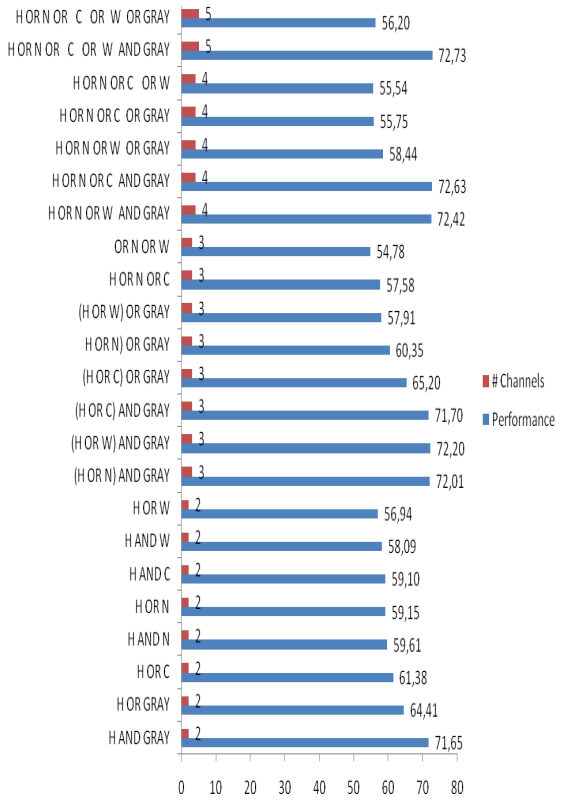



Figure 2. Analysis of the adopted combinations.


Figure 3. Results related to: a) Highway; b) Fountain; c) Pets2006; d)Bootstrap; e)Office.

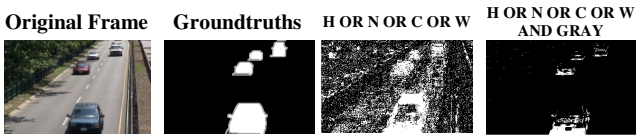



Figure 4. Results obtained introducing Gray scale information

All the above results have been obtained through the hardware system illustrated in Figure 5. The latter is based on the Raspberry Pi equipped with a camera module, able to capture RGB and grayscale images, and a Broadcom BCM2835 system on chip, consisting of an ARM1176JZF-S 700 MHz processor, a VideoCore IV GPU, 512 MB of RAM, and an SD card for long term storage and booting.

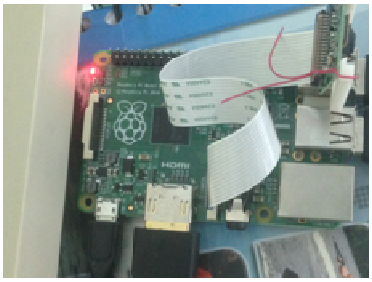


Figure 5. The hardware system used for tests

## VI. CONCLUSION

This paper has empirically compared the suitability of sets of CI combinations. Some of them include Gray scale. The tests measured the performance of the combinations

referring to indoor and outdoor experimental environments, demonstrating that the Gray scale insertion mitigates the problem of misclassified pixel. H and Gray scale combination provides the highest performance with respect to other combinations with the benefit of include particularly only two channels. Gray color model leads to background with less noise. On the contrary, CIs increase the noise due to the transformation operations, but, the combination with Gray color space allows achieving high effectiveness in the BS. These characteristics can be efficiently introduced in the algorithms for the image segmentation.

## REFERENCES


[1] A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," Computer Vision and Image Understanding, vol. 122, pp. 4-21., 2014.

[2] T. Bouwmans, "Traditional and recent approaches in background modeling for foreground detection: An overview," Computer Science Review, vol. 11, pp. 31 – 66, 2014.

[3] H. Zhou, Y. Chen, and R. Feng, "A novel background subtraction method based on color invariants," Computer Vision and Image Understanding, vol. 117, no 11, pp. 1589–1597, 2013.

[4] P. Kumar, K. Sengupta, and A. Lee, "A comparative study of different color spaces for foreground and detection for traffic monitoring system," In Intelligent Transportation Systems, 2002. Proceedings. The IEEE 5th International Conference on. IEEE, pp. 100-105, 2002.

[5] E. Khan and E. Reinhard, "Evaluation of color spaces for edge classification in outdoor scenes," In Image Processing, 2005. ICIP 2005. IEEE International Conference on. IEEE,, pp. III-952-5, 2005.

[6] C. Benedek and T. Szirányi, "Study on color space selection for detecting cast shadows in video surveillance," International Journal of Imaging Systems and Technology, vol. 17, no 3, pp. 190-201, 2007.

[7] P. Sebastian, Y. Vooi, and R. Comley, "Colour space effect on tracking in video surveillance," International Journal on Electrical Engineering and Informatics, vol. 2, no 4, pp. 298-312, 2010.

[8] T. Gevers and A. W. M. Smeulders, "Color-based object recognition," Pattern Recognition, vol. 32, no 3, pp. 453–464, 1999.

[9] O. Schreer, I. Feldmann, U. Goelz, and P. Kauff, "Fast and robust shadow detection in videoconference applications," in Video/Image Processing and Multimedia Comuunications 4th EURASIP-IEEE Region 8 International Symposium on VIPromCom. IEEE, , pp. 371–375, 2002.

[10] T. Horprasert, D. Harwood, and L.S. Davis, "Statistical approach for real-time robust background subtraction and shadow detection," In Proceedings IEEE International Conference on Computer Vision, , pp. 1-19, 1999.

[11] E. Salvador, A. Cavallaro, and T. Ebrahimi, "Cast shadow segmentation using invariant color features," Computer vision and image understanding, vol. 95, no 2, pp. 238-259, 2004.

[12] H. Zhou, Y. Chen, and R. Feng, "A novel background subtraction method based on color invariants," Computer Vision and Image Understanding, vol. 117, no 11, pp. 1589–1597, 2013.

[13] L. Guachi, G. Cocorullo, P. Corsonello, F. Frustaci, and S. Perri, "A novel background subtraction method based on color invariants and grayscale levels," In Security Technology (ICCST), 2014 International Carnahan Conference on. IEEE,, pp. 1-5, 2014.

[14] J. M. Geusebroek, R. van den Boomgaard, A. W. M. Smeulders, and H. Geerts, "Color invariance," Pattern Analysis and Machine Intelligence, IEEE Transactions On, vol. 23, no 12, pp.1338-1350,2001.

[15] L. M. J. Florack, B.M. ter Haar Romeny, J.J. Koenderink, and M.A. Viergever, "Scale and the Differential Structure of Images," Image and Vision Computing, vol. 10, no. 6, pp. 376-388, 1992.

[16] R. Gonzalez and R. E. Woods, "Digital Image Processing using Matlab, " 2004: Pearson Prentice Hall.

[17] N. Goyette, P. M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "Changedetection. net: A new change detection benchmark dataset," In Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on. IEEE, pp. 1-8, 2012.

[18] Statistical Modeling of Complex Background for Foreground Object Detection. [Online]. Available from: http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html. [accessed July 2016]

[19] Test Images for Wallflower Paper. [Online]. Available from: http://research.microsoft.com/en-us/um/people/jckrumm/wallflower/testimages.htm. [accessed July 2016]

[20] ChangeDetection.NET (CDNET). [Online]. Available from: http://www.changedetection.net. [accessed July 2016]

[21] P. D. Z. Varcheie, M. Sills-Lavoie, and A. Bilodeau, "A multiscale region-based motion detection and background subtraction algorithm," Sensors, vol. 10, no 2, pp. 1041-1061, 2010.

[22] S. Jabri, Z. Duric, H. Wechsler, and A. Rosenfeld, "Detection and location of people in video images using adaptive fusion of color and edge information," In Pattern Recognition, 2000. Proceedings. 15th International Conference on. IEEE, pp. 627-630, 2000.


 ISBN: 978-1-61208-496-1 

# A Hotspot Detection Method Based on Approximate String Search

Shuma Tamagawa*, Ryo Fujimoto†, Masato Inagi‡, Shinobu Nagayama*, Shin'ichi Wakabayashi*
*‡Graduate School of Information Sciences, Hiroshima City University
†Faculty of Information Sciences, Hiroshima City University
3-4-1 Ozuka-higashi, Asaminami-ku, Hiroshima, 731-3194 Japan
Email: ‡inagi@hiroshima-cu.ac.jp

***Abstract*—In this paper, we propose an approximate string search-based method for detecting hotspots on mask patterns used in very-large-scale integration (VLSI). In mask patterns for manufacturing VLSI chips, there are some local patterns which induce open/short circuits and thus failures. Such patterns are called hotspots. They are detected by optical simulation before manufacturing. Since it, however, requires very long time, it is desirable to detect hotspot candidates in order to limit the target regions of optical simulation. For the detection, methods which check matching between each pattern from a pre-defined hotspot library and the mask pattern are attracting attention. At first, our proposed method transforms the mask pattern and the pre-defined hotspot patterns, which are two-dimensional, into one-dimensional strings. Then, it finds hotspot candidates by using approximate string search to detect patterns similar to hotspot patterns. The transformation is performed in a particular way to efficiently realize quasi-two-dimensional search by using string search. In addition, we focus on the distance between wires as a metric to find hotspot candidates, and give a priority to patterns which have wires with a shorter distance. To evaluate the effectiveness of our method, we conducted some experiments.**

***Keywords–lithography; hotspot; optical simulation; approximate string matching.***

## I. INTRODUCTION

In recent decades, the density of semiconductor chips has greatly been increased with advances in very-large-scale integration (VLSI) technology. In lithography process for manufacturing VLSI chips [1], the circuit pattern (mask pattern) drawn on a photomask is transferred to a silicon wafer using photolithography machine. While 193nm wave length laser is currently used for the lithography process, the minimum pitch between wires is becoming smaller and has reached 14nm. Thus, because of diffraction during exposure, transfer of the mask pattern drawn on the photomask to the wafer sometimes fails. Hotspot is a place on the mask pattern where such a failure of transfer is likely to occur.

For manufacturing VLSI chips, a photomask is necessary. However, the manufacturing cost of the photomask is very high. To avoid remanufacturing photomasks, it is desirable to remove hotspots before manufacturing the photomask. Thus, optical simulation is conducted to detect hotspots. However, it takes very long time to conduct the optical simulation to the entire mask pattern. It is possible to shorten the simulation time by applying the optical simulation only to the hotspot candidates on the mask pattern. Thus, some methods for detecting hotspot candidates have been studied [2]-[6].

The method proposed in [2] is based on pattern matching [7]. The others [3]-[6] build a decision model, such as artificial neuron network, by learning hotspot/non-hotspot patterns, and detect hotspot candidates based on the model. [8] summarized related work including most of the methods.

In this study, we consider a method which conducts two-dimensional pattern matching by using string matching. In string matching, there are a lot of variation problems. String search problems are its variations to find the substrings same to a given pattern in a given string. An algorithm proposed by Knuth *et al.* [7] is well-known as an algorithm for the basic problem of string search. Hardware algorithms for the problems also have been proposed [9]-[11]. Our proposed method is based on approximate string search [12][13], which is a variation of string search and in which substrings similar to a given pattern are searched in a given string. Our method conducts approximate string search by transforming the two-dimensional mask pattern and the two-dimensional hotspot patterns into one-dimensional strings. When transforming these patterns, it is necessary to pad a lot of don't care characters to the hotspot patterns in order to match the widths of the mask and hotspot patterns. For efficient string search, we introduce a quantitative don't care character which represents a number of consecutive don't care characters. In addition, we confirmed that a pattern (in the mask pattern) similar to a hotspot pattern is more likely to be a hotspot if there are two wires in the pattern whose distance is smaller than that between the corresponding wires in the hotspot pattern. Thus, we propose a hotspot detection considering the distance between wires. Finally, we confirm the effectiveness of our proposed method comparing with a method based on two-dimensional simple matching.

The rest of this paper is organized as follows. First, in Section II, lithography, hotspot detection problem and related work, approximate string search problem, and an algorithm for the string matching problem are explained. Section III presents our hotspot detection method based on approximate string search. In Section VI, experimental results are shown. Finally, conclusions are described in Section V.

## II. PRELIMINARIES

### A. Lithography

Lithography (photolithography) is one of the processes for VLSI manufacturing, and a technology to transfer a circuit pattern drawn on a photomask, which is the master to replicate the circuit pattern, to a silicon wafer using a photolithography machine (exposure device).

The basis of lithography is shown in Figure 1. In lithography process, light is shed on the photomask and the mask

 ISBN: 978-1-61208-496-1 

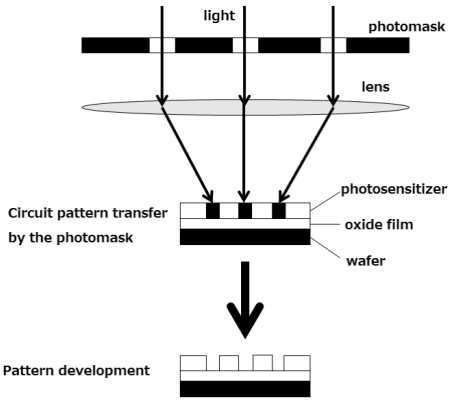



Figure 1. Lithography

(1) (2)

Figure 2. Hotspot: (1) a hotspot pattern, (2) transferred image

pattern drawn on the photomask is transferred to the wafer via lenses. While 193nm wave length laser is currently used for the lithography process, the minimum pitch between wires is becoming smaller and has reached 14nm. Thus, diffraction occurs during exposure. For this reason, there are cases in which the mask pattern drawn on the photomask cannot be correctly transferred to the wafer. Hotspot is a place on the mask pattern where such a failure of transfer is likely to occur. Let us consider a pattern (1) in Figure 2 as an example. Figure 2 (2) is a transferred image of the pattern (1). In the image (2), wires are fused at an unintended position. Thus, the pattern (1) is a pattern which induces a failure and considered as a hotspot.

### *B. Hotspot Detection Problem*

There exist two problems relating to hotspots: hotspot decision and hotspot detection (hotspot search). Hotspot decision problem is to determine whether a given small pattern is a hotspot (candidate) or not, while hotspot detection problem is to detect hotspots (hotspot candidates) from a given large mask pattern. In hotspot detection problems, a mask pattern and a known hotspot pattern (or a set of known hotspot patterns) are given, and patterns similar to the known hotspot pattern (or one of the known hotspot patterns) are searched in the mask pattern.

### *C. Related Work*

Most existing methods [2]-[6] solve a hotspot detection problem by scanning the mask pattern and solving hotspot decision problems.

The existing methods can be classified into [2], [3][4], and [5][6]. The main difference among the three groups is the information used for detection. In [2], a binary image of a pattern is encoded to a vector and used for detection. In [3][4], information such as the length and width of each wire is extracted from the pattern, and encoded to a vector. In [5][6], a pattern is encoded as a density-based data, and used for detection.

From another viewpoint, the existing methods can briefly be explained by their base algorithms: [2] is based on pattern matching, [3] is based on machine learning, [4] is a hybrid method combining machine learning and pattern matching, and [5][6] are based on fuzzy matching.

Here, we explain the overview of each group of the existing methods. In this paragraph, we describe the flow of a pattern matching-based method [2]. First, overlay a coarse grid on the mask pattern. Then, grid points on wire segments are colored. The coloring of the grid is transformed into a matrix, and then it is again transformed into a vector. Each hotspot pattern is transformed into vector in the same way. Then, the mask pattern vector and a hotspot pattern vector are matched by using the Knuth-Morris-Pratt (KMP) string matching algorithm [2][7]. If a hotspot pattern vector matches a part of the mask pattern vector, range pattern matching (RPM) is performed to them using a finer grid. In RPM, users can specify the ranges of the length and width of a segment of a wire in the hotspot pattern, etc., and thus flexible matching can be realized. This method realizes high accuracy by the hierarchical matching, in addition to low memory usage by the coarse global grid.

Second, we describe the flow of a machine learning-based method [3]. First, extract feature quantities for each known hotspot pattern considering the length, width, etc. of wire segments in the pattern, and construct a compact vector from the feature quantities of each pattern. Then, hierarchically build both artificial neural network (ANN) models and a support vector machine (SVM) models training by the vectors for accurate and robust decision-making. Then, a target pattern is applied hierarchical machine learning-based matching using the hierarchical model. This method realizes high accuracy, high robustness and low false-alarm ratio by the hierarchical model, in addition to short runtime by their compact feature vectors.

Next, we describe the flow of a hybrid method [4]. First, perform pattern matching for a target pattern and the hotspot patterns. If it matches one of the hotspot patterns, they conclude that it is a hotspot. Otherwise, matching based on machine learning is performed to the pattern. A pattern which is determined as a non-hotspot pattern in this stage is concluded to be a non-hotspot. A pattern which is determined as a hotspot pattern is applied machine learning-based matching again, using a different machine learning model to accurately examine if it is likely to be a hotspot or not. Only if it is determined as a hotspot by the second machine learning-based matching, they conclude that it is a hotspot. This method realizes high accuracy and short runtime by its hybrid strategy.

Then, we explain the common flow of fuzzy matching-based methods [5][6]. First, divide the mask and pre-defined hotspot patterns into tiles and calculate the density of wires in each tile as a feature quantity of the tile. Next, build a fuzzy matching model from the vectors of the feature quantities of the hotspot patterns. To build the model, calculate a distance, called city block distance (CBD), between every two hotspot patterns. If the CBD between two hotspot patterns is smaller

 ISBN: 978-1-61208-496-1 

than a threshold, they are placed in the same hotspot group. Next, a fuzzy region is extracted from each hotspot group in the fuzzy space. Then, fuzzy regions are expanded as far as they are legal as fuzzy regions. This is necessary to divide the fuzzy space into hotspot regions and non-hotspot regions in a fuzzy manner. Finally, determine if a given pattern is a hotspot or not, using the fuzzy matching model. If an inequality is satisfied on the model with a given pattern, it is determined as a hotspot. When building and using the model, a heavy weight is imposed to the center area of each pattern, because failures occur in the center area of hotspot patterns with a higher possibility than outside the center area. For hotspot detection in [5][6], candidate patterns, which include suspicious polygons (wires), are extracted and applied fuzzy matching. This method realizes high accuracy and low false-alarm ratio when using fine tiles.

Our proposed method is based on string matching. In the method, every pixel image of the mask and known hotspot patterns is sliced and converted into a string, and then approximate string search is performed. We expect that approximate string search realizes more flexible hotspot search since it can find not only the substrings exactly same as the given pattern but also similar substrings. Before approximate string search, each known hotspot pattern is analyzed, and congested areas, where wires are close to each other, are given heavy weights to find severe patterns in priority. We consider a hierarchical approach, which most existing methods employ, is promising. Therefore, this study is the first step of an attempt to develop a better method to replace a stage of hierarchical methods.

### D. Approximate String Matching Problem

Approximate string matching problem [12][13] is one of the string matching problems, and is a problem to determine if two given strings are similar or not. In this study, the similarity between strings is measured by the edit distance explained in the next subsection. If the edit distance is less than or equal to a given threshold, we consider they are similar each other.

### E. Edit Distance

Let us consider a pair of characters $(a, b)(\neq (\epsilon, \epsilon))$, where $\epsilon$ is an *empty character*, which represents nonexistence of any character. The operation transforming character $a$ in a string into $b$ is called an *edit operation*, and is denoted by $a \rightarrow b$. For example, let us consider a string $A = gzh$. If an edit operation $g \rightarrow f$ is applied to the first character of $A$, we get $A^{'} = fzh$ as the resultant string of the operation. If an edit operation $z \rightarrow \epsilon$ is applied to the second character of $A$, we get $A^{'} = gh$. If an edit operation $\epsilon \rightarrow j$ is applied to the empty character between the second and third characters of $A$, we get $A^{'} = gzjh$. Hereinafter, we call an operation $a \rightarrow b$ a *substitution* if $a \neq \epsilon$ and $b \neq \epsilon$. Likewise, we call an operation $a \rightarrow \epsilon$ a *deletion*, and call an operation $\epsilon \rightarrow b$ an *insertion*. Any string can be transformed into an arbitrary string by applying the edit operations. An edit operation has its cost denoted by $\gamma(a \rightarrow b)$. We assume the costs of edit operations satisfy the equation below.

$$\gamma(a \rightarrow a) = 0$$

$$\gamma(a \rightarrow b) + \gamma(b \rightarrow c) \geq \gamma(a \rightarrow c)$$

Suppose strings $A$ and $B$ on alphabets $\Sigma$ are given. A sequence of edit operations to transform $A$ into $B$ is denoted as $S = s_1, s_2, \ldots, s_m$. The cost of $S$ is defined as

$$\gamma(S) = \sum_{i=1}^{m} \gamma(s_i).$$

The minimum value among the costs of all the sequences each of which transforms $A$ into $B$ is defined as *the edit distance between $A$ and $B$* [12].

### F. Approximate String Search Problem

Approximate string search is to find substrings similar to a given pattern in a long input sequence. More precisely, approximate string search is to find all the substrings whose edit distance to the pattern $P$ are the minimum among all the substrings (or less then the given threshold $k$), in the input sequence $S$.

We here explain a dynamic programming-based algorithm for approximate string search [12][13]. Prepare an $(n + 1) \times (m + 1)$ two-dimensional array $D$, where $n$ is the length of the pattern $P = a_1 a_2 \cdots a_n$, and $m$ is the length of the input sequence $S = b_1 b_2 \cdots b_m$. The element $D(i, j)$ of $D$ is defined by the equation below. Then, $D(n, j)(1 \leq j \leq m)$ give the edit distances of substrings from the pattern. If the value is the minimum among all the $D(n, j)(1 \leq j \leq m)$ (or less than the user-defined threshold $k$), substrings whose terminal is $b_j$ are considered as those similar to the pattern. The initials of the substrings can be found by backtracing on $D$. The details of the identification of similar substrings of our proposed method are explained later.

$$D(0, 0) = 0, \quad D(0, j) = 0,$$

$$D(i, 0) = D(i - 1, 0) + del(a_i),$$

$$\begin{aligned} D(i, j) = \min\{&D(i - 1, j) + del(a_i), \\ &D(i, j - 1) + ins(b_j), \\ &D(i - 1, j - 1) + sub(a_i, b_j)\} \end{aligned}$$

$$sub(a_i, b_j) = \gamma(a_i \rightarrow b_j),$$

where the function $ins$, $del$, and $sub$ denote the insertion, deletion, and substitution costs.

Let us consider an example of searching substrings of $S = tttytzmy$ similar to $P = tyzm$. Here, we set the every cost of the insertion, deletion and substitution operations to 1. Then, we obtain an array $D$ shown in Figure 3 from the recursive definition of $D(i, j)$. Next, we search the minimum element of the bottom row of $D$. In this example, $D(4, 7) = 1$ has the minimum value. Next, we trace the edit operations back from the element. If a deletion operation was applied there, *i.e.,* the minimum function in the recursive definition chose a deletion operation, we move to the upper element. Likewise, if an insertion (substitution) operation was applied there, we move to the left (upper-left) element. Finally, we obtain $S < 3, 7 >= tytzm$ as a substring similar to the pattern $P$, where $S < i, k >= b_i b_2 \cdots b_k$. As shown in the example, in approximate string search, substrings similar to the pattern are found by calculating the edit distances from the middle of the input sequence. The time-complexity of this algorithm is $O(nm)$.

 ISBN: 978-1-61208-496-1 

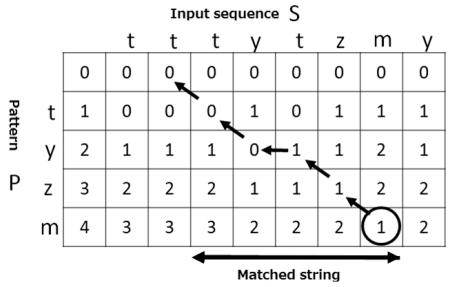



Figure 3. Edit distance array $D$

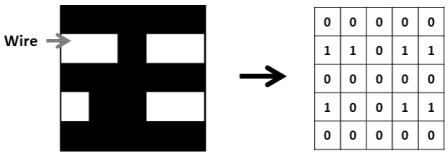



Figure 4. Image data and its corresponding array

## III. Hotspot Detection Based on Approximate String Search

In this section, we present our hotspot detection method based on approximate string search. In this method, the mask pattern and a hotspot pattern, which are both two-dimensional data, are transformed into one-dimensional strings to apply approximate string search calculating array $D$ by dynamic programming. In addition, we propose an extension of the method to give priority to be picked up to patterns more likely to be hotspots.

### A. Transformation into One-dimensional Data

Mask patterns and hotspot patterns are image data. We transform them into two-dimensional array of characters, in which wire area is represented by 1 and empty area is represented by 0.

An example is shown in Figure 4. In the left image in it, white areas show wires, and black areas show empty areas.

We transforms the two-dimensional arrays into one-dimensional data. In the transformation of the mask pattern, the two-dimensional array of the mask pattern is divided into rows. Then, the tail of the first row and the head of the second row is connected. And, the tail of the second row is connected to the head of the third row. Connecting all the rows according to the procedure, the two-dimensional mask pattern data is transformed into one-dimensional data. Next, we explain how to transform the two-dimensional hotspot pattern into one-dimensional data. The procedure of the transformation is shown in Figure 5. First, the array of the hotspot pattern is divided into rows, like the transformation of the mask pattern. Next, don't care characters are inserted to each row to fit the width of the hotspot pattern to that of the mask pattern. Then, the rows are connected into one-dimensional data, like the procedure of the mask pattern.

The insertion of don't care characters virtually realizes two-dimensional matching by one-dimensional matching. Figure 6 illustrates hotspot search with don't care characters. A don't care character can be represented by setting its substitution

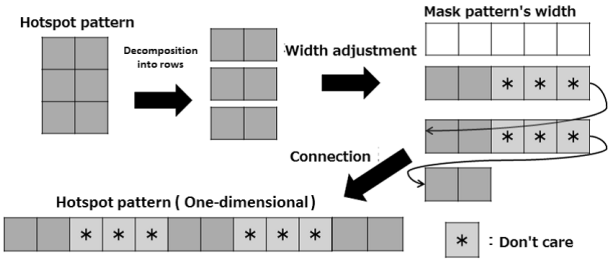



Figure 5. Transformation of hotspot data

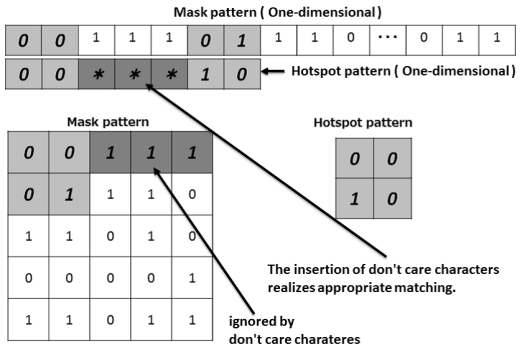



Figure 6. Matching after inserting don't care characters

cost to 0. To make the explanation simple, let us consider the case of matching the two-dimensional data (*i.e.*, not searching but matching). By padding don't care characters to the one-dimensional data of the hotspot pattern in order to fit the width of the hotspot pattern to that of the mask pattern, the values of the elements of the hotspot pattern and the values of the elements in the corresponding area of the two-dimensional mask pattern are matched (*i.e.*, compared). In the figure, the values from the third element to the fifth element in the first row of the mask pattern are matched with the inserted don't care characters in the hotspot pattern. (That is, the values from the third element to the fifth element of the mask pattern are ignored.) Therefore, the values from the first element to the second element in the second row of the two-dimensional mask pattern are appropriately matched with the values from the first element to the second element in the second row of the two-dimensional hotspot pattern by one-dimensional matching. This way, the hotspot pattern is matched with the corresponding area of the two-dimensional mask pattern.

### B. Dynamic Programming

In our method, since hotspot candidates are searched by using approximate string search, array $D$ is calculated by using the dynamic programming shown in the previous section. Except the first row and column, the value of each element of the array $D$ is calculated by using the value of its upper, left and upper-left elements. These calculations are done line by line from the top to the bottom.

Let us focus on the calculation of the element where the character of the hotspot pattern is don't care. In the case, consecutive $x$ don't care characters require $x$ rows of the calculations. Figure 7 illustrates the calculation of array

 ISBN: 978-1-61208-496-1

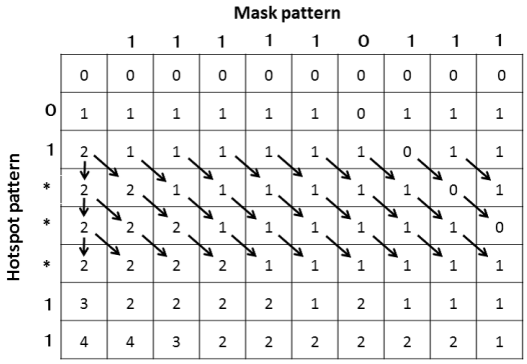



Figure 7. Calculation of array $D$ with consecutive don't care characters

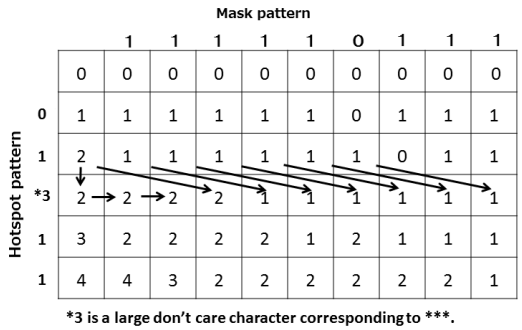



Figure 8. Calculation of array $D$ with a large don't care character

$D$ with consecutive don't care characters. In our method, such consecutive $x$ don't care characters are merged to one character, called a *large don't care*. If the character of the hotspot pattern is a don't care character, each element of the corresponding row is equal to its upper-left element. (Each element of the leftmost column is equal to its upper element.) On the other hand, if the character of the hotspot pattern is with a large don't care character, each element of the corresponding row is equal to the element at the $x$-th left column at the upper row. In this way, $x$ lines of calculations of don't care characters are realized by a line of calculations of a large don't care character. Figure 8 illustrates the calculation of array $D$ with a large don't care character.

### C. Detection of Hotspot Candidates

After calculating array $D$, substrings similar to the hotspot pattern are detected as hotspot candidates. To detect hotspot candidates, first, we focus on the elements with the minimum value in the bottom row of $D$. Each of these elements is considered as the terminal character of a hotspot candidate. Since we assume the hotspot candidate has the length same as the hotspot pattern, the initial character can be identified from the terminal character. The assumption is based on the fact that a hotspot pattern and candidates similar to the pattern are originally two-dimensional images, and have the same size or almost same sizes. Next, we focus on the elements with the minimal values in the bottom row of $D$. The elements whose values are less than or equal to a use-defined value (threshold) $k$ are chosen from the elements with the minimal values, and processed in the same way.

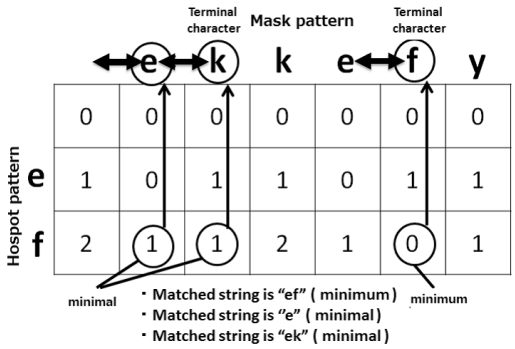



Figure 9. Detection of hotspot candidate

Figure 9 illustrates an example of hotspot candidate detection. First, the minimum values in the bottom row of array $D$ are found. In Figure 9, the value 0 of the 6th column is the minimum. The column with the minimum value corresponds to the character $f$ of the mask pattern. Thus, the substring $ef$ which has the terminal character $f$ and the length same as the pattern $ef$ is detected as a hotspot candidate. Next, the minimal values no more than the threshold $k$ in the bottom row of array $D$ are found. In this example, we set $k = 1$. The second and third columns in the bottom row satisfy the condition. Thus, the substring $e(= \epsilon e)$ whose terminal is in the second column is found as a hotspot candidate. Likewise, the substring $ek$ whose terminal is in the 3rd column is found as a hotspot candidate.

It seems that array $D$ requires a large amount of memory area. However, since we need only the bottom raw of $D$ to detect the hotspot candidates, the memory area for each raw (except the bottom raw) can be released after the calculation of the next raw. That is, we need to memorize only the current and previous rows at a time.

### D. Detection of Severe Patterns Considering Distance between Wires

Hotspot patterns are those which induce short or open circuits. If there exist patterns in the mask pattern similar to a hotspot pattern, they are likely to be hotspots. We focus on the distance between wires as a measure to determine the criticality of hotspot candidates. We performed simulations for two patterns by using an optical simulator [14]. The results of the simulations for the patterns (1) and (2) in Figure 10 are shown in Figure 11 (1) and (2), respectively. In Figure 10, pattern (1) and (2) are variations of the original hotspot pattern. In pattern (1) ((2)), the horizontal distance between the left and right wires is shorter (longer) than the original one (the horizontal segment of the right wire is longer (shorter) then that of the original one). As shown in Figure 11, pattern (1) clearly caused a short circuit, and thus it is a hotspot. Although pattern (2) might be a hotspot, too, it is not clear from the simulation. These results indicate that among patterns which have the same similarity to the original hotspot pattern, patterns with shorter distances between wires are more critical. Therefore, we propose an extension of our hotspot detection method to detect more critical patterns in priority among patterns with the same similarity, considering the distances between wires.

Let us consider two hotspot candidates similar to a hotspot

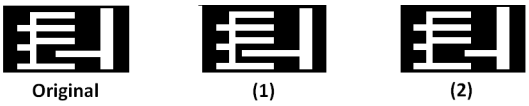



Figure 10. Patterns

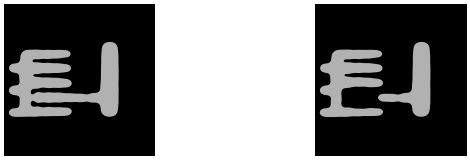



Figure 11. Transferred images

pattern. There exists the (horizontally) shortest distance between wires in the hotspot pattern. One of the hotspot candidates (shown in Figure 12(1)) has the shorter distance between wires at the corresponding area than the hotspot pattern, and the other (shown in Figure 12(2)) has the longer distance between wires at the corresponding area with the same difference. As mentioned above, the candidate with shorter distance is more likely to be a hotspot. However, these candidates have the same similarity (*i.e.*, edit distance) to the hotspot pattern. In our extension, more critical hotspot candidates are found by considering the distance between wires and increasing the costs of the edit operations around the corresponding area (between wire segments with the minimum distance). Although there are vertical and horizontal distances, it is difficult to consider both the distances at the same time because our method is base on one-dimensional matching. Thus, we here discuss only horizontal distance between wires. Vertical distance can be handled in the same way.

The procedure is as follows. First, find the place where the distance between wires is the minimum in the hotspot pattern, by scanning the hotspot pattern. Then, calculate the threshold by multiplying the minimum distance and a user-defined coefficient. The threshold decides if the distance is short or not. Next, find the places where the distance between wires is less than or equal to the threshold in the hotspot pattern, by scanning the hotspot pattern again. We assume the places are likely to cause short circuits. Next, decrease the substitution costs of the places (between wires). If the places of the hotspot pattern are applied substitution operations, the non-wire pixels (= 0s) are substituted by wire pixels (= 1s). Thus, patterns more likely to be hotspots with shorter distance between wires can be found. The procedure is depicted in Figure 13.

## IV. Experimental Evaluation

We performed experiments to evaluate the effectiveness of our method. We evaluated the runtime of a two-dimensional template matching-based method and our one, and evaluated the hotspot candidates detected by ours. The two-dimensional template matching-based method is a method to detect hotspot candidates by matching the hotspot pattern and the mask pattern moving the hotspot pattern from the top-left corner to the bottom-left corner on the mask pattern pixel by pixel.

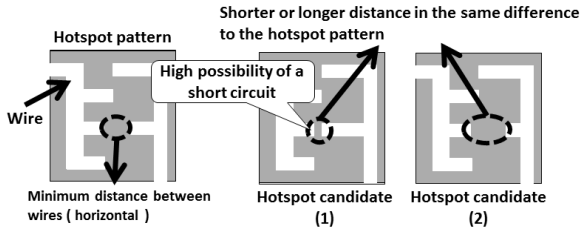



Figure 12. Hotspot pattern and its candidates

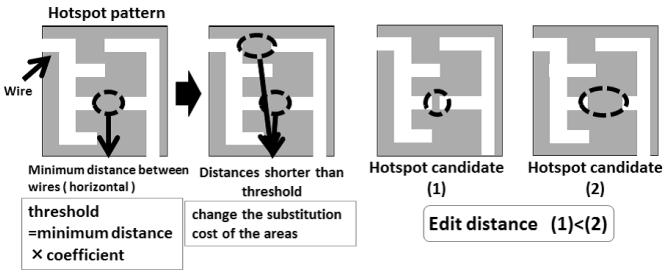



Figure 13. Detection of severe patterns

TABLE I. RUNTIME

| | Runtime (sec) |
|---|---|
| 2D matching | 152.02 |
| Non-LD method | 906.54 |
| Proposed method | 512.41 |

In addition, we also evaluated the runtime of non-LD method which is based on our method but without large don't care characters, to evaluate the effectiveness of large don't care characters. The experiments are conducted on a Linux PC (CentOS release 6.3) equipped with Intel(R) Core(TM) i7-3770 CPU @ 3.40GHz CPU.

### A. Comparison of Runtime

To evaluate runtime, we performed detection of hotspot candidates with a 1020x1020-pixel mask pattern and a 250x250-pixel hotspot pattern.

The runtime of each method is shown in Table I. The runtime of our method was longer than that of two-dimensional matching. However, we confirmed that large don't care characters improved the runtime.

Two-dimensional template matching starts its matching at the first column at the first row of the mask pattern, and moves to the next column until the right edge of the hotspot pattern reaches the right edge of the mask pattern. When the right edge of the hotspot pattern reaches the right edge of the mask pattern, it moves to the first column at the next row. This is repeated until the bottom-left corner of the hotspot pattern reaches that of the mask pattern. Thus, the time-complexity of two-dimensional matching is ((the width of the mask pattern) - (the width of the hotspot pattern)) $\times$ ((the height of the mask pattern) - (the height of the hotspot pattern)) $\times$ (the number of characters of the hotspot pattern)). On the other hand, the time-complexity of our proposed method is (the number of characters of the hotspot pattern) $\times$ (the number of characters

 ISBN: 978-1-61208-496-1 

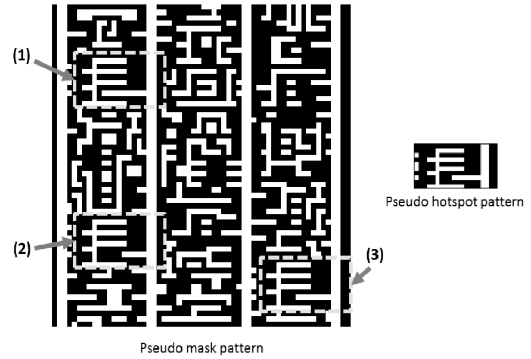



Figure 14. Pseudo patterns

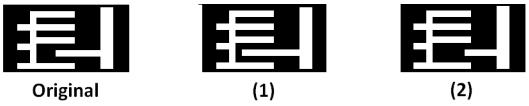



Figure 15. Detected candidates

of the mask pattern). In addition, in the dynamic programming in our proposed method, three elements are referenced in the calculation of each element, and thus it takes time to calculate the value of an element compared to the case of template matching. They are the reasons why the runtime of our proposed method was longer than that of two-dimensional matching.

### *B. Comparison of Detected patterns*

Next, we confirmed that if the detection of severe hotspot candidates considering the distances between wires is realized by our extension or not, using a pseudo mask pattern (60x60) and a pseudo hotspot pattern (10x20).

The standard cost of edit operations was set to 10, and the cost of each substitution operation at the places with distance between wires which is no more than the threshold was set to 5. The mask and hotspot patterns are shown in Figure 14, and the detected hotspot candidates are shown in Figure 15. The pattern (1) shown in Figure 15 is the same as the hotspot pattern, and thus their edit distance was 0. The edit distance of the pattern (2) is less than that of the pattern (3). As a result, the pattern (2) was detected in priority.

## V. CONCLUSIONS

In this paper, we proposed a hotspot detection method based on approximate string search, and an extension which detects severe hotspot candidates in priority. Our future work includes more in-depth experiments using open benchmark data, comparison with other existing methods, analysis of patterns more likely to be hotspots, and extensions of our method based on the analysis.

## REFERENCES

[1] Tatsuhiko Higashiki and Yasunobu Onishi, “Trends in semiconductor lithography technologies and Toshiba’s approach,” TOSHIBA review, Vol.67, No.4, 2012, pp.2–6.

[2] H. Yao, S. Sinha, J. Xu, C. Chiang, and X. Hong, “Efficient range pattern matching algorithm for process-hotspot detection,” in Proc. IET Circuits Devices Syst., 2008, pp. 2–15.

[3] D. Ding, A. J. Torres, F. G. Pikus, and D. Z. Pan, “High performance lithographic hotspot detection using hierarchically refined machine learning,” in Proc. 16th Asia South-Pacific Design Autom. Conf. (ASP-DAC), Yokohama, Japan, 2011, pp. 775–780.

[4] Jen-Yi Wuu, Fedor G. Pikus, and M. M-Sadowska, “Efficient approach to early detection of lithographic hotspots using machine learning systems and pattern matching,” in Proc. SPIE 7974, Design for Manufacturability through Design-Process Integration V, 79740U, April 04, 2011.

[5] S.-Y. Lin, J.-Y. Chen, J.-C. Li, W.-Y. Wen, and S.-C. Chang, “A novel fuzzy matching model for lithography hotspot detection,” in Proc. Design Autom. Conf. (DAC), Austin, TX, USA, 2013, pp. 1–6.

[6] W. Wen, J. Li, S. Lin, J. Chen, and S. Chang, “A fuzzy-matching model with grid reduction for lithography hotspot detection,” IEEE Trans. on CAD, Vol. 33, No. 11, Nov. 2014, pp.1671–1679.

[7] D. E. Knuth, J. H. Morris, Jr., and V. R. Pratt, “Fast pattern matching in strings,” SIAM J. Comput., vol.6, no.2, 1977, pp.323–350.

[8] Jhih-Rong Gao, Bei Yu, Duo Ding, and David Z. Pan, “Lithography Hotspot Detection and Mitigation in Nanometer VLSI,” in Proc. IEEE International Conference on ASIC (ASICON), 2013, pp.1–4.

[9] J. T. L. Ho and G. G. F. Lemieux, “PERG: A scalable FPGA-based pattern-matching engine with consolidated bloomier filters,” in Proc. 2008 IEEE International Conference on Field Programmable Technology, Dec. 2008, pp.73–80.

[10] Y. Sugawara, M. Inaba, and K. Hiraki, “Over 10Gbps string matching mechanism for multi-stream packet scanning system,” in Proc. International Conference on Field Programmable Logic and Applications, Aug. 2004, pp.484–493.

[11] B. C. Brogie, R. K. Cytron, and D. E. Taylor, “A scalable architecture for high throughput regularexpression pattern matching,” in Proc. 33rd International Symposium on Computer Architecture, 2006, pp.191–202.

[12] Yuichiro Utan, Shin’ichi Wakabayashi, and Shinobu Nagayama, “An FPGA-based text search engine for approximate regular expression matching,” in Proc. International Conference on Field-Programmable Technology, Dec. 2010, pp.184–191.

[13] Yuichiro Utan, Shin’ichi Wakabayashi, and Shinobu Nagayama, “A systolic algorithm for approximate regular expression matching and its FPGA implementation”, IEICE Journal D, Vol. J94-D, No.6, June 2011, pp.935–944. (in Japanese)

[14] Zhuo Li, et al., “ICCAD 2013 Contest,” http://cad-contest.cs.nctu.edu.tw/CAD-contest-at-ICCAD2013/problem_c/ [retrieved: June 2016].

 ISBN: 978-1-61208-496-1 

# Improving the Performance of a SOM-Based FPGA-Placement-Algorithm Using SIMD-Hardware

Timm Bostelmann and Sergei Sawitzki

FH Wedel (University of Applied Sciences)
Wedel, Germany
Email: bos@fh-wedel.de, saw@fh-wedel.de

***Abstract*—Programmable circuits and nowadays especially field-programmable gate arrays (FPGAs) are widely applied in demanding signal processing applications. In a previous work, we have introduced a method to improve the results of the netlist-placement for FPGAs with a self-organizing map (SOM). However, the presented algorithm conveys a comparably high computational effort. Considering modern, agile hardware / software codesign approaches, a slow design automation process can act as a kind of show-stopper, because software compilation is already distinctly faster. Thus, in this conceptual work, we present and evaluate different approaches to exploit the inherent parallelism of the SOM to increase the computation-speed. These approaches are based on using the single instruction multiple data (SIMD) capabilities of the central processing unit (CPU) and the graphics processing unit (GPU) for vector operations. Furthermore, we present benchmark results of an optimized implementation, based on using the CPU's SIMD units and introduce a concept for a GPU-accelerated implementation as work in progress.**

***Keywords–FPGA; netlist placement; GPU computing; parallelization; SIMD.***

## I. Introduction and Background

The ever-growing complexity of FPGAs has a high impact on the performance of electronic design automation (EDA) tools. A complete compilation from a hardware description language to a bitstream can take several hours. One step highly affected by the vast size of netlists is the NP-complete placement process. It consists of selecting a resource cell (position) on the FPGA for every cell of the applications netlist. Many current solutions optimize the placement iteratively. The academic EDA toolchain *Verilog-to-Routing (VTR) Project for FPGAs* [1] for example utilizes the simulated annealing [2] algorithm to solve the placement problem. Roughly described it starts with a random initial placement and applies random changes iteratively. The key of simulated annealing is the gradual reduction of the probability to keep disadvantageous changes over the time. Thereby the algorithm is able to leave local optima in early stages as well as to provide a fine optimization in later stages.

In [3], we have proposed a method to improve the placement results for FPGAs. Therefore, we have used a special SOM [4] to generate an initial placement optimized by a low temperature simulated annealing schedule. The placement generation consists of three stages (Figure 1). Initially for every cell of the netlist a training vector is generated. To guarantee that highly connected cells are represented by similar vectors, we use a hyperbolic distance function. The SOM is trained with these vectors in a random order. It consists of


Figure 1. Flowchart of the SOM-based placement algorithm.

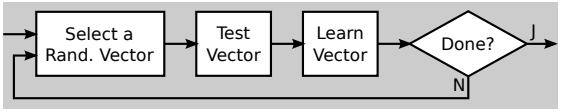



Figure 2. Flowchart of the internal SOM training.

a two-dimensional competition layer and a one-dimensional input layer. During the training the SOM clusters these vectors by similarity, so that the vectors of highly connected cells will cluster together on the competition layer of the map, thus reducing the prospective connection-lengths between the respective cells. Finally after the training has completed, the placement learned by the SOM is exported. Figure 2 shows the details of the training process. In the test stage, the neuron with the smallest Euclidean distance to the selected vector is determined. It is called the winning neuron. In the learning-stage the weights of the winning neuron and its neighbors are pulled towards the current stimulation. This makes them more susceptible to similar stimulations in the future and thereby induces the clustering of similar vectors. To prevent illegal placement results – which contain multiple occupations of a single resource cell – we temporarily block former winning neurons in the test process, so that they can not win again. The learning process is not affected by this measure. Consequentially the clusters of highly connected cells grow and displace other cells until they occupy the necessary space.

In [3], we have compared this method (including the final low temperature simulated annealing) to the regular simulated annealing process of VPR [5]. We were able to reduce the clock-rate determining critical path by six percent in average, for a set of common benchmark netlists. However – as mentioned in our previous work – we did not consider the computation time yet, knowing that our prototypic "proof of concept" implementation was not competitive regarding speed. In fact the proposed method – and especially the training of the SOM – conveys a high amount of computational effort. Fortunately, it has also a high amount of inherent parallelism, which makes a parallel computation approach very promising.

The rest of this paper is organized as follows. In Section II, we analyze the necessary computation time of our original im-

 ISBN: 978-1-61208-496-1 

TABLE I. Profiling results of the unoptimzed SOM implementation.

| Netlist | | FPGA | Relative Computation Time | | |
|---|---|---|---|---|---|
| Name | Size | Size | Test | Learning | Others |
| net16 | 256 | 16 × 16 | 69.0 % | 29.0 % | 2.0 % |
| ex5p | 1064 | 33 × 33 | 73.9 % | 25.8 % | 0.3 % |
| ex1010 | 4598 | 68 × 68 | 75.4 % | 24.6 % | 0.0 % |
| | | Average | 72.8 % | 26.5 % | 0.7 % |

plementation, as basis for further optimization. In Section III, we present the results of an optimized implementation, which we have created based on our findings and give a prospect to further optimizations using GPU-computation with OpenCL [6]. Finally in Section IV, we summarize our findings and give a prospect to further work.

## II. Parallelization

Table I summarizes profiling results of the unoptimized SOM implementation for different netlists from a Microelectronics Center of North Carolina (MCNC) benchmark-set [7]. It shows the relative computation times of the steps introduced above. Especially for larger netlists, the test and learning functions together consume almost 100 percent of the computation time. In this work, we focused on the test process because (with in average 73 percent) it consumes the highest amount of time. In the test process, the Euclidean distance between the stimulating vector and every neuron is determined. The neuron with the lowest distance is selected as winning neuron. The subfunction for the calculation of the distance consumes more than 99 percent of the test process (for example 368 seconds out of 369 seconds for the ex5p netlist). Based on these numbers, we have identified two levels of parallelism that could be exploited:

1) The vector operation to determine the distance $d$ between the stimulating vector $\vec{v}$ and a neuron's weight $\vec{w}$ as described in (1), assuming $\vec{v}$ and $\vec{w}$ have $N$ elements.
2) The calculation of all the distances and the selection of the lowest distance.

$$d = \sum_{i=0}^{N} (\vec{v}_i - \vec{w}_i)^2 \tag{1}$$

In a first attempt, we have optimized our SOM implementation by exploiting the parallelism of the vector operations. Therefore, we have created two alternative, parallel implementations of the distance function used heavily in the test loop. One implementation is using the processor's *Streaming SIMD Extensions (SSE)* for vector operations, the other is delegating the vector operations to the GPU using OpenCL.

## III. Results

Table II shows the results of the parallel implementations for different vector sizes. As this is only a preliminary test, we have used a desktop computer with an "Intel® Core™2 Duo E8400" processor. For a rough lower-bound approximation of the expectable GPU performance, we have used a "NVIDIA® GeForce® GTX 650" GPU. In comparison to the unoptimized implementation, the SSE implementation breaks even between vector sizes of 100 and 1000 cells, whereas

TABLE II. Time consumption of the parallel implementations of the distance function (1) for different vector sizes.

| | CPU | CPU SSE | | GPU OpenCL | |
|---|---|---|---|---|---|
| Vector Size | Time | Time | Speedup | Time | Speedup |
| 100 cells | 27 µs | 64 µs | 0.4 | 170 µs | 0.2 |
| 1000 cells | 200 µs | 74 µs | 2.7 | 300 µs | 0.7 |
| 10000 cells | 2000 µs | 112 µs | 17.9 | 400 µs | 5.0 |
| 100000 cells | 23 ms | 458 µs | 50.2 | 454 µs | 50.7 |
| 1000000 cells | 238 ms | 7000 µs | 34.0 | 669 µs | 355.8 |

TABLE III. Comparison of the computation times for one training cycle of our original SOM implementation and an improved version using SSE-accelerated vector operations.

| Netlist | | FPGA | Computation Time | | |
|---|---|---|---|---|---|
| Name | Size | Size | SOM CPU | SOM SSE | Speedup |
| net16 | 256 | 16 × 16 | 5 s | 2 s | 2.5 |
| e64 | 273 | 33 × 33 | 23 s | 7 s | 3.3 |
| ex5p | 1064 | 33 × 33 | 350 s | 31 s | 11.3 |
| seq | 1750 | 42 × 42 | 1476 s | 95 s | 15.5 |
| ex1010 | 4598 | 68 × 68 | 27211 s | 1259 s | 21.6 |

the GPU implementation brakes even between 1000 an 10000 cells. The SSE implementation and the GPU implementation break even at a vector size of 100000 cells. Even tough there are commercial FPGAs available with more than a million CLBs today, the netlists are typically partitioned to a smaller size before the placement and need much faster placement algorithms anyways. Further analysis has shown that the main problem of the tested OpenCL implementation lies in the small complexity of a single distance calculation. This causes a relatively large overhead for the memory transfer between host and GPU memory.

Based on these findings, we have improved our prototypic SOM implementation by using SSE for all vector operations. Table III shows the computation times of both SOM implementations for a subset of the netlists used in our previous work. The time is given for one training cycle, meaning the training of every vector. We have increased the overall speed of the training process by a factor of up to 20. Especially the larger netlists benefit from the parallelization, because the wider vectors give a better utilization of the SIMD hardware. However, the simulated annealing algorithm of VPR is still about one hundred times faster then our proposed SSE implementation. To bridge this gap, we propose to utilize the higher level of parallelism described above with an OpenCL implementation on a GPU. Thereby, we create bigger chunks of computational work and minimize the overhead for memory transfer between host and GPU. Ideally the complete training loop takes place on the GPU, so that a memory transfer is only necessary after the vector generation and for the placement export.

## IV. Conclusion

We have presented an evaluation of different approaches to exploit the inherent parallelism of a SOM-based FPGA-placenment-algorithm to increase the computation-speed. These approaches are based on using the SIMD capabilities of the CPU and the GPU for vector operations. Furthermore, we have present benchmark results of an optimized SOM-implementation based on using the CPUs SIMD units and

 ISBN: 978-1-61208-496-1

outlined a concept for an even faster GPU-accelerated implementation. In the future, we are planing to evaluate the latest OpenCL to FPGA synthesis approaches [8]. Those are especially auspicious because of the flexible memory structure and extended inter-kernel communication.

Another promising lead has been published recently in [9]. The authors present a fast SOM-based FPGA-placement-algorithm, that is using the Shimbel Index [10] for the vector-generation and thereby reduces the vector-size distinctly. We will evaluate if the optimizations presented in this paper are fitting to speedup this new approach even further.

## REFERENCES

[1] J. Rose et al., "The VTR project: Architecture and CAD for FPGAs from Verilog to routing," in ACM/SIGDA International Symposium on Field-Programmable Gate Arrays (FPGA), 2012, pp. 77–86.

[2] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," Science, vol. 220, May 1983, pp. 671–680.

[3] T. Bostelmann and S. Sawitzki, "Improving FPGA placement with a self-organizing map," in International Conference on Reconfigurable Computing and FPGAs (ReConFig), Dec 2013, pp. 1–6.

[4] T. Kohonen, Self-Organizing Maps. Springer, 1995.

[5] V. Betz and J. Rose, "VPR: A new packing, placement and routing tool for FPGA research," in International Conference on Field Programmable Logic and Applications (FPL). Springer, 1997, pp. 213–222.

[6] J. E. Stone, D. Gohara, and G. Shi, "OpenCL: A parallel programming standard for heterogeneous computing systems," IEEE Des. Test, vol. 12, no. 3, May 2010, pp. 66–73.

[7] S. Yang, "Logic synthesis and optimization benchmarks user guide version 3.0," Microelectronics Center of North Carolina, Tech. Rep., 1991.

[8] T. S. Czajkowski et al., "From OpenCL to high-performance hardware on FPGAs," in International Conference on Field Programmable Logic and Applications (FPL), Aug 2012, pp. 531–534.

[9] M. Amagasaki, M. Iida, M. Kuga, and T. Sueyoshi, "FPGA placement based on self-organizing maps," International Journal of Innovative Computing, Information and Control, vol. 11, no. 6, 2015, pp. 2001–2012.

[10] A. Shimbel, "Structural parameters of communication networks," The bulletin of mathematical biophysics, vol. 15, no. 4, 1953, pp. 501–507.

 ISBN: 978-1-61208-496-1 

# A Cost Model for SMARTLAM

Max Dobler*, James Gourlay‡, Steffen Scholz†, Andreas Schmidt†*
* Department of Computer Science and Business Information Systems,
Karlsruhe University of Applied Sciences
Karlsruhe, Germany
Email: max.dobler@gmail.com
‡ Design LED Products Ltd
Alba Innovation Centre
Livingston, United Kingdom
Email: james.gourlay@designledproducts.com
† Institute for Applied Computer Science
Karlsruhe Institute of Technology
Karlsruhe, Germany
Email: {steffen.scholz, andreas.schmidt}@kit.edu

***Abstract*—In this paper, we present a cost model for the production of small and medium batch series based on the SMARTLAM 3D-Integration approach, which allows fast prototyping without the normally neccessary allocation of tools. To accomplish this task, we use the concept of *total value*, which also includes the benefit. Because the majority of costs in SMARTLAM are caused by manufacturing overhead costs, we use an overhead calculation based on the machine-hour rate. Our model includes production cost calculation for the different used modules (i.e., Excimer- and $CO_2$ laser, aerosol-jet 3D printer, laminator, micro-assembly module) as well as for the coordination module and the whole manufactoring plant.**

***Keywords–micro-systems; additive manufacturing equipment; cost model.***

## I. INTRODUCTION

SMARTLAM (Smart production of Microsystems based on Laminated Polymer Films) is a project funded by the European Commission, with the goal to develop and integrate a complete set of micro technology based fabrication modules into a full production line without the normally required allocation and or modification of tools, masks, and hardware adaptations [1]. The aim is to develop a highly flexible and reconfigurable hot pluggable production unit which allows for the production of smaller lot sizes even down to customized tailored parts with lot size down to one part. With this approach a fast and flexible prototyping can be realized, but also competitive production of individual items or small batch series and mass customisation was in the scope of this project. The SMARTLAM 3D-Integration approach (3D-I) combines new material properties with state of the art, scalable 3D technologies such as aerosol jet printing, laser based micro structuring, laser welding, processes for combined micro milling, surface functionalisation, micro welding, and micro cutting [2]. Important questions in this context are how to calculate the costs of a micro technical component and if the production costs are competitive compared to established production methods. The SMARTLAM production cell is illustrated in Figure 1.

To answer these questions, the SMARTLAM production system will be analyzed and evaluated with the methods of cost accounting. The paper is structured as follows: In Section II, we present the different costs, which must be considered in our model. The main ideas of this paper are presented in Section III, where we develop the actual cost model. Starting with generic costs and the costs of the different modules (Excimer-Laser, $CO_2$-Laser, Aerosol-Jet 3-D printer, Laminator, and Micro-assembly) we develop the cost function of the overall process. Further, the paper will finish with a short conclusion in Section IV.

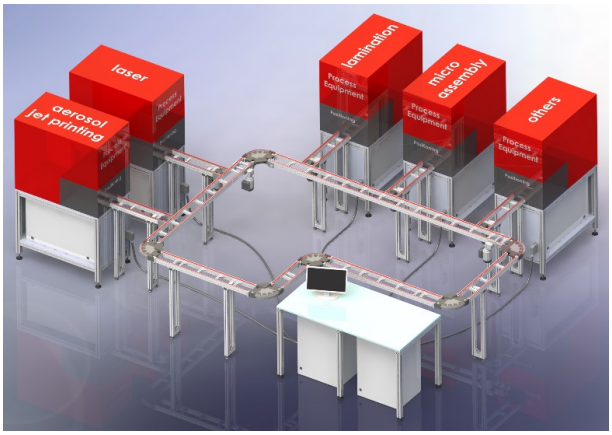



Figure 1. Simulation of SMARTLAM Production cell

## II. COST CONCEPTS

### A. Cost concept of total value

The literature embraces different definitions of cost concepts: the *cash-based* cost concept and the concept of *total value* [3], [4]. The concept of *total value* differs in the valuation of the consumption of goods compared to the first concept. The *total value* approach uses the final utility for evaluation; the *cash-based* concept uses the actual value on the buying market.

This paper establishes a cost model of the manufacturing process. The cost concept of *total value* is used in this paper to identify the costs, because it includes the benefits. In the following, all types of costs in SMARTLAM will be specified.

 ISBN: 978-1-61208-496-1 

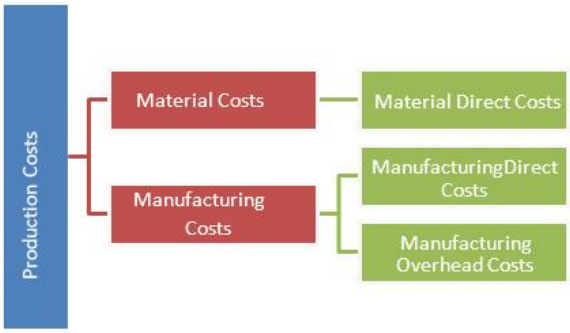



Figure 2. Types of costs for the production costs accountable in SMARTLAM (from [5])

### B. Production Costs

The consideration of the production costs is the essential basis of the carried out analyses. This is why in the following, this specific cost structure is addressed in detail. The production costs accrue during fabrication. There are differences between costs for the material (material costs) and arising costs during manufacturing a product.

Figure 2 shows a simplified presentation of the approach of Walter and Wünsche [3], since the material overhead costs and extra production direct costs were not considered. The production costs in SMARTLAM are divided into material and manufacturing costs, the latter further differed in direct and overhead costs. Direct costs will be assigned directly to a product or cost centre, overhead costs are assigned indirectly, as these costs are caused by several products or cost centres. The material overhead cost consists mainly of procurement of the material and administrative costs for this material. As this work considers only costs generated during production, these costs are not taken into account. For the evaluation of the manufacturing process in SMARTLAM, only the material direct costs and the attributable manufacturing direct and overhead costs are relevant.

In the following, these types of costs are explained in detail.

- Material direct costs: The direct costs of material are valued on basis of the value on the buying market.
- Manufacturing direct costs: These costs occur during the production and can directly be ascribed to a product.
- Manufacturing overhead costs: The majority of costs in SMARTLAM fall into this category. The overhead costs occur for several products and are spread prorated between the products. Following costs are assigned to this category especially for SMARTLAM:
  - Wage costs
  - Costs for supplies
  - Energy costs
  - Depreciation costs
  - Interest charges
  - Costs for operating materials

The following section describes the method for spreading the overhead costs.

### C. Production Costs

*1) Overhead calculation:* In addition to the direct costs there are overhead costs of the production. To get these costs split in a correct way among the products, the overhead calculation method is used. In the following, this method will be described in detail.

There are several overhead rates to characterize the overhead costs:

- Material overhead
- Manufacturing overhead
- Administration costs
- Selling expenses

For this paper, only the manufacturing costs are relevant, because these are the only ones who are collected in the production costs.

The SMARTLAM process provides a high flexibility and a module-based design, which makes machine-hour rate calculations the most suitable solution for SAMRTLAM. In the following, the method of machine-hour rate calculation will be explained in detail.

*2) Machine-hour rate calculation:* Due to high automation and new, flexible manufacturing plants the wages in production decreased significantely, followed by an increase in the manufacturing overheads (e.g., plant construction and controlling). The limits of the overhead calculation method are reached with products using the manufacturing plant in different ways and time, because the overhead rates depend mainly on the wages, which are not proportional to the production time.

Non machine-dependent overhead costs are not considered in this method and will be allocated to the direct costs via overhead calculation [4]. The SMARTLAM production has different products and a very flexible production planning, which allows to split the overhead costs precisely to each product depending on their production time.

The following calculation model in (1) is used for the machine-hours [3]):

$$machine\_hour\_rate(MHR) = \frac{machine\_overhead}{machine\_runtime} \quad (1)$$

To calculate the machine hour rate (MHR), the machine-dependent overhead costs should be determined very precisely. The following list displays the necessary parameters for calculation the MHR [4]:

- Machine replacement value ($C_{repl}$)
- Useful life of machine in years ($t_{life}$)
- Interest rate for replacement value (I)
- Repair and maintenance costs of machine per year ($C_{maint}$)
- Work center costs of machine per year ($C_{space}$)
- Fixed auxiliary material of machine per year (energy) ($C_{fix}$)
- Variable auxiliary material of machine per year (energy) ($C_{var}$)

The MHR for the manufacturing plant is calculated in the following way (2):

 ISBN: 978-1-61208-496-1

$$MHS = \frac{\frac{C_{repl}}{t_{life}} + C_{repl} * I + C_{maint} + C_{space} + C_{fix} + C_{var}}{t_{machine\_runtime}} \quad (2)$$

The methods of overhead calculation on basis of machine-hour rate calculation are suitable for the modelling of the manufacturing costs calculation above, because the overhead costs of the manufacturing plant can be registered through the production time which complements perfectly the flexible manufacturing process of SMARTLAM.

A calculation model for the manufacturing costs is sought, which shows the cost structures of the process and enables the calculation of the accumulating costs.

## III. Elaboration of the SMARTLAM Cost Model

This section shows the development of a model to identify and calculate the manufacturing costs of the production with the SMARTLAM 3D-I process.

### A. Generic costs

*1) Overhead costs of the SMARTLAM production:* In this section, all overhead costs of the manufacturing plant will be defined, i.e., all costs, which can be attributed to each product or are produced on the SMARTLAM plant. The production plant works autonomous most of the time; there is only one worker per shift necessary for operating of the facility. This worker refills the material (production raw material, auxiliary supplies, working supplies), modifies the plant between products, reacts to errors and picks up completed products and stores them in the warehouse.

The wages of the worker will be allocated on the products with the wages-rate. This value is calculated by (3):

$$wages\_rate(C_{wages}) = \frac{Wages}{machine\_running\_time} \quad (3)$$

The wage of one worker is about 80,000 €per year and the machine running time is 2,000 hours per year (assumption: work to capacity). This leads to a wage rate ($C_{wages}$) of 0.67 €/min. (4).

$$C_{wages} = \frac{80,000}{2,000 * 60}\ €/min = 0.67\ €/min. \quad (4)$$

The wages are charged to the product costs via their operating time on the manufacturing plant. This is done with the time the product is processed on the plant, multiplied with the wages-rate. Within the product cost calculation, these generic costs are added to the caused costs of a module. In the following, the unconsidered manufacturing plant-related costs will be discussed.

*2) Collective plant costs:* Besides the modules for each technology, the manufacturing plant also owns a material-flow system, which connects the modules and a collective control computer for the complete plant. This computer optimizes the production flow and coordinates the modules. As these parts are working with each module, the costs are allocated equally to each module. On basis of the machine-minute-rate ($C_M$) and the process time ($t_{pi}$) the variable machine overhead costs can be charged. In Table 1, all costs of the manufacturing plant are listed [6]. These costs can be allocated to each module.

TABLE I. OVERHEAD COSTS OF THE COLLECTIVE MANUFACTURING PLANT OF SMARTLAM

| Types of cost | Variable | Value |
|---|---|---|
| Replacement value of the machine | $C_{repl}$ | 65,000 € |
| Useful life of the machine | $t_{life}$ | 5 years |
| Interest rate of the replacement value | I | 5% |
| Work centre costs of the machine | $C_{wrk}$ | $25m^2 * 72\ €/m^2 = 1,800$ € |
| Variable auxiliary material (energy) | $C_{var}$ | 5 €/hour |

These costs accrue annually for the use of the machine and are calculated with an operating time of 2,000 hours per year. The model displayed in (5) is used to calculate the machine-hour-rate (MHR):

$$MHR = \frac{\frac{C_{repl}}{t_{life}} + \frac{C_{repl} * I}{2} + C_{wrk} + C_{var}}{t_{running}} \quad (5)$$

With the values of Table I, the MHR can be calculated with 13.23 € per machine-hour.

The production time in SMARTLAM is measured in minutes, therefore the machine-hour-rate is converted to the machine-minute-rate ($C_M$) with the value 0.22 €per production minute.

By adding up the wages costs (4), the overhead per minute raises to the value of 0.89 €.

### B. Functions for the production cost of the modules

The SMARTLAM 3D-I process enables a flexible production of several products with different specifications. Due to the modular concept of the manufacturing plant it is possible to reorganize the modules for each product. This makes it more difficult to calculate the production costs, because on one hand the calculation model has to be flexible and generic. On the other hand the model must fit all production processes. Variables, which stay unchanged for each product, will already be valued in the generic model with values of the production.

The rejects, the products with defects from the production, are separately considered for each module. In the microsystem technology it is called Yield ($Y$) matches the yield in percent (%) of the final products [7].

The production costs per film ($C_{sheet}$) of a process step, divided by the yield results in the total production costs ($C_{total}$) and reject costs ($C_y$) are shown in (6):

$$\begin{aligned} C_{total} &= \frac{C_{sheet}}{Y} \\ C_y &= \frac{C_{total}}{Y} - C_{total} \end{aligned} \quad (6)$$

The time of transport between the modules and for configuration of the module is considered in the time $t_{tc}$ and has a fixed value of 30 seconds.

*1) Costs of the Excimer Laser module:* The Excimer Laser module includes the Excimer Laser technology.

*a) Manufacturing overhead costs:* The Excimer Laser costs 65,000 € and is the most expensive laser technology of SMARTLAM. Additional costs of 12,000 € incur for the module housing and controlling. The life time is five years. The depreciation costs ($C_{Depr}$) for this module are calculated as shown in (7):

$$C_{Depr} = \frac{65,000 + 12,000€}{5\ years} = 15,400€/year \quad (7)$$

This yields to depreciation cost ($C_{InvExLas}$) of 0.13 €/min.

The acquisition costs above do not include all costs for optics and lenses, because some objects wear out multiple times a year. Every year there are wear costs of 23,000 € or 0.19 € per minute. Added to the acquisition costs, this results in overhead costs ($Cr_{ExLas}$) of 0.32 € per minute for the Excimer Laser module.

Several gases and masks are used for laser operations. The costs therefore are collected in the rate $C_G$ for gas (0.03 € per minute) and $C_M$ for the masks.

The overhead costs are allocated to the products using their process time. The time depends on the ablation volume ($V$) and the speed of the laser ($v_{ExLas}$), which depends on the material. Added to the time for the transport and changing the worktables ($t_{tc}$) it results in the formula for the process time ($t$) as shown in (8):

$$t = \frac{V}{v_{ExLas}} + t_{tc} \quad (8)$$

*b) Total cost function for Excimer Laser:* Cutting off a lot of material leads to disposal costs ($C_d$), which must be added to the production cost of the module.

For SMARTLAM there is a yield of 95% for operations with the Excimer Laser. Together with the manufacturing overhead costs it results in the total cost function for calculating the production costs of the Excimer Laser module as shown in (9):

$$C_{ExLas} = \frac{(Cr_{ExLas} + C_M + C_G) * \frac{V}{V_{ExLas}} + t_{tc} + C_d}{Y} \quad (9)$$

*2) Costs of the $CO_2$ Laser module:* The $CO_2$-Laser module includes the $CO_2$ Laser technology. Compared to the Excimer Laser the acquisition value of the $CO_2$ Laser of 45,000 € is significantly cheaper and so, the depreciation per year is only 11,400 € ($C_{InvCo_2Las} = 0.10e/min.$).

Costs for Gas ($C_G$) amounts to 0.01 € per minute and are therefore also cheaper compared to the Excimer Laser.

*a) Total cost function for the $CO_2$ Laser module:* The total cost function is similar the Excimer Laser, but there are no costs for masks. The yield is also 95%. Equation 10 shows the formula (with ablation speed $v_{co_2Las}$, ablation volume $V$):

$$C_{Co_2Las} = \frac{(C_{InvCo_2Las} + C_G) * (\frac{V}{v_{Co_2Las}} + t_{tc}) + C_d}{Y} \quad (10)$$

*3) Costs of the Aerosol-Jet 3D printer module:* The Aerosol-Jet 3D printer module includes the Aerosol-Jet 3D printer technology. In the following, the cost model of this module will be defined.

*a) Manufacturing overhead costs:* The replacement value of this module includes the acquisition costs of the 3D printer and housing costs for the module. The acquisition costs are today 130,000 €, whereas the module costs are 12,000 €. The technology has a normal lifetime of five years, but depending on its operations it can last much longer [8]. The following formula shows the deprecation costs of the Aerosol-Jet 3D printer module.

Depreciation costs per minute are calculated in (11):

$$Depreciation = \frac{130,000\ € + 12,000\ €}{5\ years} = 28,400\ €/year \quad (11)$$

This yields to depreciation costs ($C_{InvAero}$) of 0.24 € per minute.

The process time can be measured based on the ink and the printing speed of the 3D printer. The amount of ink depends on the printing area and how thick the ink should be applied on the material. This can vary depending on the ink and the thermal sensitivity of the material [8].

The following formula shows the calculation of the process time, depending on printing volume $V$), ink quality $k_{ink}$, printing speed $v_{aero}$, and the transport time between the modules $t_{tc}$.

$$t = \frac{V * k_{ink}}{v_{aero}} + t_{tc} \quad (12)$$

The Aerosol-Jet 3D Printer needs a nitrogen supply. The costs for this gas are displayed in (13):

$$C_{N_2} * t = 0.001\ €/min * t \quad (13)$$

*b) Material costs:* Material costs for ink (14) include the exercise price multiplied with the printing area and the factor for the application thickness of the ink.

$$C_{ink} * V * k_{ink} \quad (14)$$

The exercise price depends on the raw material costs and the costs for the ink catridge.

*c) Total cost function Aerosol-Jet 3D printer:* The Aerosol-Jet 3D Printer has a yield of 90% for SMARTLAM. The following formula includes the manufacturing overhead costs and material costs of the printer, together in one total cost function.

$$C_{Aero} = \frac{(C_{InvAero} + C_{N_2}) * (\frac{V * k_{ink}}{V_{Aero}} + t_{tc})}{Y} + \frac{C_{ink} * V * k_{ink}}{Y} \quad (15)$$

*4) Costs for the Laminator module:* The Laminator module includes the laminating technology. In the following, a cost model for this module will be defined.

*a) Manufacturing overhead costs:* Compared to the other technologies in SMARTLAM, the laminator ranges in the midfield with a total of 65,000 €. This module is featured with a long life time of 10 years. For calculating the depreciation costs, the costs of housing and controlling of the module should be considered.

$$Depreciation = \frac{65,000\ € + 12,000\ €}{10\ years} = 7,700\ €/year \tag{16}$$

This corresponds to a depreciation value ($C_{InvLam}$) of 0.06 € per minute.

The overhead costs are allocated to the products throughout their operating time on the laminator module. The process time depends on the material (resistance to heat) and the necessary accuracy. If there is only one layer with functionality, the alignment time of the layers is nearly negligible. If both layers possess functionalities, the alignment is crucial for this process. The process time also depends on the length of the layer and the speed of the laminator. If none of the films has an adhesive layer, additional adhesive has to be applied, which causes further costs of auxiliary material in the amount of $C_{ad} * A_{film}$. After the lamination process the adhesive has to be cured.

$$t = (\frac{L_{film}}{v_{Lam}} + t_{pos} + t_{ad}) + t_{tc} \tag{17}$$

*b) Material costs:* In most cases, only one layer has functionality. This has the advantage that no exact alignment is necessary for a good quality. The costs of the second film belong to the material costs. If the second layer is also edited, the costs for material are already collected and therefore not listed in the laminator costs.

*c) Total cost function for the Laminator module:* Equation 18 combines the manufacturing overhead and material costs of the laminator module in one function. Table II shows the used variables.

TABLE II. PARAMETER FOR THE LAMINATING MODULE

| Parameter | Meaning |
|---|---|
| $C_{InvLam}$ | Depreciation costs for the adhesive in €/$min$. |
| $C_{film}$ | Material costs of the second layer |
| $Y$ | Yield |
| $C_{ad}$ | Material costs for the adhesive in €/min. |
| $t$ | Process time in sec. |
| $v_{lam}$ | Lamination speed im $mm^2/sec$. |
| $t_{pos}$ | Positioning of both layers in sec. |
| $t_{ad}$ | Applying time of the adhesive in sec. |
| $A_{film}$ | Laminating area in $mm^2$ |
| $L_{film}$ | Length of film in $mm$ |

$$C_{Lam} = \frac{C_{InvLam} * \frac{\frac{L_{film}}{v_{Lam}} + t_{pos} + t_{ad} + t_{tc}}{60}}{Y} + \frac{C_{film} + C_{Ad} * A_{film}}{Y} \tag{18}$$

*5) Costs of the Micro-Assembly module:* Applying adhesive, putting materials (e.g., LED chips) in place or curing of adhesive, as well as curing of the laminated films is done by this technology. In the following, a cost model for this module will be defined.

*a) Manufacturing overhead costs:* This technology has an acquisition value of 60,000 € and a lifetime of five years. With the costs for housing and controlling of the module (12,000 €) the depreciation costs are calculated

$$Depreciation = \frac{60,000\ € + 12,000\ €}{5\ years} = 14,400\ €/year \tag{19}$$

This yields to depreciation costs ($C_{InvMon}$) of 0.12 € per minute.

The manufacturing overhead costs are allocated to the products through their process time in the Micro-Assembly module. The time depends on the operation and further on material, adhesive and components. Furthermore factors like precision, thermal sensitivity and number of parts per film are important for calculation. The process time of this module is split into time for adhesive application, positioning of components and curing of adhesive. Necessary time frames are measured individually per part and must therefore be multiplied with the number of parts per film.

$$t = n * (t_{adh} + t_{pos} + t_{cur}) + t_{tc} \tag{20}$$

*b) Material costs:* There are material costs for adhesive ($C_{MatDis} * V * n$) and components ($C_{MatComp} * n$).

*c) Total cost function for the Micro-Assembly Module:* Equation 21 shows the manufacturing overhead costs and the material cost combined in one function. The yield of this technology is 99.38%. The used parameters can bee seen in Table III.

TABLE III. PARAMETER FOR THE MICRO ASSEMBLY MODULE

| Parameter | Meaning |
|---|---|
| $Y$ | Yield |
| $C_{InvMon}$ | Depreciation costs of the module in €/min. |
| $C_{ad}$ | Material costs for the adhesive in €/$mm^3$ |
| $t$ | Process time in sec. |
| $n$ | Number of operations in min. |
| $V$ | Volume of pockests in $mm^3$ |
| $t_{adh}$ | Time for filling one socket in min. |
| $t_{pos}$ | Positioning per socket in min. |
| $t_{cur}$ | Curing time of adhesive per pocket in min. |

$$C_{Mon} = \frac{C_{InvMon} * n * (t_{adh} + t_{pos} + t_{cur} + t_{tc})}{Y} + \frac{n * (C_{Ad} * V + C_{MatComp})}{Y} \tag{21}$$

## C. Cost function of an overall process

The production cost function of an overall process is put together with the cost functions of the modules. Depending on the process some module are used multiple times, which makes the overall function an accumulation of all cost functions. Costs for transport between the modules and controlling the

facility are covered with the time $t_{tc}$, which is used in every module.

The overhead costs of the manufacturing plant are allocated to the products with their production time on the facility. Normally there are several products on the plant at the same time, which makes the slowest module decisive for the overhead costs of a product.

The following function shows the calculation of the production costs of a film in SMARTLAM (use parameters explained in Table IV):

TABLE IV. PARAMETER FOR THE COST CALCULATION OF THE OVERALL PROCESS

| Parameter | Meaning |
|---|---|
| $C_{unit}$ | Costs per unit |
| $C_{process}$ | Costs of all process steps |
| $C_{Mtotal}$ | Machine-related overhead costs of a product |
| $C_{pi}$ | Costs of the process step $i$ |
| $C_{oh}$ | Overhead cost rate in €$/min.$ |
| $t_{i_{max}}$ | Time of the proces step with the longest operation in sec. |
| $n$ | Number of products on a film |

$$C_{Sheet} = C_{Process} + C_{Mtotal} = \sum_{i=1}^{n} Cpi + C_{oh} * \frac{t_{i_{max}}}{60} \quad (22)$$

The production costs per unit results from the division of the production cost of a film by the number of products on the film which are shown in (23).

$$C_{UNIT} = \frac{C_{sheet}}{n_{output}} \quad (23)$$

## IV. CONCLUSION

With the generic costs described in Section III-A and the cost functions of the modules described in Section III-B, which are combined in the overall cost function in Section III-C, the calculation of the production costs of any product, manufactured on the SMARTLAM production plant can be done. This model is applicable to any production using these six modules. If more technologies are added to the manufacturing plant the model has to be reworked.

Costs for transport between the modules, inspection and controlling of the plant are detected in the fixed time $t_{tc}$ in every module. Within the implementation of the first production plant new knowledge should replace the fixed time.

If there are new functionalities or an updated technology, the cost function of the modules must be expanded. Each addition can be done in the function of a module and does not change anything on the overall cost function.

## ACKNOWLEDGMENT

SMARTLAM is funded by the European Commission under FP7 Cooperation Programme Grant agreement No. 314580.

## REFERENCES

[1] S. Scholz, T. Mueller, M. Plasch, H. Limbeck, R. Adamietz, T. Iseringhausen, D. Kimmig, M. Dickerhof, and C. Woegerer, "A modular flexible scalable and reconfigurable system for manufacturing of Microsystems based on additive manufacturing and e-printing," Robotics and Computer-Integrated Manufacturing, vol. 40, 2015, pp. 14–23.

[2] "SMARTLAM - Home," http://www.smartlam.eu, [last accessed: 17.5.2016].

[3] W. G. Walter and I. Wünsche, Einführung in die moderne Kostenrechnung: Grundlagen - Methoden - Neue Ansätze (english translation: Introduction into modern Cost Accounting: Basic Principles – Methods – New Approaches. Springer, 2013.

[4] A. G. Coenenberg, T. M. Fischer, and T. Günther, Kostenrechnung und Kostenanalyse, (english translation: Cost Accounting and Cost Analysis. Schäffel Poeschel Verlag Stuttgart, 2009.

[5] M. Dobler, "Analyse der Leistungsfähigkeit und Wettbewerbsfähigkeit von Fertigungstechnologien im EU-Projekt SMARTLAM (english translation: Analysis of the Efficiency and Competitive Capacity of the Manufacturing Technology in the EU-Project SMARTLAM)," Bachelor Thesis, Karlsruhe University of Applied Sciences, 2013.

[6] "Interview Max Dobler with R. Adamietz: Anlagekosten in SMARTLAM," 2013.

[7] M. Dickerhof, "Ein neues Konzept für das bedarfsgerechte Informations- und Wissensmanagement in Unternehmenskooperationen der Multimaterial-Mikrosystemtechnik (english translation: A new Concept for an Appropriate Information and Knowledgemanagement for Business Cooperation in the Field of Multimaterial-Microsystem Technology)," Ph.D. dissertation, Institute for Applied Computer Science, Karlsruhe Institute of Technology, 2009. [Online]. Available: http://dx.doi.org/10.5445/KSP/1000011501

[8] "Interview Max Dobler with M. Hedges," 2013.

 ISBN: 978-1-61208-496-1

# Novel Conductive Inks For 3D Printing

## Preliminary Studies on Silver Ink Development and Curing Strategies

Ayala Kabla, Abd El Razek, Fernando de la Vega
PV Nano Cell Ltd.
Migdal Haemek, Israel
e-mail: ayala@pvnanocell.com, abd@pvnancell.com, fernando@pvnanocell.com

Leo Schranzhofer
Profactor GmbH
Steyr-Gleink, Austria
e-mail: leo.schranzhofer@profactor.at

***Abstract*— UV (ultraviolet) curable nanoparticle conductive inks for 3D printing present a novel solution for production of a wide variety of parts with integrated conductive circuits. Such inks are difficult to design due to the complex tradeoff between achieving high conductivity, obtainable by a high metal content ink, and requiring low ink viscosity in order to meet jetting requirements. Adding to that is the complexity of managing the competing curing and sintering processes of the printed polymer and metal constituents. In the current work, UV curable nanoparticle conductive inks have been formulated and tested by different curing and sintering techniques. It has been demonstrated that a low resistivity of less than 10 times the silver bulk resistivity is attainable by thermal sintering, even with non-conductive organic content of 60%. Photonic sintering demonstrated the ability to obtain a resistivity of below 60 times the silver bulk resistivity, which is higher than the resistivity obtained by thermal sintering, but achievable in shorter processing times within milliseconds.**

***Keywords*— printed electrtonics; 3D printing; silver ink; UV curable ink; curing; sintering; photonic sintering**

## I. INTRODUCTION

PV Nano Cell Ltd. provides a wide range of solvent based nano conductive Sicrys™ inks [1], available for printed electronics applications with 50 wt% (weight percent) solid concentration and six times bulk resistivity at 150 °C, 30 minutes thermal treatment. In order to expand its capabilities, PV Nano Cell is taking part in the EC (European commission) funded DIMAP project (Novel nanoparticle enhanced Digital Materials for 3D Printing and their application shown for the robotic and electronic industry), dealing with new materials for 3D printing [2]. PV Nano Cell will develop a novel UV curable nanoparticle conductive ink for 3D printing by PolyJet technology. The ink is based on PV Nano Cell's Sicrys™ silver nanoparticles, dispersed in a monomeric matrix from Stratasys Ltd. or GEO® specialty chemicals, formulated for reaching high conductivity of the cured/sintered tracks.

UV-curable inks have attracted increasing attention due to environmental considerations, fast curing speed, low energy consumption, fast overall manufacturing process, and the possibility to work with thermal sensitive substrates [3]. Moreover, UV irradiation has been demonstrated to be a viable alternative to thermal sintering. Polzinger et al. have shown resistivity values of about four times bulk silver after 80 seconds UV-sintering of commercially available Ag ink [4]. Hu Yating et al. have demonstrated the sintering of copper ink by UV irradiation for a few seconds, achieving a resistivity of $4\times10^{-5}$ Ω·cm for antenna applications [5]. The ability of UV-irradiation to sinter Ag nanoparticles has opened up new paths to incorporate conductive features into the polymeric media. This makes UV-curable conductive inks attractive as they will benefit from both, fast curing and fast sintering.

In the market, however, mainly pastes with high viscosity and only a limited number of UV conductive inks are available. Among these, Polychem UV/EB International Corp. [6] and Tanaka Kikinzoku International (Europe) GmbH [7] pastes can be found, as well as a UV Ag flexographic ink from Gwent Group Ltd. [8]. The published resistivity values of these pastes and inks are 70-100 μΩ·cm.

The preparation of UV-curable conductive inks has been demonstrated in the literature. Sangermano et al. prepared UV-curable acrylic inks for printing resistors by incorporating Ag nanoparticles into a monomer; and reported a resistivity of $5x10^{3}$ Ω·cm$^{-1}$, which is an improvement of two orders of magnitude as compared to the pure polymeric matrix [3]. They also prepared UV-curable ink using Ag precursors; starting from silver hexafluoroantimonate and in-situ production of the Ag nanoparticles during the UV-irradiation [9], [10]. A hybrid Ag based ink prepared for an antenna printing application by Chiolerio et al. has shown 5% conductivity of bulk silver after thermal sintering at 250 °C for 30 minutes [11]. Zhai et al. shown conductivity of $7.14\times10^{-5}$ Ω·cm for UV-curable ink after UV-curing and thermal sintering; they achieved the conductivity for the ink-jet printed ink after evaporation of water on a heating stage, curing under UV spot light source and sintering in a muffle furnace. Thus, they showed that even after curing, thermal treatment might lead to sintering of nanoparticles within the polymeric media [12]. Designing UV-curable conductive inks is challenging due to the complex tradeoff between good conductivity, achievable with high metal content, while maintaining both a low viscosity ink that meets jetting requirements and realizing fast curing rates. All the while, the ink constituents of metal nanoparticles and polymeric media must remain stable in the formulation.

 ISBN: 978-1-61208-496-1 

TABLE I: FORMULATION COMPOSITION

| | Wt.% | Wt.% from solids | Density (g/ml) | Vol % | Vol % from solids |
|---|---|---|---|---|---|
| Ag | 50 | 86.06 | 10.49 | 7.34 | 40.37 |
| Dispersant | 1.1 | 1.89 | 1.21 | 1.40 | 7.70 |
| PI | 2 | 3.44 | 1.2 | 2.57 | 14.12 |
| Monomer | 5 | 8.61 | 1.12 | 6.88 | 37.81 |
| Solvent | 41.9 | 0 | 0.789 | 81.81 | 0.00 |

The main challenges facing this project include the ability to obtain a high solid content ink in a monomer matrix with suitable viscosity for jetting, and reaching percolation of particles in the sintered and cured printed pattern that will allow low resistivity. In order to overcome these hurdles, the formulation is carefully adjusted with appropriate additives and solvents for jetability, the solid content is raised so that percolation can be reached in the dried and cured samples, and alternative curing/sintering technologies to standard oven thermal treatments are being researched for short processing times.

The aim of this work was to investigate the curing and sintering behavior of the monomeric silver inks and understand the effect of the different monomers on the resistivity in order to choose the most promising monomers for further formulation optimizations. In order to do so, several curing/sintering approaches were tested, including: thermal treatment in an oven, UV irradiation, photonic sintering and curing monitoring by FTIR (Fourier-transformed infrared) spectroscopy.

Photonic sintering is a technique that uses pulsed light of a Xenon flash lamp to sinter thin films of particles.[13]–[19] It is a fast sintering method with common processing time in the range of milliseconds that operates at room temperature. The delivered light energy is selectively absorbed by the target particles and transferred to heat that triggers coalescence. Due to the pulsed shape of the flash, the heat transfer and therefore the damage to the substrate is low.

TABLE II: FORMULATIONS PREPARED WITH PI

| sample # | monomer | viscosity @ 25 °C (cP) |
|---|---|---|
| | ***GEO*** | |
| D006 | GM1 | 5.6 |
| D009 | GM2 | 5.6 |
| D010 | GM3 | 5.3 |
| D011 | GM4 | 5.7 |
| D012 | GM5 | 5.8 |
| | ***STRATASYS*** | |
| D013 | SM1 | 5.4 |
| D014 | SM2 | 5.4 |
| D015 | SM3 | 5.9 |

Thus, heat sensitive materials like polymers, paper or textiles can be used as substrate materials for printed electronics. The structure of this paper is as followed: in section II the performed experiments are described, followed by the discussion of the results in section III. In conclusion, section IV summarizes the presented work and gives a brief outlook on future work.

## II. EXPERIMENTAL

Several different UV curable monomers were received from GEO (GM1-5) and Stratasys (SM1-3) and used as received, Ag nano particles were synthesized and prepared by PV Nano Cell.

### *A. Sample preparation:*

#### *1) Formulation preparation:*

Sicrys™ Ag particles were incorporated into monomers at a 50 wt% Ag concentration. A volatile co-solvent was used in order to lower the viscosity of the formulation and evaporate quickly out of the printed tracks, resulting in a high solid Ag content of ~85 wt% Ag in the dried printed pattern (40 vol%). Formulations D006 through D015 (see also Table II), consisting of Ag, dispersant, photoinitiator (PI), monomer, and solvent constituents, were prepared according to the formulation composition indicated in Table I. For curing rate characterization of the monomers, samples were prepared keeping the same component ratio but excluding Ag and dispersant.

#### *2) Curing:*

Formulations D006-D015 were printed with a dispenser and dried on a hot plate at 50 ºC. The samples were then cured/sintered by each of the following methods:

- UV LED
- UV-LED + Oven
- Oven
- Photonic sintering

UV curing was performed with a 395 nm LED (light emitting diode) lamp and 2 min exposure time, and thermal treatment was performed in a box oven at 200 ºC for 1hr.

#### *3) Sintering:*

Ink samples on glass were freshly prepared by draw down bar printing (40 µm) on glass substrates followed by sintering. Photonic sintering was performed using a Heraeus flash system [20], with an emission spectrum between 200 to 1000 nm. The lamp setting was fixed for all the different ink formulations.

### *B. Sample Characterization:*

The ink formulations were characterized for viscosity with a viscometer (Brookfield), and for particle size distribution and sedimentation rates using a centrifugal measurement technique (Lumisizer). In addition, curing rate was monitored by FTIR-spectroscopy (Brucker Tensor 37). Electrical performance was obtained by a four-point probe measurement set-up for resistance, and cross-section measurements were obtained using a surface profiler (Dektak IIA).

 ISBN: 978-1-61208-496-1 

TABLE III: PRELIMINARY OBTAINED RESISTIVITY OF INDIVIDUAL TECHNIQUES

| Sample | Resistivity (μΩ·cm) | | | |
|---|---|---|---|---|
| | *UV LED* | *UV LED + Oven* | *Oven* | *Photonic Sintering* |
| D006 | N.A. | 13 - 28 | 13 - 20 | N.A. |
| D009 | 17 - 7e9 | 20 | 21 - 43 | 112 – 231 |
| D010 | 212 - 2e9 | 10 – 13 | 10 - 15 | 638 |
| D011 | 1e9 - 3e9 | 16 - 30 | 16-26 | 258 |
| D012 | 900 | 50 - 82 | 77-92 | 62-85 |
| D013 | 80 - 4e7 | 17 - 23 | 23 | 9-78 |
| D014 | 50 - 83 | 38 - 72 | 50 | 723-1300 |
| D015 | 15 - 270 | 12 – 17 | 15-18 | 87 |

## III. RESULTS AND DISCUSSION

### A. *Ink viscosities:*

Table II lists the formulations (D006 – D015) and their measured viscosity. The formulations slightly differ from each other due different GEO monomer (GM) and Stratasys monomer (SM) types

### B. *Oven and UV curing/sintering*

As seen in Table III, the resistivity values resulting from the UV LED curing step alone are very high and the variation range of results is quite broad. After UV curing, the measured resistance from four-point probe was in the range of MΩ, though some measurement points showed lower resistances. The oven treatment, however, (with and without the UV curing step), significantly reduced the resistivity, along with a reduced variation in results.

Figure 1 shows the resistivity of the printed patterns as a function of line thickness for formulations prepared from different monomers, where UV cured and thermal oven treated samples are shown as filled data points, and thermally treated samples without UV curing are shown as un-filled data points.

As seen in the graphs, as the thickness of the samples increased, the resistivity increased as well. Thereby, the variation range of resistivity results indicated in Table III may have been due to line thickness variations of the printed samples.

### C. *Photonic Sintering*

In addition to thermal and UV treatment, freshly prepared samples were photonically sintered. Hypothetically, matrix curing and particle sintering can take place in a single step and in a short period of time. In that way, samples D009, D012, D013 and D014 became conductive (low conductivity was determined) within a single flash experiment, although D014 showing the highest resistivity of the flashed samples using the stated setting. Samples D010, D011 and D015 on the other hand became conductive after a second flash. For Sample D006 no conductivity could be established.

### D. *Monitoring Curing via FTIR:*

To investigate polymerization progress of the matrix, samples were prepared as described in the previous chapter. These samples were treated with UV light using a self-made UV-LED array in one minute steps. After each single step the irradiation was stopped for recording an IR-spectrum and irradiation was continued afterwards. For example, Figure 2 shows the decrease of the transmission-band at 1640 $cm^{-1}$ corresponding to an α,β–unsaturated ketone (functional part of the acrylic monomer) over time for SM3. As can be seen, after two minutes of irradiation, there is no significant change indicating curing of most of the monomer content.

## IV. SUMMARY AND CONCLUSION

It has been demonstrated that UV-curable low viscosity inks are feasible. Even with 60% by volume of non-conductive organic material, a resistivity of 10 μΩ·cm (6.3 times bulk silver resitivity) was achieved. In this work the curing step (UV radiation) did not sinter the Ag nano particles. This may stem from there being two competitive processes: the curing (cross-linking of the monomer) and the sintering of the nanoparticles; the curing being much faster than the sintering process therefore resistivities are high. As seen from thermal and UV LED sintering/curing treatments, there appears to be a dependence of the line thickness on the resulting resistivity, where thin lines achieve lower resistivity values. In addition, the UV LED conditions of 395 nm did not provide resistivity reduction of the printed tracks, and the oven treatment was a necessary step for achieving conductivity.

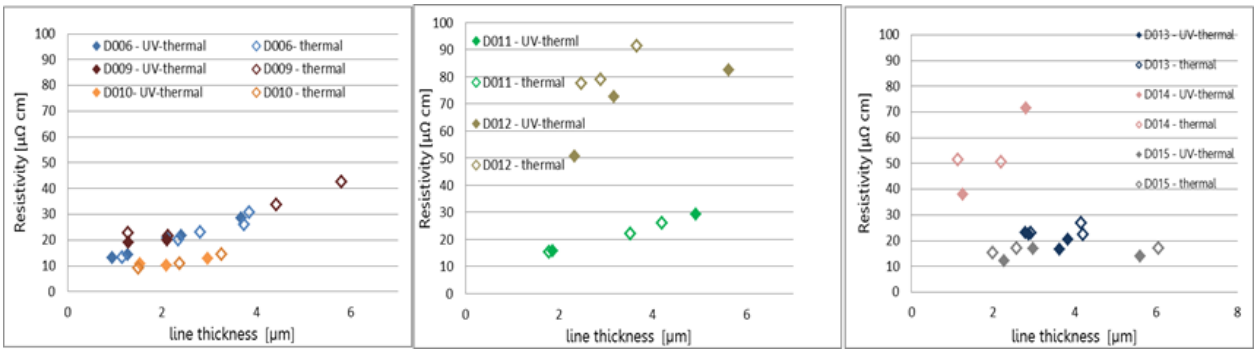



Figure 1: Resistivity of D006 – D015 formulations after UV-Thermal, and Thermal curing/sintering treatments.

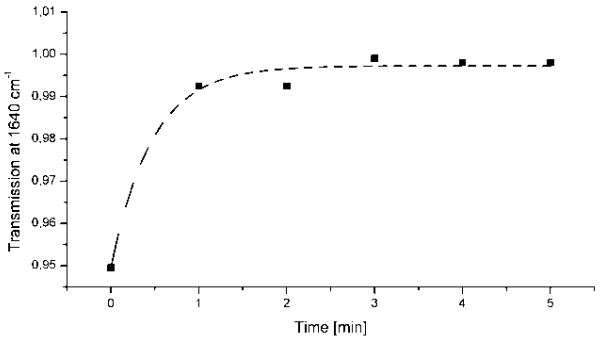



Figure 2. Curing progress based on FTIR spectroscopy. The data points correspond to the transmission recorded at 1640 $cm^{-1}$.

Another conclusion from the work is that the resistivity of the formulations differed from monomer to monomer. This might demonstrate the role of the monomer type on the percolation threshold or on the nano particle arrangement in the printed pattern. In addition, photonic sintering appeared to be a suitable method to cure and sinter the inks in one single step. Resistivity below 300 μΩ·cm was obtained after one and two flash experiments. Compared to the results for UV curing and thermal treatment, the measured resistances obtained from photonic sintering were higher. However, conductivity from flash sintering was achieved in a shorter period of time (within seconds).

Since the monomer matrix system was different in each of the ink formulations, the curing rates are expected to be different. The ideal setting for each ink has to be determined individually. The results shown here serve as a proof of concept for potentially using a photonic sintering strategy for monomeric inks. However, the flash parameters were not fully optimized in this preliminary experiment. Therefore, further work is necessary to find the optimal conditions for curing and sintering for each individual ink.

## ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon 2020 Programme for research and innovation under Grant Agreement No. 685937.

## REFERENCES

[1] 'PVNanoCell - Home'. [Online]. Available: http://www.pvnanocell.com/. [Accessed: 17-May-2016].

[2] 'DIMAP: DIMAP'. [Online]. Available: http://www.dimap-project.eu/. [Accessed: 17-May-2016].

[3] M. Sangermano, A. Chiolerio, G. Marti, and P. Martino, 'UV-Cured Acrylic Conductive Inks for Microelectronic Devices', Macromol. Mater. Eng., vol. 298, no. 6, 2013, pp. 607–611.

[4] B. Polzinger, et al., 'UV-sintering of inkjet-printed conductive silver tracks', 2011, pp. 201–204.

[5] Y. Hu, B. An, C. Niu, W. Lv, and Y. Wu, 'Application of nano copper conductive ink for printed electronics', 2014, pp. 1565–1567.

[6] 'Welcome to Polychem UV/EB International Corp::Products (UV Curable Silver Conductive Inks)'. [Online]. Available: http://www.polychem.tw/index.php?module=product&mn=1&f=content&tid=30276. [Accessed: 17-May-2016].

[7] 'Tanaka Europe GmbH: Conductive Pastes'. [Online]. Available: http://www.tanaka-europe.eu/products/thin-thick-film-metallization/thick-film-pastes-precious-metal-powders/conductive-pastes/. [Accessed: 17-May-2016].

[8] 'The Gwent Group, Leaders in paste manufacturing, sensor/ biosensor development and Instrumentation .' [Online]. Available: http://www.gwent.org/gem_flexographics_pastes.html. [Accessed: 17-May-2016].

[9] A. Chiolerio, L. Vescovo, and M. Sangermano, 'Conductive UV-Cured Acrylic Inks for Resistor Fabrication: Models for their Electrical Properties', *Macromol. Chem. Phys.*, vol. 211, no. 18, 2010, pp. 2008–2016.

[10] A. Chiolerio, et al., 'Ag nanoparticle-based inkjet printed planar transmission lines for RF and microwave applications: Considerations on ink composition, nanoparticle size distribution and sintering time', *Microelectron. Eng.*, vol. 97, 2012, pp. 8–15.

[11] A. Chiolerio, V. Camarchia, R. Quaglia, M. Pirola, P. Pandolfi, and C. F. Pirri, 'Hybrid Ag-based inks for nanocomposite inkjet printed lines: RF properties', *J. Alloys Compd.*, vol. 615, 2014, pp. S501–S504.

[12] D. Zhai, T. Zhang, J. Guo, X. Fang, and J. Wei, 'Water-based ultraviolet curable conductive inkjet ink containing silver nano-colloids for flexible electronics', *Colloids Surf. Physicochem. Eng. Asp.*, vol. 424, 2013, pp. 1–9.

[13] J. West, J. Sears, M. Carter, and S. Smith, *Photonic sintering of silver nanoparticles: comparison of experiment and theory*. INTECH Open Access Publisher, 2012.

[14] R. Abbel, T. van Lammeren, R. Hendriks, J. Ploegmakers, E. J. Rubingh, E. R. Meinders, and W. A. Groen, 'Photonic flash sintering of silver nanoparticle inks: a fast and convenient method for the preparation of highly conductive structures on foil', *MRS Commun.*, vol. 2, no. 4, 2012, pp. 145–150.

[15] D. J. Lee, S. H. Park, S. Jang, H. S. Kim, J. H. Oh, and Y. W. Song, 'Pulsed light sintering characteristics of inkjet-printed nanosilver films on a polymer substrate', *J. Micromechanics Microengineering*, vol. 21, no. 12, 2011, p. 125023.

[16] J. Niittynen, E. Sowade, H. Kang, R. R. Baumann, and M. Mäntysalo, 'Comparison of laser and intense pulsed light sintering (IPL) for inkjet-printed copper nanoparticle layers', *Sci. Rep.*, vol. 5, 2015, p. 8832.

[17] H.-J. Hwang, K.-H. Oh, and H.-S. Kim, 'All-photonic drying and sintering process via flash white light combined with deep-UV and near-infrared irradiation for highly conductive copper nano-ink', *Sci. Rep.*, vol. 6, 2016, p. 19696.

[18] K. C. Yung, X. Gu, C. P. Lee, and H. S. Choy, 'Ink-jet printing and camera flash sintering of silver tracks on different substrates', *J. Mater. Process. Technol.*, vol. 210, no. 15, 2010, pp. 2268–2272.

 ISBN: 978-1-61208-496-1 

[19] Nanotechnology Conference and Trade Show, M. Laudon, B. F. Romanowicz, and Nano Science and Technology Institute, Eds., *NSTI Nanotech 2006: 2006 NSTI Nanotechnology Conference and Trade Show, Boston, May 7-11, 2006*. Boston [Mass.]: Nano Science and Technology Institute, 2006.

[20] 'Heraeus Flash Lamp System'. [Online]. Available: https://www.heraeus.com/en/hng/products_and_solutions/arc_and_flash_lamps/flash_lamp_systems.aspx. [Accessed: 17-May-2016].

 ISBN: 978-1-61208-496-1 

# Novel Nanoparticle Enhanced Digital Materials for 3D Printing and their Application Shown for the Robotic and Electronic Industry

Introduction on the DIMAP project

Steffen Scholz, Adrien Brunet, Tobias Müller
Institute for Applied Computer Science
Karlsruhe Institute of Technology
Karlsruhe, Germany
Email:
[steffen.scholz;adrien.brunet;tobias.mueller2]@kit.edu

Anita Fuchsbauer
Profactor GmbH
Steyr-Gleink, Austria
Email: anita.fuchsbauer@profactor.at

*Abstract*— **In this paper the general idea of the DIMAP project is presented. The ultimate goal of the project is to fabricate innovative applications using new additive manufacturing technologies. In order to do so, four new materials and a new generation of 3D printers are developed. The eight main objectives of the project are described, in order to present the backbone of the concept. Then a status of the project is given and the future work is explained. The DIMAP project is application-driven and intends to lead to advances in additive manufacturing technologies, therefore this paper is strongly relevant to the ADAAM symposium.**

***Keywords: Digital Material, 3D Printing, Robotic, Electronic, Nanoparticule***

## I. INTRODUCTION

The here described DIMAP (Novel nanoparticle enhanced Digital Materials for 3D Printing and their application) idea focuses on the development of novel ink materials for 3D multi-material printing by PolyJet technology [1]. The state-of-the art of Additive Manufacturing (AM) will be advanced through modifications of the fundamental material properties, mainly by using nanoscale material enhanced inks [2, 3]. This widens the range of current available additive AM materials and implements functionalities in final objects. Therefore, applications will not be limited to rapid prototyping but can be used directly in production processes. DIMAP will show this transition in two selected application fields: the production of soft robotic arms/ joints and customized luminaires. In order to cope with these new material classes, the existing PolyJet technology [4] is further developed and therefore improved. The DIMAP project has the following objectives: additive manufactured joints, additive manufactured luminaires, ceramic enhanced materials, electrically conducting materials, light-weight polymeric materials, high-strength polymeric materials, novel multi-material 3D-printer and safe by design. With the development of novel ink materials based on nanotechnology improvement of the mechanical properties (ceramic enhanced and high strength polymeric inks), the electrical conductivity (metal enhanced inks) and the weight (light-weight polymeric materials) are achieved. Based on the voxel printing by PolyJet, these new materials lead to a huge broadening of the range of available digital material combinations. Further focus points during the material and printer development are safe by design approaches, work place safety, risk assessment, collaboration with European Union (EU) safety cluster and life cycle assessment. An established roadmap at the end of project enables the identification of future development needs in related fields order to allow Europe also in the future to compete at the forefront of the additive manufacturing revolution. In Section II, the eight main objectives of the project are described. A status of the project is given in Section III. At the end, in Section IV, the future strategy is explained.

## II. THE EIGHT DIMAP OBJECTIVES

The DIMAP project intends to implement an idea born among various industrial and research centers across Europe. The overall objective is to enhance digital materials with novel nanoparticles for 3D Printing. Four different inks are investigated, electrically conductive inks, ceramic inks, high strength polymeric inks and lightweight polymeric inks. The idea is to implement those materials into two concrete demonstrators (luminaires and bio inspires robotic joints) requiring specific processes. In parallel to the ink and process developments, safe by design and work place safety approaches are conducted in order to minimize the risk due to nanoparticles use. The whole concept can be differentiated into 8 objectives, summarized in Figure 1.

### *A. Objective 1: Additive manufactured joints*

Environmental interaction, the so-called man-machine interaction, is a great challenge for mechanical devices and modern robot design. Currently, research suggests “low impedance” as key strategy and an approach is provided by bionics [5]. Flexibility and low weight are two main aspects of reducing impedance that are perfectly realized in biological systems. 3D Printing technologies offer freedom of design to manufacturing and therefore are predestined for

 ISBN: 978-1-61208-496-1

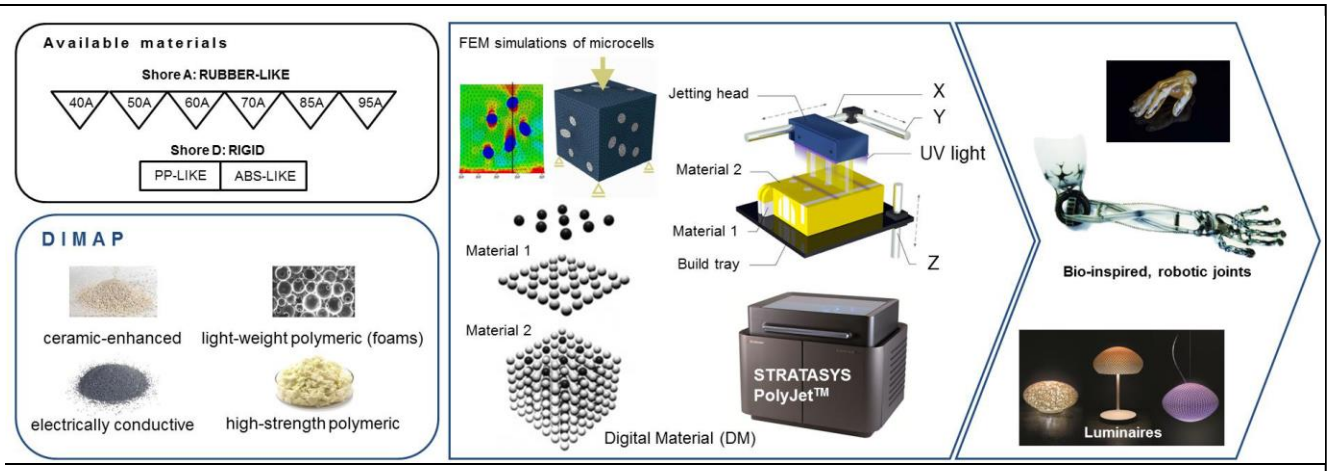



Figure 1. Concept for novel nanoparticle enhanced Digtial Materials for 3D Printing and their application shown for the robotic and electronic industry

the production of bio-inspired solutions, which are the prime example for such designs. Significant limitations in AM technologies are found in available materials, in particular functional materials to solve above-mentioned technical questions. Within DIMAP the manufacturing of bio-inspired joints, solutions derived from arthropod "jointed legs", will show novel solutions in robot design and man-machine interaction. These innovations will be reached with the development of aforementioned functional materials and technical improvements to the 3D-printer, e.g., combined thermal and Ultraviolet (UV) curing strategies, to be able to print the novel materials.

### *B. Objective 2: Additive manufactured luminaires*

One of the main drawbacks in the development of lightning applications is the lead time, which can take up to 6 months. This is due to the use of injection molding that also adds high costs for the required tooling (up to 1.5 million Euros) and limits the applicability to customized small series production. The implementation of AM is expected to increase the ability of companies to introduce new and improved products at accelerated rates as well as addressing the market for customization due to highly value added product diversity. Within DIMAP the material portfolio for the PolyJet Printing will be expanded to achieve higher mechanical strength, electrical conductivity and flexibility in order to be able to manufacture a lighting application with movable parts for a user-controllable light orientation. DIMAP will show that the new developed ink systems and improvements to the printing process are able to lead the way to manufacturing small to medium series of customized luminaires.

### *C. Objective 3: Ceramic enhanced materials*

The addition of ceramic particles allows the addition of new functionalities in products. The mechanical, electrical or magnetic properties of the resulting polymer matrix after solidification can be tailored. In the last couple of years, ceramic filled photo-curable inks have been used in stereolithography (SLA). Here solid loading up top 50 Vol% have been reported [6]. Based on the structuring method higher viscosities in the range of several 1000 mPas are acceptable in this case. In order to use the advantage of creating digital materials with PolyJet Printing the requirements in terms of viscosity are different: an ink viscosity below 100 mPas is crucial. However, high nanoparticle loading is necessary to enable huge property adjustments. Within DIMAP this will be achieved by using nanoparticles not smaller than 100 nm as smaller nanoparticles come along with larger specific surface areas leading to a significant viscosity increase. In addition higher solid loadings (up to 50 Vol% without solvent) retaining low viscosity values should be achieved via selection of suitable additives such as liquefiers and dispersants.

### *D. Objective 4: Electrically conducting materials*

Nowadays, electric conductors are made in industrial environment mainly by screen printing or photolithography which carry severe environmental concerns due to the low metal ink yield and the large quantity of hazardous waste as well as limited applicability for customization. Electrically conductive inks for digital inkjet conductive printing are a challenge as its necessary to balance several counter influence properties, such as: high metal content, low viscosity, small non agglomerated particles, surface tension properties compatible with the print heads and substrates. The only viable way to achieve this balance of properties is by formulating inks with nano-metal particles. Such inks for digital conductive 3D Printing are currently not available on the market. Within DIMAP UV curable inks will be developed which address these issues - printed paths containing a monomer will be cured (polymerized) while at the same time keeping suitable conductive properties (e.g., resistivity below 10 μΩ·cm). These goals will be achieved by a selection of suitable monomers, optionally solvents and photo initiators in metallic nanoparticle ink formulation.

 ISBN: 978-1-61208-496-1 

### E. Objective 5: High strength polymeric materials

Since both demonstrators may be exposed to mechanical stresses a high strength material will be part of the material development. Also, a high temperature tolerance is favorable, especially in case of the lighting device. Therefore, a new thermoplastic material will be added to the portfolio of PolyJet materials, opening up new possibilities in functionality and mechanical strength. Strategies to realize 3D Printing of these materials are based on the preparation of highly filled 3D-inks made of soluble polymer particles or by developing reactive 3D-inks based on precursors.

### F. Objective 6: Light weight polymeric materials

Various polymers can be designed to uptake high mechanical forces comparable to metal, glass and ceramics used in industrial application. A further reduction of weight from the solid polymeric object can be achieved with foams and foam-like structures. The strength of hard foams is found in the combination of different positive properties e.g., their light-weight structure and at the same time possessing high mechanical load capacity. Currently, 3D Printing of foam-like structures requires several working steps, e.g., Fused Deposition Modelling (FDM) printing and adjacent leaching in water for several days, obtaining soft foam-like objects. 3D Printing of foams with PolyJet technologies has not been realized yet. Within DIMAP printing of hard foams in one production step will be implemented to PolyJet 3D Printing. Thus, especially designed 3D-printable ink based on foamable core-shell particles and enhancements concerning the PolyJet-3D-printer will be developed. Printing of hard, closed foams will enable several new possibilities in AM including the print of light-weight, highly mechanically robust objects or the use of the closed foam as supporting material that allows printing on top of "hollow" objects, e.g., the top-surface of a 3D-printed cube.

### G. Objective 7: Novel multi-material 3D-printer

The PolyJet technology is based on a multi-jet and multi-head inkjet technology where objects are constructed layer-by-layer. PolyJets' uniqueness among other AM technologies is its ability to control the material deposition for building an object to the level of a single voxel allowing a precise control of the structure and of the properties of object printed. In order to address new markets widening the range of available materials and implementation of functions into the printed objects is crucial. As shown on the two applications this will allow SMEs to manufacture functional objects with tabletop multimaterials printers. In addition to the novel materials mentioned before a further development of the technology in order to cope with the ink requirements is needed. DIMAP will perform investigations and integration of novel printhead systems for higher ink viscosities (100 mPas at printing temperature) including ink circulation systems. Furthermore, an inline control of the printing process is needed to investigate the quality of the joint between already hardened drops and new deposited drops to be able to predict the toughness of the final 3D printed product. Thermographic flash method is known as a proper method for the investigation of the mechanical consistency of manufactured parts. DIMAP will investigate the appropriateness of this method for PolyJet 3D Printing.

### H. Objective 8: Safe by design

Nanotechnology is a key enabling technology for the development of innovative products of a large variety of industries. Impact and interaction of nanomaterials on environment and human health is a widely discussed question. Considering ink-formulations as proposed within DIMAP the handling of the base materials prior to ink formulation and excess materials after the printing process are the main challenges. In DIMAP a concern-driven guidance for investigating potential risks of engineered nanomaterials (ENM) will be established according to the NanoSafetyClusters' Research strategy 2015 – 2025. This allows a focused research on exposure levels and exposure routes, material properties, various in silico models as well as hazard and biokinetic data [7, 8]. In order to reach a Safe-by-Design approach for the process chains proposed within DIMAP assessment of the exposure rates during the manufacturing processes, usage time and end of life will be carried out. Furthermore, the implementation of a highly efficient risk management system will be targeted. Based on the data of existing databases and studies the process chain and materials will be evaluated and potential hazards will be identified. In combination with assessment of exposure rates mentioned above a strategy for the control and reduction of exposure rates will be proposed to reach a less hazardous product life cycle. The information gathered within DIMAP will be shared with the scientific community to further improve the knowledge base.

## III. CURRENT STATUS

During the first six months of the project, the main focus of the consortium was on the assessment of requirements for the demonstrators, materials and processes that will be used throughout the runtime of DIMAP. The base information needed for the developments within the project were the envisaged properties of the robotic arm and the luminaire system. Therefore, the main mechanical, electrical and thermal properties of the products themselves as well as the operating conditions and the limitations given by the current state of the PolyJet technique were collected. Based on these assumptions the development for the novel material classes and improvement of the printing technology is currently performed.

In close cooperation with a powder supplier (*TECNAN S.L.*)[9] the development of ceramic filled inks at the *Karlsruhe Institute of Technology (KIT)* is currently focusing on selecting the most suitable components. Ceramic nanoparticles for mechanical stability ($Al_2O_3$ or $ZrO_2$) or thermal conductivity (AlN or BN) are currently customized in order to adapt the particle size distribution and surface properties. Additionally, a wide variety of

 ISBN: 978-1-61208-496-1 

binder materials and surfactants will be examined for their suitability to reduce the ink viscosity to a processable level.

*PVNanoCell Ltd*. [10] develops conductive materials, mainly conductive inks for digital conductive printing based on silver nanoparticles. Silver-based inks will be used for printing conducting paths that are required in both applications. Currently, the main focus of the research is on the composition of the ink, as well as on the stability of the ink, reducing the percolation of the dense silver particles. In parallel, first printing test will be performed in the near future using the first promising ink compositions available.

As described above, the use of a new class of thermoplastics as a high strength polymer is a completely new approach in context of the PolyJet technology. Therefore, investigations at the *Soreq* [11] research center had to start from scratch by determining the most promising approach to create a suitable ink. Two different routes were evaluated, one using pre-polymer powders and the second by cross-linking low-molecular weight oligomers. At the current state, the second approach seems to show better compliance to the desired properties, which is why this route will be followed.

The last material is the foam-like material for the bone structure. At the *Johannes-Kepler University of Linz* and *Borealis AG* two different types of foams will be analyzed. The first type is a closed cell foam based on the principle of blowing agents that are dispersed in a polymer matrix and that forms voids with a outer shell upon UV- or heat treatment. The second type is an open cell foam that is produced by using organic blowing agents within the ink matrix.

All developments on the ink systems are carried out in close cooperation with *TIGER Coatings GmbH & Co KG* [12], who are focusing their work on the composition of the ink matrix, and *Profactor GmbH*, that are responsible for the ink curing strategies, that have to be individually adapted for each material. Finally, the practical test and printer development is done by *Stratasys Ltd* [13]. since some materials may not be fitting to the current PolyJet standards and may require new generation print heads or combined dispensing techniques.

## IV. CONCLUSION AND FUTURE WORK

DIMAP will provide a new range of materials as well as significant improvements concerning the dispensing technique to the PolyJet Printing within the runtime of the project. Two exemplary applications will be produced to show the capabilities of the development of the process chain. Although the project has just been running for 6 months, significant steps have already been made by defining the material and product properties in detail and first investigations in the field of ink development with very promising results so far.

The next steps are the development of suitable ink formulations using the designated nano materials defined in the previous months, ink stability testing and, in some cases, also first printing and modeling trials using the newly developed materials.

The printing technology has to be adapted for the expected higher values of viscosity. Therefore, the print heads used in the inkjet printer will be changed to upgrade the capability for higher viscous inks wherever it is needed. In addition, a possible combination with other dispensing techniques is currently thought of, depending on the properties of the final inks.

For the safety aspect, personal monitoring devices that measure workplace exposures will be implemented in the ink production processes to determine the possible influence of nano materials in the workplace safety.

## REFERENCES

[1] http://www.stratasys.com/3d-printers/technologies/polyjet-technology [accessed June 2016]

[2] T. Hanemann, D.V. Szabó, "Polymer-Nanoparticle Composites: From Synthesis to Modern Applications", Materials, 3, pp. 3468-3517 (2010), doi: 10.3390/ma3063468.

[3] R. Zichner, E. Sowade, R. R. Baumann, "Inkjet printed WLAN antenna for an application in smartphones", Jpn. J. Appl. Phys. 2014, 53, 05HB06.

[4] http://www.stratasys.com/fr/imprimantes-3d/technologies/polyjet-technology

[5] S. P. Buerger, "Stable, high-force, low-impedance robotic actuators for human-interactive machines", Massachusetts Institute of Technology, 2005.

[6] C. Provin, S. Monneret, H. Le Gall and S. Corbel, "Three-Dimensional Ceramic Microcomponents Made Using Microstereolithography"Adv. Mater. 2003, 15, pp. 994–997.

[7] A. G. Oomen et al., " Concern-driven integrated approaches to nanomaterial testing and assessment – report of the NanoSafety Cluster Working Group 10", Nanotoxicology 2014, 8, pp. 334–348, doi: 10.3109/17435390.2013.802387

[8] V. Stone et al., "ITS-NANO – Prioritising nanosafety research to develop a stakeholder driven intelligent testing strategy", Particle and fibre toxicology, 2014, 11, 9, doi: 10.1186/1743-8977-11-9

[9] https://tecnan-nanomat.es/ [accessed June 2016]

[10] http://www.pvnanocell.com/ [accessed June 2016]

[11] http://soreq.gov.il [accessed June 2016]

[12] http://www.tiger-coatings.com/ [accessed June 2016]

[13] www.stratasys.com [accessed June 2016]

 ISBN: 978-1-61208-496-1 

# Requirements for 3D Printed Applications using Novel Nanoparticle Enhanced Digital Materials

3D printed robotic and electronic applications

Adrien Brunet, Tobias Müller, Steffen Scholz
Institute for Applied Computer Science
Karlsruhe Institute of Technology
Karlsruhe, Germany
E-Mail:
[adrien.brunet;tobias.mueller2;steffen.scholz]@kit.edu

Anita Fuchsbauer
Profactor GmbH
Steyr-Gleink, Austria
Email: anita.fuchsbauer@profactor.at

***Abstract*— The DIMAP idea focuses on the development of innovative applications using additive manufacturing technologies. This paper describes the requirements for a 3D printed robotic arm and a luminaire application. In addition, the advances on novel ink materials for 3D multi-material printing by PolyJet technology are reported as well as advances on the printing equipment. The approach is application-driven and the first results of the project show advances in additive manufacturing technologies.**

***Keywords: Digital Material, 3D Printing, Robotic, Electronic, Nanoparticule***

## I. INTRODUCTION ON DIMAP CONCEPT

The DIMAP (novel nanoparticle enhanced DIgital MAterials for 3D Printing) [1] project intends to implement an idea born among various industrial and research centres across Europe. DIMAP aims to develop applications not only limited to rapid prototyping but that address production processes. Two applications of high interest are being developed to demonstrate the printing feasibility of functional inks. The first application is a bio-inspired robotic arm developed by Festo [2] and Cirp [3], the second is a shape changing luminaire developed by Philips [4]. The challenges created by applications, materials and printing processes requirement interdependencies are numerous and complex. Namely, the development of novel ink systems with incorporated nanoparticles is especially challenging. In order to cope with these new material classes, the existing PolyJet technology is further developed and therefore improved.

The overall objective of the project is to enhance digital materials with novel nanoparticles for 3D Printing in order to increase design possibilities. Indeed, developing robots poses particular challenges in terms of design [5]. DIMAP proposes to learn from nature to create bio-inspired robotic joints and it appears that the additive manufacturing (AM) provides a suitable basis to mimic this approach [6]. As well, additive manufacturing enables a high-customized production. It has been demonstrated that customers perceive customized luminaires as high value products compare to standard products [7,8]. In order to develop those innovative added-value products, four different inks are investigated: electrically conductive inks, ceramic inks, high strength polymeric inks and lightweight polymeric inks. In parallel to the ink and process developments, safe by design and work place safety approaches are conducted in order to minimize the risk due to nanoparticles use. The

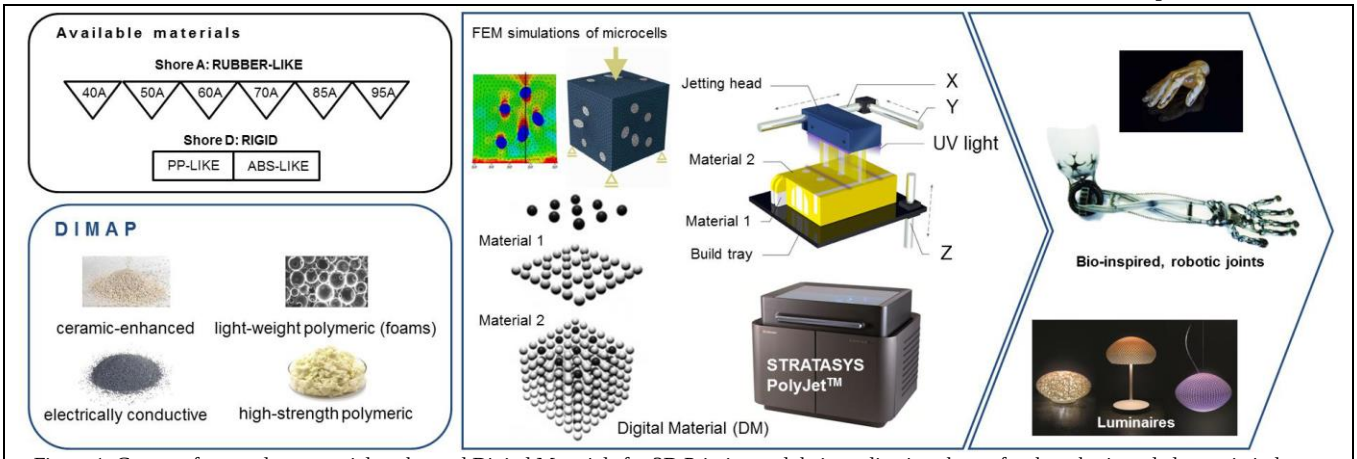



Figure 1. Concept for novel nanoparticle enhanced Digital Materials for 3D Printing and their application shown for the robotic and electronic industry

 ISBN: 978-1-61208-496-1 

whole can be differentiated into 8 objectives [1,9]. Those objectives (four novel digital material developments, novel multi-material 3D printer, safe by design approach and two innovative demonstrators) are summarized in Figure 1. Within this paper, the requirements elaborated during the starting phase of the project and the challenges in material development are presented. The paper is structured as follows: In Section II, the designs and the concepts of the 3D printed robotic arm and luminaire are described. In Section III, the faced hurdles and the envisaged solutions are discussed. Section IV concludes the paper.

## II. DESIGN SPECIFICATION AND MATERIAL REQUIREMENTS FOR ROBOTIC ARM AND LUMINAIRE DEMONSTRATOR

### A. *Additive manufactured robotic arm*

DIMAP intends to 3D print bio inspired robotic joints. Man-machine interaction is a great challenge for mechanical devices and modern robot design. Low impedance and high force-to-weight ratio is currently seen as a key approach for bionics [10]. Flexibility and low weight are two *sine qua non* aspects of reducing impedance. 3D printing technologies offer freedom of design to manufacturing and therefore are predestined for the production of bio-inspired solutions. Nevertheless, each AM-technology currently available is limited within the available materials. Typically, only a limited number of polymer types are applicable when compared to other manufacturing techniques such as injection moulding or hot pressing. On the other hand, specialized technologies like selective laser sintering (SLS) or –melting (SLM) are capable of producing metal parts, but are heavily dependent on the quality and reproducibility of the base material/powder.

DIMAP application requirements create the basis to solve those technical limitations. The chosen solution is derived from Arthropod "jointed legs" and require development of functional materials (with high strength, light-weight, dielectric, magnetism or conductive properties), as well as technical improvements to the 3D-printer (combining thermal and Ultraviolet (UV) curing strategies).

The first draft of the joint (i.e., Figure 2) consists of two bellows surrounding the actual hinge, operated by compressed air. The current design features a material combination of a rubbery, flexible and deformable material on the one hand and a rigid material on the other hand. Mechanical properties mainly characterized the two materials. The second main part is the hinge, which is proposed to resemble to a human elbow joint. Therefore, a hard material is necessary to withstand the tribological stress generated by the movement of the connected elements without wearing out too fast. The last element is the arm structure; it should be made of a rigid, yet light material in order to reduce weight when compared with current designs being produced with other manufacturing techniques. Therefore, a combination of a hard shell filled with a porous (or foam-like) structure is considered. Finally, conductive tracts are intended to be printed on the structure to connect sensors (pressure, position, etc.). The length of the tracks is expected to be superior to 300 mm, which implies new challenges.

The development of such application to be 3D printed requires to conduct research activities on the material side as well as on the printer itself.

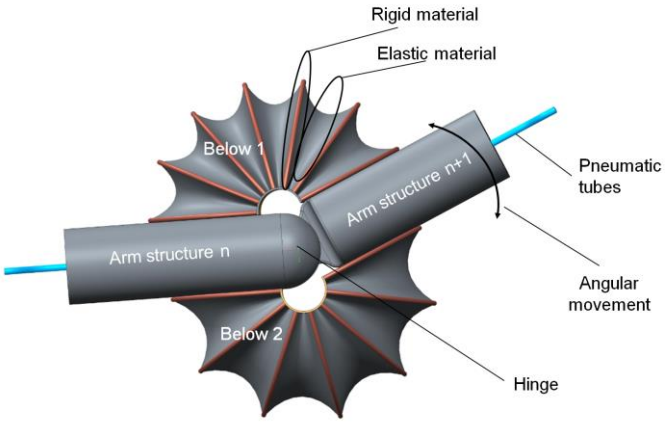



Figure 2: Conceptual drawing of robotic joint actuator

### B. *Additive manufactured luminaires*

In lightning applications, the lead times from design to available finished products is long even for low volume productions (up to 6 months). This is based on the currently used injection moulding technique, which in addition is rather costly. Depending on the complexity of part and mould, the design and manufacturing of the mould can cost up to 1.5 million Euros. The implementation of AM is expected to increase the ability of companies to introduce new and improved products at accelerated rates. The profession retail lighting market is an interesting market for customization due to highly value added product diversity. There is a variety of different levels, on which one can customize parts and products using novel digital materials. Customization can happen at the material level, on the part level and finally on the modular level. In terms of mechanical strength and electrical conductivity, material requirements for this application have to be fulfilled, in order to accelerate the transition from mere prototyping towards production. DIMAP will show that the new developed ink systems lead the way to additive manufacturing of customized luminaires.

To that end, DIMAP will develop a luminaire demonstrator. Figure 3 shows a sketch of this demonstrator developed by Philips. It is a linear array of LED distant of 100 mm from each other. One pattern of the array is composed of one LED and of surrounding moveable elements, shaped as a V and dedicated to direct the light. The parts of the V facing the LED are reflective in order to minimize light losses. The movement is performed by pressurized air coming through a lower channel.

 ISBN: 978-1-61208-496-1 

The approach chosen within DIMAP is to separate the development in two parts. The first one is the Printed Circuit Board (PCB) receiving the LED. The PCB part is made of a thermally conductive layer, used as a substrate for the rest of the structure and as a heat spreader. The second layer is made of a dielectric material to avoid sparking and electric bridging. On top of it, the electrical pattern is printed with conductive ink. A driver and an optical source (the LED itself) are mounted in further steps. Finally, a highly reflective layer (printed or coated) will cover the PCB.

The second part is the LED array; its function is to control the light effect, changing the angle of the V. The main challenge in this part is to print simultaneously hard and soft material. Indeed the reflective part has to maintain its shape and therefore high-strength ink should be used.

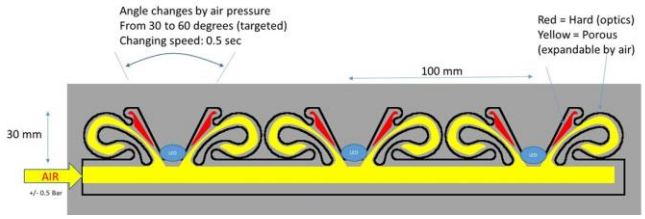



Figure 3: Concept sketch of hard/soft combination activated by pressurized air

## III. HURDLES AND SOLUTIONS

Additive manufacturing techniques offer an unchallenged design freedom. Nevertheless, the material variety is limited. One of the main challenge faced by DIMAP concerns the ceramic inks being researched in Karlsruhe Institute of Technology in the group of Prof. Thomas Hanemann [11]. The ceramic inks developed in DIMAP are partially dedicated to the Printed Circuit Board on which the Light-Emitting Diode is printed. One of the main challenges for PCBs is to dissipate the heat generated by electrical current and to limit the material deformation induced by this heat. Therefore the ceramic part has to perform high heat dissipation (10 – 50 $W.mK^{-1}$) and small linear thermal expansion coefficient (inferior to 10-6 $K^{-1}$). If expansion is too high, it can lead to solder fatigue and cracks in conductive tracks.

Ceramic inks developed by DIMAP are actually based on polymer ink. Within a certain range, the physical properties of resins and polymer can be adjusted by the addition of organic or inorganic nano-sized fillers [12-14]. The addition of spherical ceramic particles with average particle size larger than several micrometres normally deteriorates the resulting mechanical properties, while ceramic fibres enhance them. The addition of ceramic nanoparticles causes an enhancement of the thermomechanical properties due the particle's very large surface area generating an interfacial layer with pronounced attractive forces [15-17].

As explained in the previous paragraph, the addition of nanofiller can be used to tune thermomechanical properties of polymer-based ink. The specifications given by Philips and Festo imply a high load that will increase the ink viscosity significantly. Now, the ink viscosity at jetting temperature (60-90°C) excides the 20 mPas recommended by Stratasys to use the actual PolyJet technology. The solutions to overcome this problem heads towards a new generation printhead accepting higher viscosity or a hybrid solution including existing technologies (i.e advanced ink-jet (e.g., SIJ Technology), air pressure-based multi-nozzle dispenser, Aerosol-type (e.g., Optomec) [18], syringe-type dispenser). Due to their high hardness, ceramic materials are also considered for printing the hard-shelve structure of the robotic arm hinge. Since materials like $ZrO_2$ and $Al_2O_3$ are also brittle and relatively heavy, DIMAP proposes to fill the ceramic hard-shelve with a foam-like material. The purpose of a foam-like material is to reinforce the ceramic structure and maintain an overall low-weight. For example, solid PS has a density of 1050 kg m-3, but expanded PS can have a density as low as 15 kg m-3 [19-21]. The development of such inks is challenging and therefore different approaches are envisaged. The first approach developed by the University Johannes Kepler of Linz [22], is to include in polymer based inks Microspheres. The microspheres are made of a shell-polymer and of a core blowing agent. Under heat or UV exposure, the core material blows, resulting in expanded microspheres with thinner polymer shell. To maximize the chance of developing foam-like ink, an open-cell foam approach is also considered. Unlike the microsphere approach, this does not require synthesis of polymeric material. The blowing agent is ground down to an acceptable size and uniformly dispersed in the matrix ink. Depending if the compound is organic or inorganic, the agent will be blown using heat or UV exposure. During the expansion process, the bubbles can freely merge creating a porous foam and gas permeable structure.

## IV. CONCLUSION

New printing techniques allow the creation of so-called digital materials with which multiple material combinations and novel composites with predictable physical properties can be produced. As explained in the previous parts, DIMAP intends to use the design freedom given by AM and the PolyJet technology to produce demonstrators with innovative design and exceptional mechanical properties. Different material properties can be achieved and used to implement functional materials to the portfolio of printable materials, either by adding nanoparticles or by adjusting the polymeric ink composition and the curing processes;.

In the upcoming months the main focus of the project will be on the development of suitable ink formulations with incorporated nanoparticles. In addition, the ink stability and curing strategies will be developed and first printing tests will be performed.

Since the filler materials raise the viscosity of the inks, an adaptation of the printing technology will be mandatory. Upgraded print heads that are able to cope with the new

 ISBN: 978-1-61208-496-1 

materials are one way this will be achieved during the runtime of DIMAP. A different approach could be a combination of different dispensing techniques. Depending on the final ink properties a appropriate solution will be implemented.

With regards to the safe-by-design approach and safety aspects, each material will be evaluated and a Safety Data Sheet (SDS) will be developed. Personal monitoring devices will be used during the material preparation and printing steps to analyze exposure to nanoparticles in these stages of the whole process. Also, a life cycle assessment will be carried out to determine the impact of the printed product throughout the whole production, usage and disposal span.

## REFERENCES


[1] http://www.dimap-project.eu/ [accessed June 2016]

[2] www.festo.com [accessed June 2016]

[3] www.cirp.de [accessed June 2016]

[4] www.philips.com [accessed June 2016]

[5] R. Neumann, A. Hildebrandt and E. M. Knubben (Eds.) "Fluid Power in Motion", 2008.

[6] K. Liu and L. Jiang, "Multifunctional Integration: From Biological to Bio-Inspired Materials", ACS nano 2011, 5(9), pp. 6786–6790.

[7] B. Soars, "Driving sales through shoppers' sense of sound, sight, smell and touch", Intl J of Retail & Distrib Mgt 2009, 37(3), pp. 286–298

[8] D. Hinks and R. Shamey, "Review of Retail Store Lighting: Implications for Colour Control of Products", Coloration Technology, 2011, 127, pp. 121–128.

[9] S. Scholz, A. Brunet., T. Müller and A. Fuchsbauer, "Novel Nanoparticle Enhanced Digital Materials for 3D Printing and their Application Shown for the Robotic and Electronic Industry", proceding of the CENICS 2016 conference.

[10] S.P. Buerger, Stable, high-force, "low-impedance robotic actuator for human-interactive machines", Ph.D. Thesis, Massachusetts Institute of Technology, Dept. of Mechanical Engineering, 2005.

[11] T. Hanemann, W. Bauer, R. Knitter and P. Woias, "Rapid Prototyping and Rapid Tooling Techniques for the Manufacturing of Silicon, Polymer, Metal and Ceramic Microdevices", in: "MEMS/NEMS" (Ed.: C. Leondes), Springer US, 2006, pp. 801-869, ISBN: 978-0-387-25786-0

[12] T. Hanemann and D.V. Szabó, "Polymer-Nanoparticle Composites: From Synthesis to Modern Applications", Materials, 2010, 3, pp. 3468-3517, doi: 10.3390/ma3063468.

[13] S. Schlabach, R. Ochs and T. Hanemann, D.V. Szabo, "Nanoparticles in polymer-matrix composites", Microsystem Technologies, 2011, 17, 183-193, doi: 10.1007/s00542-010-1176-8.

[14] D.V. Szabó and T. Hanemann, "Polymer nanocomposites for optical applications", In Advances in Polymer Nanocomposites: Types and Applications; Ed. F. Gao, Woodhead Publ., Oxford, 2012, pp. 567-604, ISBN 978-1-8456-9940-6

[15] L.S. Schadler, L. Brinson and W. Sawyer, "Polymer nanocomposites: A small part of the story", JOM Journal of the Minerals, Metals and Materials Society, 2007, 3, pp. 53-60, http://dx.doi.org/10.1007/s11837-007-0040-5.

[16] L.S. Schadler, S.K. Kumar, B.C. Benicewicz, S.L. Lewis and S. E. Harton, "Designed Interfaces in Polymer Nanocomposites: A Fundamental Viewpoin"t, MRS Bulletin, 2007, 32, pp. 335-340.

[17] W.R. Caseri, "Nanocomposites of polymers and inorganic particles: preparation, structure and properties", Materials Science and Technology, 2006, 22(7), pp. 807-817.

[18] http://www.optomec.com/ [accessed June 2016]

[19] H. Weber, I. Grave, E. Röhrl, "Foamed Plastics", in: Ullmann's encyclopedia of industrial chemistry, 2010, Wiley, Chichester.

[20] Handbook of polymer foams, Ed.: Eaves, D., 2004, Rapra Technology, Shawbury, Shrewsbury, Shropshire, U.K.

[21] J. Maul, B. G. Frushour, J. R. Kontoff, H. Eichenauer, K.-H. Ott, C. Schade, "Polystyrene and Styrene Copolymers", In: Ullmann's encyclopedia of industrial chemistry, 2010, Wiley, Chichester.

[22] http://www.jku.at/ [accessed June 2016]


 ISBN: 978-1-61208-496-1 

# ADDMANU – An Austrian Lighthouse Project for Additive Manfacturing

Christian Wögerer
PROFACTOR GmbH
Vienna, Austria
christian.woegerer@profactor.at

Michael Mühlberger / Markus Ikeda
PROFACTOR GmbH
Steyr – Gleink, Austria
Michael.muehlberger@profactor.at
Markus.ikeda@profactor.at

***Abstract*—** AddManu **is a national Austrian flagship project for research, development and the establishment of additive manufacturing. There are four topics defined as key technologies for additive manufacturing: Lithography based manufacturing (LBF), Fused Filament Fabrication (FFF), Laser beam melting (LBM) and the InkJet printing. These have, from today's perspective, the highest potential for applications and further developments. The first results (process concept and first printing trails) of Inkjet printing of the project are presented in the paper.**

***Keywords: Additive manufacturing, 3D Printing, Lithography Based Manufacturing (LBF), Fused Filament Fabrication (FFF), Laser Beam Melting (LBM), InkJet printing***

## I. INTRODUCTION

To reach a significant progress beyond the state of the art in additive manufacturing, an extensive discussion of various technologies, which can be achieved only by a large consortium with various technology know-how, is necessary. This is the idea of the lead project AddManu.

The process of InkJet printing on 3D curved surfaces has different challenges which must be solved in the project (Materials, Process, Quality Control). The issue of appropriate materials as well as a precise knowledge of the process must be integrated in a suitable system. Each component of the system must work exactly with all others and must be optimized in hard- and software. The idea of robot-based InkJet printing is the key issue and it is described in this paper, together with the material, soft and hardware requirements. The purpose of the article is also the presentation of the first results after the concept phase.

The rest of the paper is structured as follows. Section II presents an overview of the consortium, the project goals and the motivation as well as the key facts of the project. Section III shows a general overview of the technical work, followed by a more detailed description of PROFACTOR´s work inside the consortium. This is focused on the development of basics for a multimaterial hybrid manufacturing technology based on InkJet printing. In addition, the concept of robot-based ink jet printing and material development is presented. Section IV provides an overview of the first results. Section V concludes the paper.

## II. THE ADDMANU PROJECT

### A. General project description

The Lead-project AddManu [1] will form a national research network with an international scientific board in order to find recognition and acceptance within the Austrian economy. Four AM (Additive Manufacturing) technologies are brought into focus (Lithography-based AM, Fused Filament Fabrication (FFF), InkJet and Selective Laser Melting), which have the largest potential for industrial application and further development (Figure 1). The most important families of engineering materials, i.e., ceramics, polymers and metals are included. Based on longtime expertise of consortium partners and intensive research work, the project will deal with those problems, which can be considered as barriers for further developments and economic use or which have a very high innovation potential. Within AddManu.at, the R&D-activities (Research & Development) are divided in four areas: materials development, design and dimensioning, process-specific and application-oriented aspects, each for metals and non-metals. Cross-sectional issues, like system integration are covered in a separate working package.

### B. Key facts

The work of AddManu was launched in May 2015 and will run until April 2018th. The projects involve 19 different partners from research and industry, which are the key players for additive manufacturing in Austria [5]. The partners are situated along the manufacturing value chain from basic research to industrial implementation, from material development, process development to manufacturing and product development.

The following industrial sectors were addressed:

- Materials/Surface Technologies
- Plant Construction
- Machine Building Industry and Engineering
- Automotive
- Aerospace Industries
- Research, Science and Education

 ISBN: 978-1-61208-496-1 

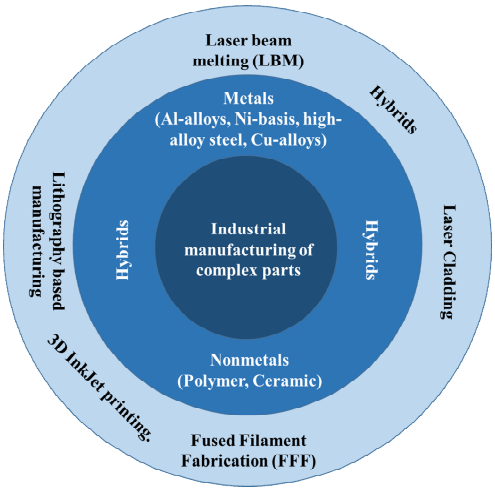



Figure 1. AddManu´s overall concept [1]

### C. *Objectives:*

The most important objectives are:

- Material developments for improved processing and service properties of AM-built components, like new powder materials and hybrids (composites, segmented structures, etc.)
- The innovation potential of AM-processes will primarily depend on the designer's creativity and the use of sophisticated FEM (Finite element method) [14] -software packages for light weight design. By adaption of methods like topology and shape optimization to AM-specific issues and coupling with extremely fine lattice structures, novel solutions are generated and new user markets can be generated.
- Process developments for AM-technologies, lithography-based AM, Fused Filament Fabrication and InkJet.
- The industrial implementation of novel AM-concepts within the fundamental R&D areas will be done in separate working packages, which are dedicated to the branches mechanical engineering, tooling, automotive engineering, semiconductor industry, refractory industry and aerospace industry. Solutions will be searched, which offer significant competitive advantages.

The most important deliverables and findings will be:

- Development of new materials (metal powders, ceramics and thermoplastic photopolymers) with significantly improved material properties.
- Development of an AM-concept to build hybrid components made of metal/ceramics, steel/aluminum
- Development of novel lithography-based AM-processes with significantly improved resolution and higher operational capacity.
- Development of post-processing-methods to improve the surface quality of AM-built products.
- Development of new industrial applications taking into account the whole processing chain.

### D. *Thesis of AddManu*

Success stories for additive manufacturing show us that a simple substitution of existing manufacturing technologies without increasing the complexity of part geometry or/and a better integration of the part functionality is not really productive.

Therefore the AddManu project is mainly based on four leading hypotheses: These four hypotheses determined the research agenda of the project, the objectives, the project structure and the consortium. These hypotheses are also aligned with the defined goals of the call and the definition of a "lead project" and with the industrial needs.

**HYP1 Freedom of Design:** Additive manufacturing allows manufacturing of complex parts with significantly less financial effort and with reduced equipment compared to conventional technologies. So, a cost neutral "Freedom of design" exists and leads to a higher integration of functionalities in parts or units. Examples are light weight constructions or manufacturing without assembly. One result of this is that, overtaking existing or classic design from the conventional manufacturing into additive manufacturing, a main benefit from additive manufacturing is not taken into account.

**HYP2 Flexibility of fabrication:** Additive manufacturing requires, in general, no tooling costs and also low set-up costs. The outcome of this is a massive advantage in logistics, so new business models for custom made parts are possible. An on-side production is of great interest again, and additive manufacturing is a key component for Europe's Re-Industrialization.

**HYP3 Rules of scaling:** The up-scaling of the part costs in additive manufacturing is directly proportional to the manufactured volume. Costs of additive manufactured parts are not really depending on lot sizes. This is in contrast to tooling based manufacturing (injection molding, forging). So, additive manufacturing has its advantage in manufacturing of high complex parts and small lot sizes.

**HYP4 Process integration and shortening the value chain:** The classic manufacturing chain consists of independent suppliers for material, design (CAD/CAM - (computer-aided design/manufacturing) and production (milling/CNC - Computerized Numerical Control - machines). Additive manufacturing benefits from fewer requirements as far as design for manufacturability is concerned, but knowledge from the whole manufacturing chain (material to production) is needed at almost every step in the manufacturing chain to be able to profit from potential benefits. Design, material and process are dependent on each other. New business models must face this fact from the beginning to be successful.

## III. PROFACTOR´S TASK

The project partner PROFACTOR [9] is inside AddManu Workpackage Leader and is developing a basis for a multi-material hybrid manufacturing technology based on InkJet

 ISBN: 978-1-61208-496-1 

printing. The challenge is that InkJet printing should be done on 3D curved surfaces of parts which are pre-manufactured with other processes or technologies developed in AddManu, i.e., an advanced FFF process. The goals of the tasks are divided in 4 different parts which enable a new process technology (Figure 2).

- Development of UV (ultra violet) curable Inks for multi-material hybrid manufacturing technologies
- Research on printing processes using InkJet printing on FFF pre-manufactured free-form parts.
- System and process development of free-form printing using a robot based Ink-Jet printing system.
- Development of a non-destructive quality control system using machine vision.

The general approach of the new multi-material hybrid manufacturing process is that additive technologies promise new functionality.

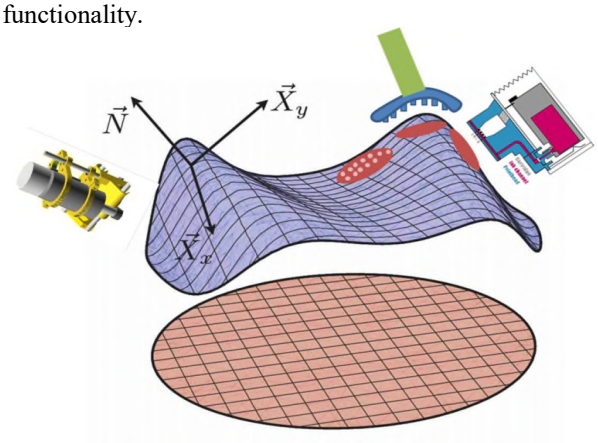



Figure 2. InkJet printing on curved sufaces with robots and inspection systems

To design a pilot system, several challenges must be solved. These challenges include: InkJet printing to apply the imprint material on the right spot (Figure 3) and 3D machine vision [2] (Figure 4) in combination with robotics [3][4] (Figure 5) to position the print head above the 3D-printed surface.

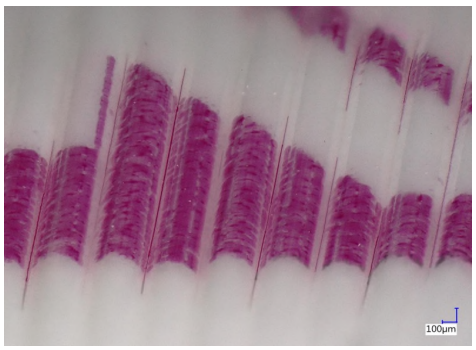



Figure 3. Optical micrograph of InkJet printed UV-curable ink on a 3D-printed surface.

The curvature resulting from the 3D-printing process can clearly be seen in Figure 3.

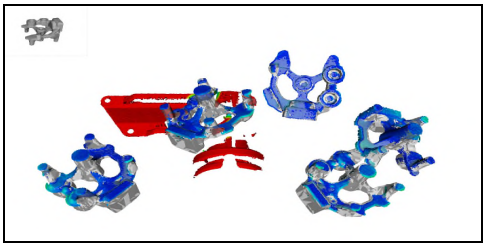


Figure 4. Illustration of the use of the Candelor software library to identify objects (grey/blue) within a 3D pointcloud (red). [2]


Figure 5. Photograph of a robot while measuring

A detailed work plan with different steps and milestones is defined to reach the goal for a prototype installation.

### *A. Description of activities*

The first step is the development of functional special inks which optimize the 3D surface of an FDM manufactured part with respect to chemical/mechanical stability and water tightness printable on 3D FFF surfaces. Additional to this, a second special Ink is developed. This Ink is for surfaces with special haptics (rubber like).

These special Inks will be printed by a robot-controlled Ink Jet head on 3D free form surfaces.

The robot assisted 3D InkJet system and the necessary process planning software respecting the critical parameters (printing distance, angle of printing head to the vertical, printing velocity and angle to the substrate) and a collision free path planning will be designed and developed. Enabling a constant coating (printing) speed flow rate and quantity will be coordinated and highly dynamic.

Recognizing the topography of the substrate to be coated and for determination and controlling of critical parameters it is necessary to implement a 3D-control system. This could be also used as quality control system. Surface roughness and other fails are registered, rated and corrected if necessary.

The goal is a proof of concept demonstrator for a multilayer printing of FFF manufactured 3D printed freeform parts using InkJet printing. It will be only printed on the areas where it is necessary for correction of geometry or functionalisation of the part. The technology to be developed is based on existing PROFACTOR Knowhow [10].

### *B. Method:*

According to concept descript above, special InkJet Inks for multi-material hybrid-manufacturing will be developed.

 ISBN: 978-1-61208-496-1 

A test bed for robot assisted 3D InkJet Printing will be built in parallel and the necessary software for the control system developed. Developed InkJet Inks will be tested and evaluated against their printing behavior and their functionality. For identification of the part topography an adequate 3D control system will be designed and implemented.

### C. *Deliverables*

- Functional special Inks with optimised properties (chemical/mechanical stability, water tightness) printable on 3D FFF surfaces.
- Functional special Inks with optimised haptics printable on 3D FFF surfaces.
- Robot controlled InkJet printing head with process control for printing on 3D free forms.
- Non-destructive quality control system based on machine vision.

## IV. FIRST RESULTS

The first results of the project are available at this time, after one year from the start of the project, and are represented mostly by theoretical/concept level results. The design of the robot based process is finished and an analysis of possible technologies for vision control systems was done. Additionally, the first experimental results of printing trials on curved surfaces are available.

### A. *Robot controled InkJet printing head with process control for free form printing*

The work was divided into software and hardware related research.

Hardware: The necessary hardware is the printing head, ink transport system, meniscus control system, drivers for the printing head with master control unit und slave control unit and the software for controlling the printing head and to integrate the head into the robot.

Parameters which were important for the decision of the printing head included only one supplier for all components and also an open software system. At the end, a Ricoh Gen 4 [11] printing head was chosen (Figure 6).

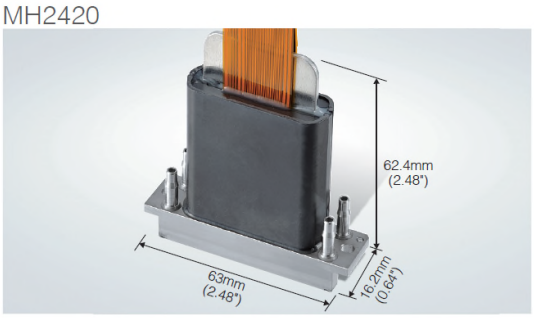



Figure 6. Ricoh Gen 4 MH2420 printing head

Reasons for the decision:

- Compact Dimension: In general, compact dimensions are an advantage for the integration on the robot. Additional it's a benefit for printing on curved surfaces if the printing area is not too broad. Other printing heads are broader, e.g., Dimatix Spectra 150mm, Xaar 501 125mm. Konica Minolta 512 printing head is match-able with 67mm wide (32,4mm printing wide). A smaller printing enables a better printed image over the total printed area, because the variation of the distance from printing head to the surface can be better controlled.
- Integrated heating: If we use a printing head with integrated heater, then inks which do not have a suitable viscosity at room temperature could also be used. It is not necessary to heat the Ink transport system and therefore the whole system could be easier integrated into the robot system.
- Dumping: The printing head could be also used in a dumped or sloped position for printing. This is an essential feature for printing on curved surfaces.
- Compatibility with project partner TIGER´s [8] Heavy Duty Ink: TIGER Inks were already tested with this printing head.

Mechanical design and construction: Based on an analyses of accessibility the most convenient robot position (Industrial Robot model Stäubli TX90L) [12] for a square working space (800x800mm) was determined (Figure 7).

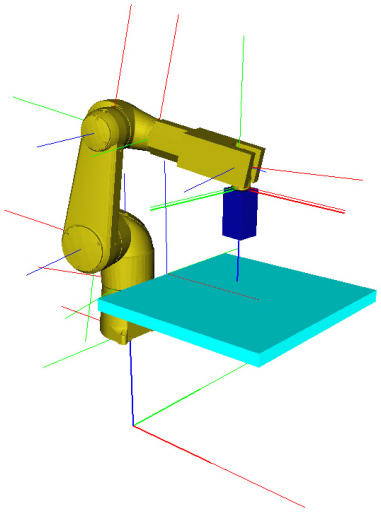


Figure 7. Simulation – Analyses of accessibility

Requirements to the tool holder are

- Automatic tooling system for an easy demounting of the printing head from the robot arm
- Mechanical fixtures for
  - Printing head (Ricoh Gen IV MH2420)
  - Driver board (Ardeje)
  - Ink-storage (2 nozzles with connecting tubes)
  - Automatic tooling system
- For safety reasons the robot system has to operate behind a mechanical disconnecting safety installation (fence)

 ISBN: 978-1-61208-496-1 

- A PC-working place is situated near the system
- The controller for the printing head must be situated close to the robot caused by limited length of wires (robot foot)

For the robot assisted InkJet printing process, the necessary system components are shown in Figure 8.

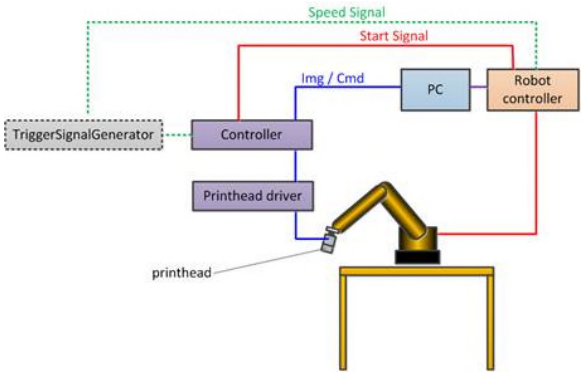



Figure 8. System overview of robot based printing system

A PC creates a robot program for guiding the printing head simultaneously with a command list for the (printing head) controller. This controller starts the real printing process. For synchronizing of single printing commands with the spatial position we cannot use path points, because the necessary high frequencies are not available. So we use between the sychronisation points a trigger signal which is generated by a frequency generator and depends on the instantaneous velocity.

Software: For the configuration and the activation of the control a firmware of the controller manufacturer is used. Figure 9 shows the suggested workflow for printing (disregarding the path planning for the robot).

Using the software module „Image composition“ a printing picture is created and prepared for the Printing Head controller by a special „Raster Image Processing“ Software - module.

In this step, vector-graphics are calculated into raster-graphics, resolutions are converted, color channels separated and color management processes executed. Precompiled information is sent to the controller. This controller needs afterwards only a start signal (Enable) and the “feed” signal (synchronized with the robot movement)

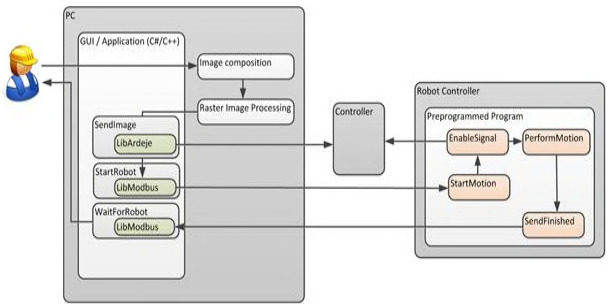



Figure 9. Work flow „priniting on an known path“

## B. *First printing trails with heavy duty ink on curved surfaces*

First InkJet printing trials were done on 3D printed surfaces (FFF) using a Heavy Dury Ink from TIGER. Used Substrate were FFF printed parts without any pre-treatment. The material was white PLA (Polylactides from Orbi-Tech) [13] which was printed with a HAGE 3Dp-A2 printer in the Labs of PROFACTOR. The thickness was, in most cases 0.4mm printed at 210°C. The geometry of the teste samples was a simply cuboid (75x25x10mm³, 30% infill). The side areas were used for the InkJet printing (Figure 10).

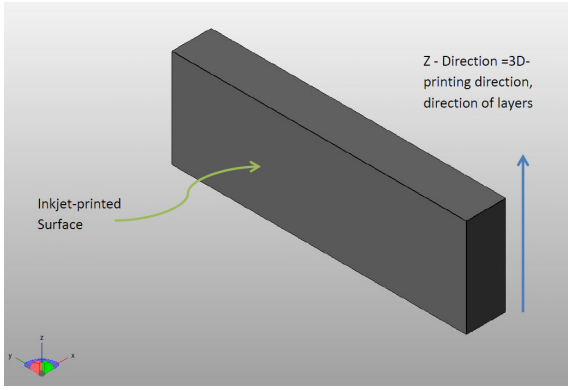



Figure 10. Explaination of test printing

In the first trails, the results are three important findings:

- Adhesion on PLA substrate is very good, even without pre-treatment of the substrate good results were reached.
- The printed image was strongly influenced by the capillary effects. These capillaries occur during the FFF process.
- Curing with UV-LEDS (Ultraviolet Light-emitting diode) (395nm) worked very well in a few seconds (also in Air).

Figures 11 -14 are showing the first test samples. Missing horizontal lines are caused by suboptimal parameters on the used Dimatix printer and are not dependent upon the 3D printed substrate.


Figure 11. Overview of printed sample, lenghth ~ 55mm

 ISBN: 978-1-61208-496-1


Figure 12. Details of printed sample

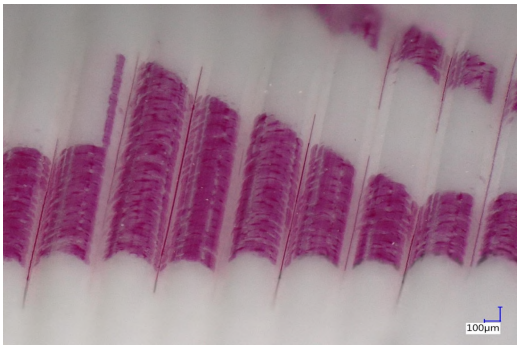



Figure 13. Angular view of sample, FFF Layers are visible well, luting caused by cappilary – effects

By choosing a greater distance between the drops the capillary effect could be reduced. It is better (for example) printing 2x50 µm drop distance than using 1x25µm. A smaller layer thickness (0.2mm instead 0.4mm) raised the possibility that drops are getting inside the small channels.

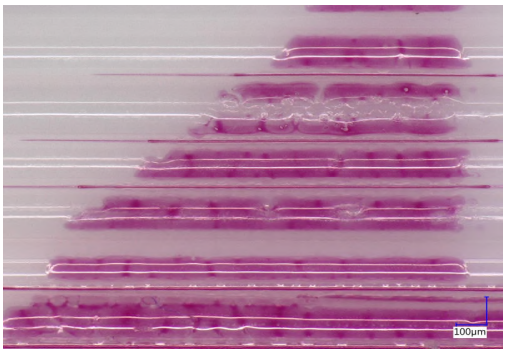



Figure 14. PLA 0.2 mm thickness-missing lines as a consequence of not using opmimized printer parameters

## V. CONCLUISION AND FURTHER WORK

### A. *Conclusion*

Multilayer printing on FFF manufactured parts needs a lot of processing and also material development to be successfully implemented and brought to Industry. The first steps of a proof of concept show promising results, but each component of the system has to be optimized on its own and in the system.

### B. *Further work and next planned steps*

Based on the studies and on the results of the first printing tests, the next steps will be:

- Installation and implementing of printing head and robot in the PROFACTOR Lab
- Ongoing printing tests and generation of parameters for further Ink Development and for advanced Robot set up
- Trails for Quality control with existing equipment of PROFACTOR and implementing the best fitting technology into the printing system
- Surveying test samples from other project partner to find a set of parameters for quality control
- Ongoing Haptic- Ink development by project partner TIGER and testing this new inks at the lab and with the robot system
- Optimisation of printing parameters and comparison of equal inks with different colours.

All steps will be done in a strong interaction with industry to have a feedback and a “closed loop” development which meets the industrial requirements.

## ACKNOWLEDGMENT

The project is funded by the Austrian Ministry for Transport, Innovation and Technology (BMVIT) [6]) and the Austrian Research Promotion Agency (FFG) [7]. It is a lighthouse project in the frame of the initiative “production of the future”. We also acknowledge the Company TIGER [8] coatings for supplying the UV-curable ink.

## REFERENCES

[1] AddManu: Avialable online at www.addmanu.at

[2] Candelor: Avialable online at http://candelor.com

[3] E. Weigl, W. Heidl, E. Lughofer, C. Eitzinger, and T. Radauer, On improving performance of surface inspection systems by on-line active learning and flexible classifier updates, Machine Vision and Applications, 27(1), 103-127, Jan. 2016, doi:10.1007/s00138-015-0731-9.

[4] E. Weigl, S. Zambal, M. Stöger, C. Eitzinger, Photometric stereo sensor for robot-assisted industrial quality inspection of coated composite material surfaces, QCAV 2015, SPIE Proceedings Volume 9534, 2015, doi:10.1117/12.2182750

[5] Austrias Key Players in Additive Manufacuring avialable online at http://www.addmanu.at/de/4886/ and http://www.addmanu.at/de/4885/

[6] BMVIT- Avialable online at www.bmvit.gv.at/en/index.html

[7] FFG – Avialable online at https://www.ffg.at/

[8] TIGER - Avialable online at http://www.tiger-coatings.com/

[9] PROFACTOR – Avialable online at www.profactor.at

[10] ANIPF project: Avialable online at: http://www.profactor.at/index.php?id=895&L=1

[11] Ricoh Gen 4: Avialable online at http://www.rpsa.ricoh.com/gen4.html

[12] Stäubli TX90L: Avialable online at: http://www.upc.edu/sct/documents_equipament/d_181_id-492.pdf

[13] Orbi-Tech: Avialable online at www.orbi-tech.de

[14] FEM Software: Avialabel online at: http://3dsim.com/

 ISBN: 978-1-61208-496-1 

# Novel FGMOS based Voltage Differencing Buffered Amplifier and its Filter Applications

Akanksha Ninawe
Indian Institute of Technology Delhi
New Delhi, India
email: iitd.akanksha@gmail.com

Himani Kanwar
Netaji Subhas Institute of Technology
New Delhi, India
email: himani.himani2@gmail.com

Richa Srivastava, Devesh Singh
A.K.G. Engineering College
Ghaziabad, India
emails: {richa_ec@yahoo.co.in, dsinghece@gmail.com}

***Abstract*—The work presents Floating Gate MOS (FGMOS) based low-voltage, low-power (LV/LP) variant of recently proposed Voltage Differencing Buffered Amplifier (VDBA). The proposed topology operates at low supply of ±1.35V with total power consumption of 0.745mW. The linearity of the Operational Transconductance Amplifier (OTA) stage of the proposed active element is observed to increase compared to the conventional VDBA and the same has been proved for several supply voltages. The application of the proposed circuit is verified through robust resistorless voltage mode universal biquad filters which are observed to implement standard filter functions. The simulations are performed through SPICE in 0.18µm technology to validate the workability of the proposed circuit. The work is intended to find applications in low-voltage, low-power battery-operated medical devices and other analog signal processing circuits.**

***Keywords-FGMOS; Voltage Differencing Buffered Amplifier (VDBA); Operational Transconductance Amplifier (OTA); Universal Filter; Analog Signal Processing***

## I. INTRODUCTION

Analog signal processing (ASP) with the use of active elements employing Voltage Mode (VM) techniques is seen in existing research. VM techniques offer several advantages [1]. Among several VM active elements, VDBA has attractive properties of current mode technology such as reduced power consumption, larger bandwidth, wider linearity and higher slew rate compared to OP-AMP [2]. Furthermore, lower output impedance of VDBA compared to OTA eliminates loading effect which is suitable for VM circuit synthesis. These evidences demand ASP circuits with VDBA as a building block. Several modifications of VDBA have been suggested by Sotner et al. [3].

The demand for LV/LP electronics is inevitable in modern portable consumer electronics and battery-operated wearable and implantable biomedical devices. FGMOS techniques have dominated in this field of research with advantages of flexibility, controllability and tunability. Moreover, narrower bandwidth and relatively lower transconductance compared to the conventional MOS transistor are attractive features for biomedical devices since biological signals have extremely low amplitude and frequency. Applications of FGMOS in voltage buffer, analog inverter, winner-take-all (WTA), neural networks, electronic programming, squarers, current mirrors, multipliers, digital-to-analog and analog-to-digital converters have been reported [4]. Several research publications have dealt with CMOS implementations of VDBA [2][3][5]. Few VM active elements have been designed using the FG technique, such as Op Amp [6], OTA [7]-[10] and class AB output stage for CMOS Op-Amps [11]. As per authors' knowledge, suitable literature related to FGMOS based VDBA was not found.

This paper introduces a favorable structure of a new FGMOS VDBA suitable for VM ASP described in Section II. Implementation of second order active filters has been possible with the utilization of the active block. Two VM filters containing the proposed VDBA have been simulated and are given in Section III. The workability of the proposed circuits is confirmed by PSPICE simulations. These are discussed in Section IV. A comparison of the previously reported and the proposed VDBA is done. The use of proposed VDBA is confirmed with its applications to first and second form biquads and their capability in generating all filter functions. Section V concludes the results obtained.

## II. PROPOSED FGMOS REALIZATION OF VDBA

Conception of the classical VDBA has been formulated in [2][5][12]. Using standard notation, port currents and voltages of CMOS VDBA can be described by (1):

$$\begin{bmatrix} I_P \\ I_N \\ I_Z \\ V_W \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ g_m & -g_m & 0 & 0 \\ 0 & 0 & \alpha & 0 \end{bmatrix} \begin{bmatrix} V_P \\ V_N \\ V_Z \\ I_W \end{bmatrix} \quad (1)$$

$g_m$ is transconductance of VDBA, α is the corresponding voltage ratio ($\alpha = 1-\varepsilon_v$) and $\varepsilon_v$ is voltage tracking error. The circuit symbol and schematic implementation of the proposed VDBA are shown in Figures 1 and 2 respectively. It has two fundamental blocks: OTA ($M_1$-$M_9$) and voltage buffer ($M_{10}$-$M_{16}$), both are realized by FGMOS variants of MOS differential pairs. FGMOS differential pair has been employed in OTA to ensure low voltage operation because of low FGMOS threshold voltage. In the differential pair

 ISBN: 978-1-61208-496-1 

formed by two floating gate (FG) transistors M1 and M2, one control input of each transistor is used for signal processing purpose, other control input is used for biasing, and hence an adjustable threshold voltage is achieved. The differential input voltage $V_{in} = V_{FG1} - V_{FG2}$, produces an output current that is voltage controlled current source (VCCS) [2]. $V_{FG1}$ is the FG voltage at M1 given by $V_{FG1} = \frac{C_1}{C_T}V_N + \frac{C_2}{C_T}V_{bias}$ and the FG voltage at M2 is given by $V_{FG2} = \frac{C_1}{C_T}V_P + \frac{C_2}{C_T}V_{bias}$, where $C_T$ is total capacitance, $C_T = C_1+C_2+C_{GS}+C_{GD}$. $C_1$ and $C_2$ are the FG capacitances. The current mirror load for the OTA is formed by M5 and M6. A voltage buffer is used in the output stage of the proposed circuit. It consists of a differential amplifier (M10 - M13) and a feedback transistor M14. The first control input of FGMOS M10 is $V_Z$ terminal voltage. The input range of the OTA increases by the use of FGMOS transistors. The assumption is that identical MOSFETs and identical FGMOS transistors are used and all operate in saturation.

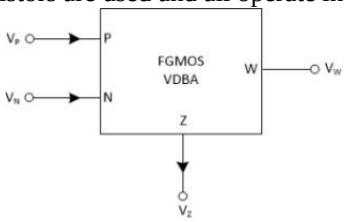



Figure 1. Circuit symbol: FGMOS VDBA

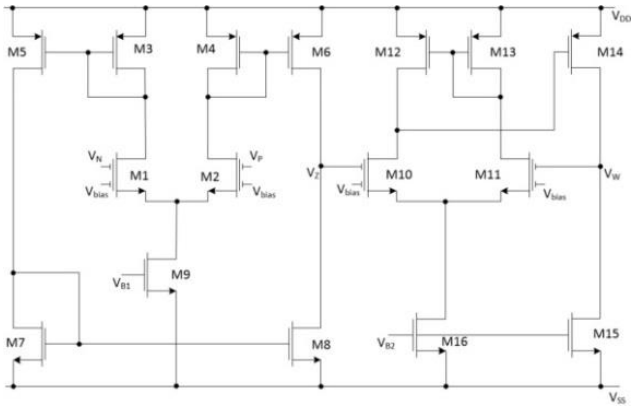



Figure 2. FGMOS implementation of the VDBA

*Calculation of input range of the proposed OTA:*

Condition for M2 to be in saturation is given by

$$V_{D2} \geq V_{FG2} - V_{T2} \quad (2)$$

and

$$V_{D2} = V_{G4} \quad (3)$$

$$V_{GS4} = V_{ov4} + V_{T4} \quad (4)$$

$V_{ovi}$ and $V_{Ti}$ are respectively the overdrive and threshold voltages of $i^{th}$ transistor. (4) can be expressed as

$$V_{G4} - V_{S4} = V_{ov4} + V_{T4} \quad (5)$$

Substituting $V_{S4}=V_{DD}$ in (5) gives

$$V_{G4} = V_{ov4} + V_{T4} + V_{DD} \quad (6)$$

Using (2), (3) and (6) gives

$$V_{ov4} + V_{T4} + V_{DD} \geq V_{FG2} - V_{T2} \quad (7)$$

From Figure 2 the FG voltage of M2 is given as

$$V_{FG2} = \frac{C_1}{C_T}V_P + \frac{C_2}{C_T}V_{bias} \quad (8)$$

where $C_1$ and $C_2$ are the capacitances of FGMOS transistor and $C_T=C_1+C_2+C_{GS}+C_{GD}$, is the total capacitance.

Substituting $V_{FG2}$ from (8) in (7) gives

$$V_{ov4} + V_{T4} + V_{DD} \geq \frac{C_1}{C_T}V_P + \frac{C_2}{C_T}V_{bias} - V_{T2} \quad (9)$$

$$V_P \leq \frac{C_T}{C_1}V_{ov4} + \frac{C_T}{C_1}V_{T4} + \frac{C_T}{C_1}V_{DD} - \frac{C_2}{C_1}V_{bias} + \frac{C_T}{C_1}V_{T2} \quad (10)$$

If $C_1=C_2$ and $C_T \approx C_1+C_2$, then (10) can be reduced to

$$V_P \leq 2V_{DD} + 2V_{ov4} + 2V_{T4} - V_{bias} + 2V_{T2} \quad (11)$$

The maximum input range of conventional VDBA is [2]

$$V_P \leq V_{DD} + V_{ov4} + V_{T4} + V_{T2} \quad (12)$$

On comparing (11) and (12), maximum input range is seen higher for proposed VDBA compared to the conventional. This has been proved through simulations in Section IV. Circuits proposed in the work are based on the simulation model of a 2 – input FGMOS transistor. It is assumed that the bulk transconductance of FGMOS is negligibly small.

## III. FILTER APPLICATIONS

This section discusses filter applications of proposed VDBA. The filters consist of two cascaded FGMOS VDBA with two capacitors as passive components. Robust biquad filter configurations employing CMOS based VDBA have been proposed by Kacar et al. [2]. The proposed biquads (Figures 3 and 4) exhibiting filter functions are obtained on similar lines with the objective of low voltage operation. The nodal analysis of the proposed filters yield transfer functions given by (13) and (14). The relations for natural frequency and quality factor of FGMOS based Biquad 1 and Biquad 2 circuits are given by (15) and (16) respectively. For both biquads to realize standard filter functions, different values of $V_1$, $V_2$ and $V_3$ can be employed as given by Kacar et al. [2].

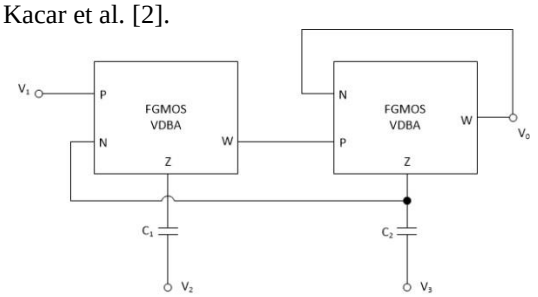



Figure 3. Proposed biquad 1 showing several filter functions

$$V_o = \frac{V_3\alpha_2 s^2 + V_2\frac{g_{mF2}\alpha_1\alpha_2}{C_2}s + V_1\frac{g_{mF1}g_{mF2}\alpha_1\alpha_2}{C_1C_2}}{s^2 + s\frac{g_{mF2}C_1\alpha_2}{C_1C_2} + \frac{g_{mF1}g_{mF2}\alpha_1}{C_1C_2}} \quad (13)$$

$$V_o = \frac{V_3\alpha_2 s^2 - V_2\frac{g_{mF2}\alpha_2}{C_2}s + V_1\frac{g_{mF1}g_{mF2}\alpha_1\alpha_2}{C_1C_2}}{s^2 + s\frac{1}{R_1C_2} + \frac{g_{mF1}g_{mF2}\alpha_1\alpha_2}{C_1C_2}} \quad (14)$$

$$f_0(\text{Biquad 1}) = \frac{1}{2\pi}\sqrt{\frac{g_{mF1}g_{mF2}\alpha_1}{C_1C_2}} \quad (15a)$$

$$Q(\text{Biquad 1}) = \frac{1}{\alpha_2}\sqrt{\frac{g_{mF1}C_2\alpha_2}{g_{mF2}C_1}} \quad (15b)$$

$$f_0(\text{Biquad 2}) = \frac{1}{2\pi}\sqrt{\frac{g_{mF1}g_{mF2}\alpha_1\alpha_2}{C_1C_2}} \quad (16a)$$

$$Q(\text{Biquad 2}) = R_1\sqrt{\frac{g_{mF1}g_{mF2}C_2\alpha_1\alpha_2}{C_1}} \quad (16b)$$

 ISBN: 978-1-61208-496-1

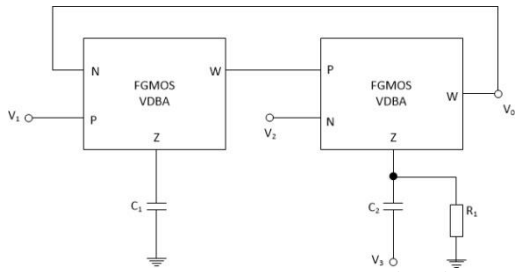



Figure 4. Proposed biquad 2 showing several filter functions

## IV. SIMULATION RESULTS

All simulations are done with TSMC CMOS 0.18 µm technology, supply voltage ±1.35V, $C_{F1} = C_{F2} = 200$ fF. Model suggested by Rodriguez-Villegas [13] has been used to overcome DC convergence error because of FG. Aspect ratio values of transistors are given in Table I. Comparison of previous topologies with the proposed is done in Table II. An increase in the input range of OTA section of the proposed VDBA is seen. Reduced power consumption and transconductance is also observed. Increased bandwidth of the OTA stage is also seen. The gain of buffer stage of the proposed VDBA is ~1 and more reliable compared to previous topologies. DC transfer characteristic of the output stage of the proposed VDBA at bias voltages $V_{B1}$ = 1V and $V_{B2}$ = 0.45V is shown in Figure 5, where the linearity extends from -1V to +1.5V. The $I_Z$ versus $V_P$ and $V_N$ is shown in Figure 6. The linearity extends in the range ±0.4V while a non-linear behavior is observed elsewhere. The operation of OTA stage of the proposed VDBA, depicting $I_Z$ against $V_P$ for different supply voltages is seen in Figure 7. Though the swing for which the OTA justifies its operation decreases by reducing the power supply; approximately ±350 µA at supply voltage of ±1.45V, ±300 µA at supply voltage of ±1.35V, ±250 µA at supply voltage of ±1.25V and ±200 µA at supply voltage of ±1.15V, the current swing provided for conventional VDBA reported in [2], is around ±140 µA, and that too for a supply voltage of ±1.5V which is high enough compared to the one simulated for the proposed VDBA. The linearity observed is in the range ±0.4V for the proposed VDBA performed at different supply voltages, while the linearity extends only upto a threshold of ±0.2V for the conventional VDBA [6]. This increase in the range of operation of the proposed VDBA can be accounted for the fact that the equivalent capacitance ratio, $C_i/C_T$ (FG capacitance $C_i$) scales down the effective input signal at FG, resulting in a wider range. The frequency response depicting the transconductance of OTA stage and complete proposed VDBA are shown in Figures 8 and 9. One out of the two inputs of FGMOS transistors in the OTA stage is grounded. The transconductance of OTA stage of the proposed VDBA at bias voltage of 0.41V is 483µA/V. The bandwidth of input stage of the OTA is found to be 385 MHz at $V_b$ = -0.41V. Biquad 1 implemented using FGMOS VDBA depicting standard filter functions are shown in Figure 10.

TABLE I. ASPECT RATIOS OF THE TRANSISTORS OF PROPOSED VDBA

| Transistors | W(µm) | L(µm) |
|---|---|---|
| $M_1$, $M_2$, $M_3$, $M_4$, $M_{10}$, $M_{11}$, $M_{15}$, $M_{16}$ | 7 | 0.35 |
| $M_5$, $M_6$ | 21 | 0.7 |
| $M_7$, $M_8$ | 7 | 0.7 |
| $M_9$ | 3.5 | 0.7 |
| $M_{12}$, $M_{13}$, $M_{14}$ | 14 | 0.35 |

TABLE II. COMPARISON OF CONVENTIONAL AND PROPOSED VDBA

| **Comparison Parameters** | ***DO-VDBA [3]*** | ***FB-VDBA [3]*** | ***Conventional VDBA [2]*** | ***Proposed VDBA*** |
|---|---|---|---|---|
| Technology (µm) | 0.18 | 0.18 | 0.35 | 0.18 |
| Supply (V) | ± 1.2 | ± 1.2 | ± 1.5 | ± 1.35 |
| No. of transistors | 8 MOS | 16 MOS | 16 MOS | 12 MOS + 4 FGMOS |
| Input range of OTA section(mV) | ± 200 | ± 200 | ±200 | ±400 |
| Power | Not reported | Not reported | 0.97 mW | 0.745 mW |
| Output impedance at $V_w$ (kΩ) | 0.053 | 130 | Not reported | 1.132 |
| Transconductance ($g_m$) (µS) | 865 at $I_B$ = 100 µA | 1740 at $I_B$ = 100 µA | 748 at $V_{B1}$ = 0.44V, $V_{B2}$ = 0.9V | 483 at $V_{B1}$ = 0.41V, $V_{B2}$ = 0.643V |
| Port z impedance (kΩ) | 170 | 130 | Not reported | 101.9 |
| Bandwidth of OTA stage (MHz) | 217 at $I_B$ = 100 µA | 70 at $I_B$ = 100 µA | Not reported | 385 at $V_b$ = -0.41V |
| Gain of buffer stage ($V_w/V_z$) | 0.962 | 0.962 | Not reported | 0.97 |

Passive component values chosen are $C_1 = C_2 = 100$pF and transconductances $g_{mF1} = g_{mF2} = 483$µA/V. VM biquad in Figure 4 were designed for $f_0$ = 0.815 MHz and Q-factor of 1. Various responses of the proposed universal filter (Figure 4) are shown in Figure 11. For simulations, equal passive capacitance values $C_1 = C_2 = 100$pF and $R_1 = 5$kΩ are chosen for natural frequency of 0.815 MHz and Q-factor of 1.024. It is seen from the frequency responses of Figures 10 and 11 that the proposed filters are capable of performing standard filter functions. The simulation results emphasize high linearity and high performance of the proposed VDBA in terms of reduced power consumption.

## V. CONCLUSION

A new FGMOS realization of VDBA that operates on low voltage with reduced power consumption was presented. Realization of resistorless voltage mode biquad filters is possible with the proposed FGMOS based VDBA and are seen to implement standard filter functions. The benefits of the new proposed VDBA on the filter performance include low-power implementation and simpler circuitry. The proposed circuit and universal filters are intended to find applications in LV/LP ASP and consumer electronics.

## ACKNOWLEDGEMENT

The work was carried out at Netaji Subhas Institute of Technology, New Delhi (University of Delhi) as a partial fulfillment of Engineering Degree.

 ISBN: 978-1-61208-496-1 

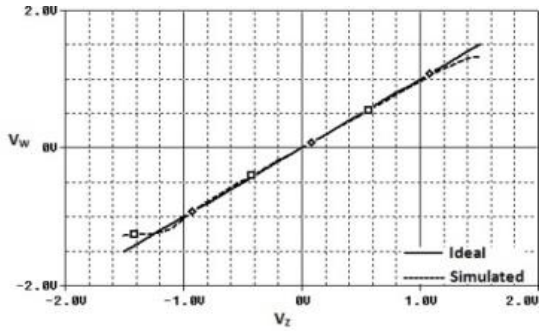



Figure 5. DC transfer characteristic $V_W$ versus $V_Z$

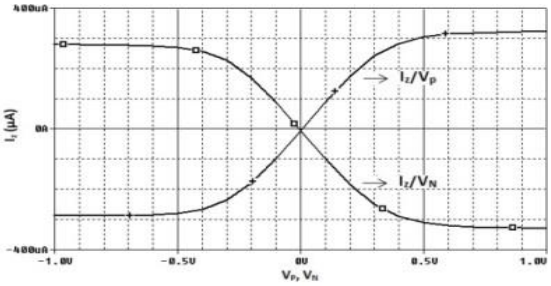



Figure 6. DC transfer characteristic $I_Z$ vs $V_P$ and $V_N$

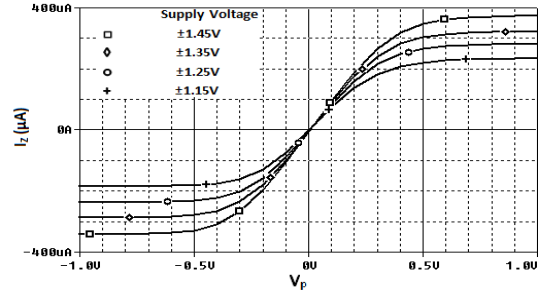



Figure 7. $I_Z$ versus $V_P$ at different supply voltages

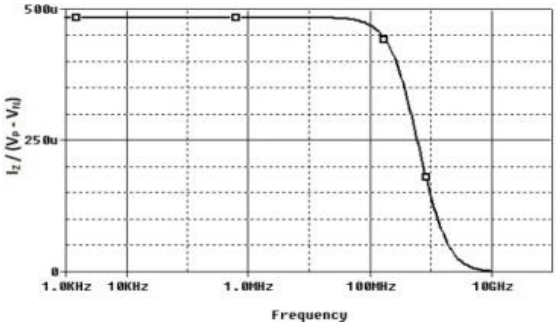



Figure 8. Frequency response of OTA stage

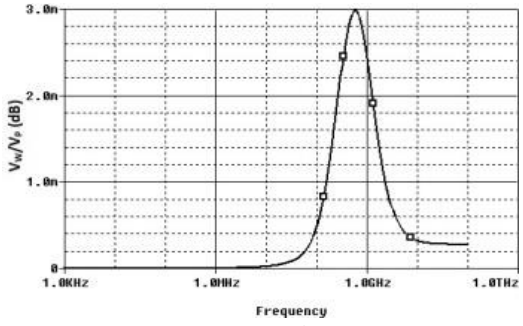



Figure 9. Frequency response of complete proposed VDBA

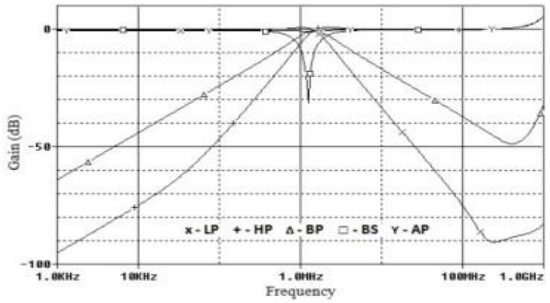



Figure 10. Various responses of proposed universal filter (Biquad 1)

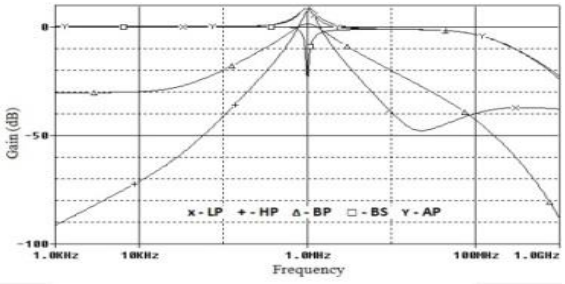



Figure 11. Various responses of proposed universal filter (Biquad 2)

## REFERENCES


[1] F. Yuan, "Voltage-mode versus current-mode: A critical comparison" in CMOS Current-Mode Circuits for Data Communications, 2007, pp. 1-12.

[2] F. Kacar, A. Yesil, and A. Noori, "New CMOS realization of voltage differencing buffered amplifier and its biquad filter applications", Radioengineering, vol. 21(1), 2012, pp. 333-339.

[3] R. Sotner, J. Jerabek, and N. Herencsar, "Voltage differencing buffered/inverted amplifiers and their applications for signal generation", Radioengineering, 22 (2), 2013 pp. 490-504.

[4] A. Ninawe, R. Srivastava, A. Dewaker, and M. Gupta, "Design of low-voltage, low-power FGMOS based voltage buffer, analog inverter and winner-take-all analog signal processing circuits", Circuits Syst, vol. 7, 2016, pp. 1-10.

[5] D. Biolek , R. Senani , V. Biolkova and, Z. Kolka, "Active elements for analog signal processing: classification, review, and new proposals", Radioengineering, vol. 17 (4), 2008, pp. 15-32.

[6] E. R. Ruotsalainen, K. Lasanen, and J. Kostamovaara, "A 1.2 V micropower CMOS op amp with floating-gate input transistors," 43rd IEEE MWCAS, vol. 2, 2000, pp. 794-797.

[7] Z. Alsibai, Floating-gate operational transconductance amplifier, 2013, IJIEE, vol. 3(4), 2013, pp. 361-364, DOI: 10.7763/IJIEE.2013.V3.335

[8] O. Naess and Y.Berg, "Tunable floating-gate low-voltage transconductor", IEEE Int. Symp. Circuits and Systems, vol.4, 2002, pp. 663-666.

[9] V.S. Babu, A. Sekhar, R. Salini Devi, and M.R. Baiju, "Floating gate MOSFET based operational transconductance amplifier and study of mismatch", 4th IEEE C IND ELECT APPL, 2009, pp. 127-132.

[10] Y. Berg, O. Naess and M. Hovin, "Ultra low voltage floating-gate transconductance amplifier with tunable gain and linearity", IEEE INT SYMP CIRC S, vol.3, 2000, pp. 343 – 346.

[11] R.G. Carvajal, A. Torralba, J. Tombs, F. Muñoz, and J. Ramírez-Angulo, "Low voltage class AB output stage for CMOS Op-amps using multiple input floating gate transistors", AICSP, vol. 36(3), 2003, pp. 245-249.

[12] V. Biolkova, Z. Kolka, and D. Biolek, "Fully balanced voltage differencing buffered amplifier and its applications", 52nd IEEE MWCAS, 2009, pp. 45- 48.

[13] E. Rodriguez-Villegas, "Low power and low voltage circuit design with the FGMOS transistor", IET CIRC DEVICE SYST, 2006. ISBN: 086341


 ISBN: 978-1-61208-496-1 

# A High-Speed Programmable Network Intrusion Detection System Based on a Multi-Byte Transition NFA

Tomoaki Hashimoto, Shin'ichi Wakabayashi*, Shinobu Nagayama†, Masato Inagi, Ryohei Koishi, Hiroki Takaguchi

Graduate School of Information Sciences, Hiroshima City University,
Hiroshima 731-3194, Japan
email: {*wakaba, †s_naga}@hiroshima-cu.ac.jp

***Abstract*—To improve the network security, when a virus pattern is updated, an arbitrary updated pattern should be quickly set in a network intrusion detection system (NIDS). This type of NIDS is called "programmable." However, present programmable NIDSs could hardly be applied to a high-speed network with more than 10 Gbps of network transmission speed due to the limitation of clock frequency of the circuit. To overcome this speed limitation, this paper proposes a programmable NIDS based on a multi-byte transition non-deterministic finite automaton (NFA). The proposed NIDS is implemented on an FPGA to evaluate its performance. The FPGA implementation results show that the proposed NIDS can achieve more than 10 Gbps of throughput.**

***Keywords-Regular expression matching; non-deterministic finite automaton; programmable hardware; network intrusion detection system; FPGA.***

## I. INTRODUCTION

Network intrusion detection systems (NIDSs) that can detect network attacks such as computer viruses and worms in real time have become indispensable nowadays to maintain network security. To detect suspicious data patterns included in packets, NIDSs perform *regular expression matching* [14] between packet payloads and data patterns predefined as regular expressions. Since regular expression matching is a time-consuming task, NIDSs tend to be bottleneck in network transmission. In addition, since regular expression patterns are frequently updated, NIDSs tend to be security holes until pattern updating is completed. Thus, 1) fast regular expression matching and 2) quick pattern updating are major requirements for NIDSs.

Software NIDSs that perform regular expression matching by a software program are popular. The Snort system [19] is well-known as an open source software NIDS, and its regular expression patterns are available at the website. Software NIDSs can update regular expression patterns quickly, but speed of regular expression matching is slow. Although various algorithms [1][3][6][7][12][14][18][21][25] have been proposed to perform regular expression matching faster, their software implementation is difficult to achieve enough performance to catch up with the speed of latest Gigabit Ethernet with more than 10Gbps.

On the other hand, hardware NIDSs that perform regular expression matching with FPGAs [4][8][10][13][15][16][17][23][24] can perform regular expression matching much faster, but they require long time for pattern updating because regular expression patterns are embedded as hardware circuits. To update patterns in such *pattern-specific* circuits, a sequence of FPGA design and implementation processes (i.e., generating HDL code, logic synthesis, place and route, etc.) should be performed again. It is well known that those FPGA design processes require a fairly long time, sometimes, a few hours. Since this time is longer than update interval [2], it is hard to keep NIDSs up to date.

To overcome this problem due to architecture of NIDSs, *programmable NIDSs* [5][11][20][22] that perform regular expression matching with *pattern-independent* circuits have been proposed as the third approach. The programmable NIDSs can update patterns more quickly because regular expression patterns are stored in registers. In addition, the programmable NIDSs can perform regular expression matching much faster than software NIDSs. However, since size of programmable NIDSs is larger than that of pattern-specific NIDSs, we use a programmable NIDS with pattern-specific ones, as hybrid NIDSs. In hybrid NIDSs, a pattern-specific NIDS performs regular expression matching for all patterns currently registered in the NIDS, while the programmable NIDS performs regular expression matching for only new patterns. In this way, we can compensate the defect of programmable NIDS in terms of hardware size. However, since throughput of the existing programmable NIDSs is at most a few Gbps, faster programmable NIDSs are still required to achieve more than 10 Gbps of throughput.

The existing programmable NIDSs process one byte of packet payload per one clock, and their throughput has been improved mainly by increasing clock frequency of circuits. But, there is a limitation to increase of clock frequency. Thus, this paper proposes a method to improve throughput of a programmable NIDS by processing $k$ bytes per one clock. We design such a high-speed programmable NIDS using a $k$-byte transition non-deterministic finite automaton (NFA) that is converted from a one-byte transition NFA.

Although a similar method has been proposed by Yamagaki et al. [23], their method is targeted to their hardware NIDS, and thus, the capability of quick pattern updating

 ISBN: 978-1-61208-496-1

is not considered. Therefore, we propose an architecture of a programmable NIDS considering quick pattern updating. As far as we know, a programmable NIDS based on $k$-byte transition NFA has not been proposed so far.

The rest of this paper is organized as follows: Section II briefly defines regular expressions and NFAs. Section III introduces $k$-byte transition NFAs. Section IV presents architecture of a programmable NIDS based on $k$-byte transition NFAs. Its FPGA implementation results are shown in Section V. Section VI concludes the paper.

## II. Preliminaries

### A. Regular Expressions

In NIDSs, suspicious data patterns to be found are described by not only simple strings, but also **regular expressions** [9]. This is because its acceptance algorithm is simple, and it has practically enough expression power [13]. A set of strings represented by regular expression is called a **regular set** or regular language. Regular expressions and their regular sets are recursively defined as follows:

*Definition 1:* Let $\Sigma$ be a finite set of characters: $\Sigma = \{a_1, a_2, \ldots, a_n\}$, called the **alphabet**, $R$ and $S$ be regular expressions on $\Sigma$, and $L(R)$ and $L(S)$ be regular sets denoted by $R$ and $S$, respectively. Then,

1) $\emptyset$ is a regular expression denoting the regular set $\emptyset$.
2) $\varepsilon$ is a regular expression denoting the regular set $\{\varepsilon\}$ (the null character).
3) A character $a_i \in \Sigma$ is a regular expression denoting the regular set $\{a_i\}$.
4) An alternation $R \mid S$ is a regular expression denoting the regular set $L(R) \cup L(S)$.
5) A concatenation $R \cdot S$ is a regular expression denoting the regular set $\{rs | r \in L(R), s \in L(S)\}$. $R^1 = R$ and $R^m = R \cdot R^{m-1}$ for $m \geq 2$. Usually, '$\cdot$' is omitted.
6) A Kleene closure $R*$ is a regular expression $\varepsilon \mid R \mid R^2 \mid \ldots$ denoting the regular set $\{\varepsilon\} \cup L(R) \cup L(R^2) \cup \ldots$, where $L(R), L(R^2), \ldots$ are regular sets denoted by $R, R^2, \ldots$, respectively.

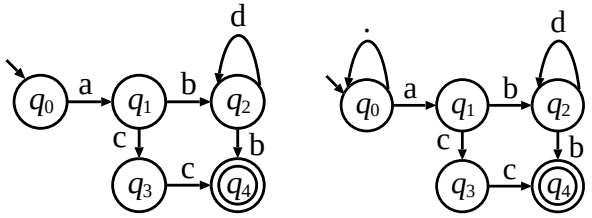
Only expressions obtained by considering the above 1) to 3) as constants, and applying the above operations 4) to 6) finite times are regular expressions on $\Sigma$.

For simplicity of expressions, this paper introduces the following operation as well:

7) A dot . denotes the set $L(.) = \Sigma \cup \{\varepsilon\}$ (don't care). □

### B. Non-Deterministic Finite Automaton

*Definition 2:* A **non-deterministic finite automaton (NFA)** [9] is defined by a 5-tuple $(Q, \Sigma, \delta, q_0, F)$, where $Q$ is a finite set of states, $\Sigma$ is the alphabet, $\delta$ is a state transition function $\delta : Q \times (\Sigma \cup \{\varepsilon\}) \to 2^Q$, $2^Q$ is the power set of $Q$, $q_0 \in Q$ is an initial state, and $F \subseteq Q$ is a set of accepting states. When $\delta : Q \times \Sigma \to 2^Q$, it is called an **$\varepsilon$-free NFA**. In the following, an NFA means an $\varepsilon$-free NFA, unless otherwise stated. □




(a) NFA for prefix matching. (b) NFA for infix matching.

Figure 1. NFAs for regular expression $a(cc \mid bd*b)$.

An arbitrary regular expression $R$ can be converted into an NFA that accepts its regular language $L(R)$. Usually, NFAs are represented as directed graphs. Since directed graphs can be represented by adjacency matrices, NFAs can also be represented by adjacency matrices as follows:

*Example 1:* Figure 1(a) shows an NFA for the regular expression $a(cc \mid bd*b)$. In this NFA, $q_0$ is the initial state, and $q_4$ is the accepting state. By considering this NFA as a directed graph, the NFA can be represented by the following adjacency matrix $M$:

$$M = \begin{pmatrix} \emptyset & a & \emptyset & \emptyset & \emptyset \\ \emptyset & \emptyset & b & c & \emptyset \\ \emptyset & \emptyset & d & \emptyset & b \\ \emptyset & \emptyset & \emptyset & \emptyset & c \\ \emptyset & \emptyset & \emptyset & \emptyset & \emptyset \end{pmatrix}$$

In this matrix, an element in $i$-th row and $j$-th column of $M$ denotes a set of transition characters from $q_i$ to $q_j$. When no state transition is defined, $\emptyset$ is entered as the corresponding element. □

### C. Regular Expression Matching Based on NFAs

Regular expression matching performed in NIDSs is often formulated as follows:

*Problem 1:* Given a regular expression $R$ and a text $T$, detect if $T$ has a string in $L(R)$ as a substring. □

Regular expression matching can be performed by iterating state transitions on an NFA for $R$ according to characters in $T$. If a substring in $T$ can cause state transitions from the initial state $q_0$ to an accepting state in $F$, then it is in $L(R)$ (i.e., a matching is achieved). Otherwise, the substring is not in $L(R)$ (i.e., matching failed).

Although the NFA in Figure 1(a) is equivalent to the regular expression $a(cc \mid bd*b)$, the NFA in Figure 1(b) is used in NIDSs to solve Problem 1. This is because the NFA can discard unmatched prefix characters by the state transition by the don't care character at $q_0$. In the following, we focus on this type of NFAs.

Regular expression matching hardware based on NFAs realizes the above behavior on a circuit. Since in an NFA, multiple states can be active, state transitions from active states caused by a character are computed in parallel in hardware. Existing programmable NIDSs based on NFAs [5][11][20][22] compute all state transitions caused by a character in a clock.

 ISBN: 978-1-61208-496-1 

## III. Multi-Byte Transition NFAs

The alphabet $\Sigma$, on which regular expressions and NFAs are defined, usually consists of one-byte (8-bit) characters. Since existing programmable NIDSs based on NFAs compute all state transitions for a character in a clock, network transmission speed $s$ that the NIDSs can be applied is $s = 8f$ bits per second, where $f$ is clock frequency of the NIDSs. Thus, network transmission speed $s$ has been increased by increasing clock frequency $f$ of circuits so far. However, there is a limitation to increase of clock frequency, and thus, it is difficult to achieve 10 Gbps of network transmission speed in this way.

To overcome the problem, we introduce multi-byte transition NFAs [23] to design of programmable NIDSs.

### A. Definition of Multi-Byte Transition NFAs

Whereas ordinary NFAs require a *one-byte character* for state transitions, multi-byte transition NFAs require a *multi-byte character*. To obtain multi-byte transition NFAs that accept the same regular sets as regular expressions $R$ defined on one-byte characters, we generate $k$-byte transition NFAs, in which state transitions are caused by $k$ *one-byte characters*, from one-byte transition NFAs for $R$. Thus, this paper defines multi-byte transition NFAs as follows:

*Definition 3:* Let a one-byte transition NFA be $(Q, \Sigma, \delta, q_0, F)$. Then, a **multi-byte ($k$-byte) transition NFA** equivalent to it is defined by $(Q_k, \Sigma^{k'}, \delta_k, q_0, F_k)$. $F_k = F \cup F'$ where $F'$ is a set of additional accepting states. $Q_k = F' \cup Q'$ where $Q' \subseteq Q$. $\Sigma^{k'} = \Sigma^k \cup \bigcup_{i=1}^{k-1} \Sigma^i \times \{\varepsilon\}^{k-i}$, where products of sets are Cartesian products. And, $\delta_k : Q_k \times \Sigma^{k'} \to 2^{Q_k}$. □

Since $k$-byte transition NFAs take in $k$ one-byte characters at a time, null characters $\varepsilon$ can be input at the end of a text when length of the text is not a multiple of $k$. Thus, in this definition, $\varepsilon$ is added to $\Sigma^{k'}$. And, for the same reason, $F'$ is added. The next subsection shows how to produce such $k$-byte transition NFAs.

### B. Conversion into Multi-Byte Transition NFA

We can generate $k$-byte transition NFAs of Definition 3 by converting $k$ state transitions (a path of length $k$) on one-byte transition NFAs into a state transition (a path of length 1). Such path conversions can be achieved by considering NFAs as directed graphs, and raising adjacency matrices of NFAs to the $k$-th power [23]. In the following, we introduce the conversion method briefly. For more details, see [23].

Given an adjacency matrix $M$ of a one-byte transition NFA, at first, add a symbol $\omega$ to diagonal elements of $M$ that are associated with accepting states. The symbol $\omega$ corresponds to $\varepsilon$ in Definition 3. Let $M^{(1)}$ be the obtained matrix. Then, obtain $M^{(k)}$ by raising $M^{(1)}$ to the $k$-th power using the matrix multiplication defined in the following:

*Definition 4:* Since all elements of $M^{(1)}$ can be considered as regular expressions, multiplications and additions on ordinary scalar matrix multiplication are computed as concatenation $\cdot$ and alternation $|$, respectively, except for the following cases:

- Multiplication with $\emptyset$ results in $\emptyset$.
- $. \times E = .E$ and $E \times . = \emptyset$, where $E$ is a regular expression for an element.
- $\omega \times E = \emptyset$ and $E \times \omega = E\omega$. □

After the matrix multiplications, apply the following operations to $M^{(k)}$ to obtain an adjacency matrix $M_a^{(k)}$ of a $k$-byte transition NFA.

1) Add $k-1$ columns for new accepting states $q_{f(1)}, q_{f(2)}, \ldots, q_{f(k-1)}$, and redefine strings including $i$ $\omega$'s as state transitions to the new accepting state $q_{f(i)}$. Note that no row is added because there is no state transition from the new accepting states.
2) Replace $\omega$ with the don't care character. Although $\omega$ can be replaced with $\varepsilon$, the don't care character is preferred because it makes hardware implementation simpler.

And, finally, eliminate states without incoming state transition, except for the initial state, to obtain an irredundant $k$-byte transition NFA.

*Example 2:* $M_a^{(2)}$ for the NFA in Figure 1(b) is as follows:

$$
\begin{aligned}
M^{(2)} &= M^{(1)} \times M^{(1)} \\
&= \begin{pmatrix} . & a & \emptyset & \emptyset & \emptyset \\ \emptyset & \emptyset & b & c & \emptyset \\ \emptyset & \emptyset & d & \emptyset & b \\ \emptyset & \emptyset & \emptyset & \emptyset & c \\ \emptyset & \emptyset & \emptyset & \emptyset & \omega \end{pmatrix} \begin{pmatrix} . & a & \emptyset & \emptyset & \emptyset \\ \emptyset & \emptyset & b & c & \emptyset \\ \emptyset & \emptyset & d & \emptyset & b \\ \emptyset & \emptyset & \emptyset & \emptyset & c \\ \emptyset & \emptyset & \emptyset & \emptyset & \omega \end{pmatrix} \\
&= \begin{pmatrix} .. & .a & ab & ac & \emptyset \\ \emptyset & \emptyset & bd & \emptyset & bb \mid cc \\ \emptyset & \emptyset & dd & \emptyset & db \mid b\omega \\ \emptyset & \emptyset & \emptyset & \emptyset & c\omega \\ \emptyset & \emptyset & \emptyset & \emptyset & \emptyset \end{pmatrix} \\
M_a^{(2)} &= \begin{pmatrix} .. & .a & ab & ac & \emptyset & \emptyset \\ \emptyset & \emptyset & bd & \emptyset & bb \mid cc & \emptyset \\ \emptyset & \emptyset & dd & \emptyset & db & b. \\ \emptyset & \emptyset & \emptyset & \emptyset & \emptyset & c. \\ \emptyset & \emptyset & \emptyset & \emptyset & \emptyset & \emptyset \end{pmatrix}
\end{aligned}
$$

Figure 2 shows a 2-byte transition NFA for $M_a^{(2)}$ □

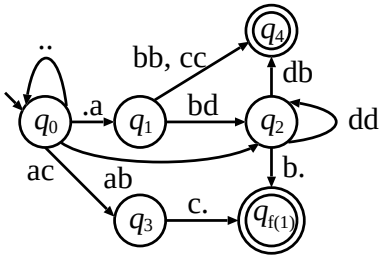



Figure 2. 2-byte transition NFA equivalent to the NFA in Figure 1(b).

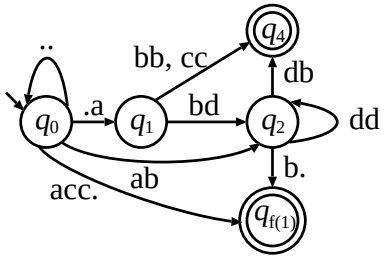



Figure 3. 2-byte STNFA equivalent to the NFA in Figure 1(b).

### C. Multi-Byte String Transition NFAs

A **string transition NFA (STNFA)** [20][22] is an extension of an NFA, and it causes state transitions by a string, not a character. By eliminating states connected linearly in an ordinary NFA, and connecting between its beginning state and ending state directly as a *string transition*, an STNFA is obtained. The STNFA obtained by this simple way is equivalent to the original NFA, and has fewer states resulting in a smaller circuit. Since many concatenations are used in regular expression patterns for NIDSs [19], linearly connected states often appear. Thus, this simple conversion method is effective to reduce circuit size [20][22].

We apply the same method to multi-byte transition NFAs to generate **multi-byte STNFAs** with fewer states. In Figure 2, the states $q_0$, $q_3$, and $q_{f(1)}$ are linearly connected. By eliminating $q_3$ and connecting $q_0$ and $q_{f(1)}$ directly, we obtain a 2-byte STNFA in Figure 3.

## IV. Programmable NIDS Based on Multi-Byte String Transition NFAs

This section presents architecture of our programmable NIDS based on multi-byte STNFAs. To realize both fast regular expression matching and quick pattern updating with a compact circuit, the proposed NIDS is designed taking advantages of multi-byte STNFAs shown in Section III.

### A. Overall Architecture of Our Programmable NIDS

Since overall architecture of our NIDS follows architecture of the NIDS shown in [20], this subsection shows only an overview of its architecture. For more details, see [20].

Figure 4 shows overall architecture of the proposed programmable NIDS. It has a two-dimensional array structure, consisting of two parts: a matching array (MA) and a feedback array (FA). The MA performs string matching needed to trigger state transitions on an STNFA. The FA is a programmable interconnection network to activate next states according to state transitions triggered by the MA.

### B. Matching Array Based on k-Byte String Transition NFAs

The MA is constructed by arranging string matching units (SMUs), shown in Figure 5, in a row. Eash state transition in an STNFA is assigned to an SMU. An SMU is constructed as a one-dimensional array of simple processing

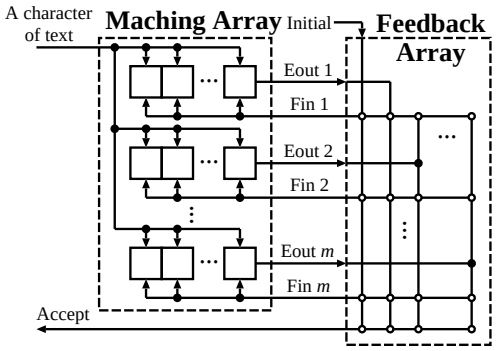



Figure 4. Overall architecture of programmable NIDS [20].

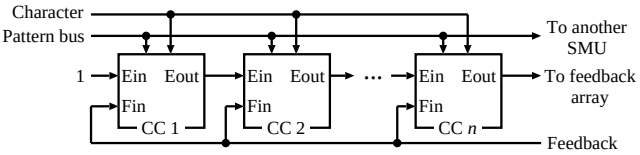



Figure 5. Architecture of string matching unit (SMU) [20].

units, called comparison cells (CCs), shown in Figure 6. Each CC performs one-character matching for a $k$-byte character fed synchronously with the clock, and transmits its matching result to the right neighbor CC via $E_{out}$.

By performing one-character matching sequentially using the pipelined CCs, an SMU performs string matching. When an input string matches with a string pattern stored in an SMU, the SMU outputs an enable signal to the FA to trigger state transitions. Then, by transmitting triggered state transitions (the enable signal) to $F_{in}$ of appropriate SMUs via the FA, next states are activated, since CCs perform one-character matching only when an enable signal is fed via $E_{in}$ or $F_{in}$. Using the pattern bus, we can program which CC receives an enable signal from, $E_{in}$ or $F_{in}$, to the register connected to the selector.

CCs perform one-character matching using character matching tables, shown in Figure 7, that are new components proposed in this paper. An input $k$-byte character is divided into each byte, and is fed to addresses of $k$ RAMs in parallel. Word width of a RAM is 1 bit, and it stores 0 or 1. If a byte $x$ given as an address matches with the $i$-th byte of a $k$-byte

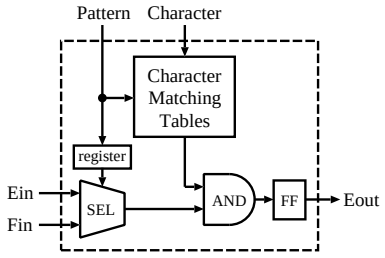



Figure 6. Architecture of comparison cell (CC).

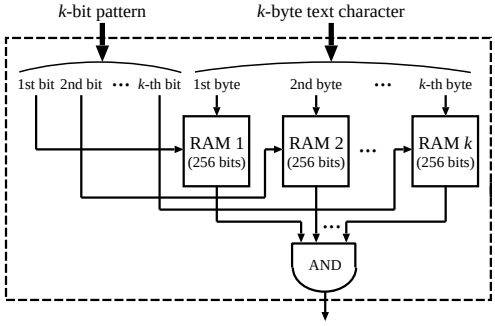



Figure 7. Architecture of character matching tables.

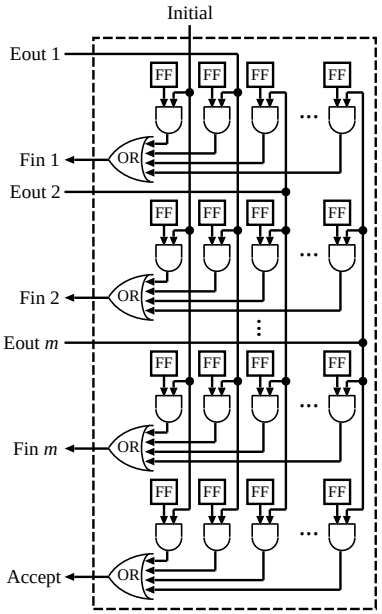



Figure 8. Architecture of feedback array.

pattern character, $\mathrm{RAM}_i[x]$ stores 1. Otherwise, $\mathrm{RAM}_i[x]$ stores 0. If the $i$-th byte of a pattern character is a don't care character, $\mathrm{RAM}_i$ stores 1 for all addresses. Contents of RAMs can be rewritten using the pattern bus.

### C. Feedback Array Based on k-Byte String Transition NFAs

The FA is a programmable interconnection network to transmit enable signals fed from the MA to appropriate SMUs. Figure 8 shows its architecture. By setting appropriate bits to the FFs, like crossbar switches, we can program arbitrarily connections. When an FF stores 1, a vertical line and a horizontal line are connected. Otherwise, lines are disconnected.

Since the proposed NIDS based on $k$-byte STNFA takes in $k$ one-byte characters at a time, network transmission speed $s$ that the NIDS can be applied is improved to $s = 8kf$ bps.

## V. Experimental Results

To experimentally evaluate performance of the proposed programmable NIDS, we used the following five regular expressions randomly chosen from rules of the SNORT [19]:

R1: /level/(0|1|2|3|4|5|6|7|8|9)+/(exec|configure)
R2: cookies\s+Monster\s+server\s+engine
R3: template\s*=\s*{$
R4: fn=..(/ |\)
R5: (((\x0*b*dyndns|\x02yi)\x03org)|((\x07dynserv|\x04mo oo)\x03com))

where \s denotes a blank character, + is a regular expression operator that means $R+ = R \cdot R*$, \x followed by a two-digit hexadecimal number denotes an ASCII code, and the others are treated as one-byte characters in the alphabet $\Sigma$.

### A. Results for Multi-Byte Transition NFAs

Table I shows the numbers of states and state transitions in a $k$-byte transition NFA for each regular expression, where columns of $k = 1$ show results of existing method [20]. From this table, we can see that the numbers of states and state transitions increase as the value of $k$ increases. This is because new accepting states are added, and the number of characters in an alphabet increases as $k$ increases.

By converting into STNFAs, they can be reduced. Since the number of state transitions corresponds to the number of SMUs in the proposed NIDS, we can reduce the circuit size by using STNFAs. When $k = 8$, the number of state transitions is not reduced so much by STNFAs. This is because an 8-byte character consists of 8 one-byte characters, and it already forms a string.

### B. FPGA Implementation Results

We designed programmable NIDSs based on the above $k$-byte STNFAs by setting design parameters as follows: the number of CCs $n = 5$ and the number of SMUs $m = 10, 50, 100, 150, 200$. Since the SNORT rules use 7-bit ASCII characters as one-byte characters, we designed each character matching table in Figure 7 as a 128-bit RAM that is implemented with 2 LUTs in an FPGA. The designed NIDSs were implemented with the Xilinx Virtex-7 XC7VX485T-2FFG1761 FPGA using the Xilinx Vivado Design Suite 2014.2 as a synthesis tool. Table II shows the FPGA implementation results and network transmission speed the proposed NIDSs can achieve.

When $k = 1$, 10 SMUs ($m = 10$) are required in order to realize any rule of the five, since the maximum number of state transitions in an STNFA is 8, as shown in Table I. Although the NIDS with $m = 10$ achieves 429 MHz of operating frequency, its transmission speed is 3.4 Gbps. This corresponds to performance of the existing programmable NIDS [20]. When $k = 2$ and 4, $m = 50$ and 100 are required, respectively. The NIDSs with $m = 50$ and 100 still do not reach 10 Gbps, even though they achieve higher speed than the existing one. When $k = 8$, $m = 150$ is required, but the NIDS can achieve more than 10 Gbps. Even if there is a regular expression that requires $m = 200$, the NIDS with $m = 200$ can achieve more than 10 Gbps when $k = 8$.

 ISBN: 978-1-61208-496-1

TABLE I. NUMBERS OF STATES AND STATE TRANSITIONS IN A $k$-BYTE TRANSITION NFA.

| Rules | Number of states | | | | | | | | Number of state transitions | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Character transition NFAs | | | | String transition NFAs | | | | Character transition NFAs | | | | String transition NFAs | | | |
| | $k=1$ | $k=2$ | $k=4$ | $k=8$ | $k=1$ | $k=2$ | $k=4$ | $k=8$ | $k=1$ | $k=2$ | $k=4$ | $k=8$ | $k=1$ | $k=2$ | $k=4$ | $k=8$ |
| R1 | 22 | 23 | 25 | 29 | 4 | 7 | 13 | 26 | 24 | 38 | 40 | 88 | 6 | 12 | 28 | 85 |
| R2 | 29 | 30 | 32 | 36 | 5 | 12 | 26 | 36 | 32 | 39 | 62 | 144 | 8 | 21 | 56 | 144 |
| R3 | 12 | 13 | 15 | 19 | 4 | 17 | 11 | 19 | 14 | 19 | 39 | 147 | 6 | 13 | 35 | 147 |
| R4 | 7 | 8 | 10 | 14 | 3 | 5 | 9 | 14 | 8 | 10 | 14 | 22 | 4 | 7 | 13 | 22 |
| R5 | 29 | 30 | 32 | 36 | 4 | 7 | 13 | 17 | 32 | 36 | 44 | 60 | 7 | 7 | 25 | 41 |

TABLE II. FPGA IMPLEMENTATION RESULTS OF THE PROPOSED NIDSS.

| $m$ | Total number of LUTs | | | | Operating frequency [MHz] | | | | Transmission speed [Gbps] | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $k=1$ | $k=2$ | $k=4$ | $k=8$ | $k=1$ | $k=2$ | $k=4$ | $k=8$ | $k=1$ | $k=2$ | $k=4$ | $k=8$ |
| 10 | 291 | 332 | 662 | 1,083 | 429 | 383 | 399 | 405 | 3.4 | 6.1 | 12.8 | 25.9 |
| 50 | 2,192 | 2,443 | 3,743 | 6,093 | 351 | 353 | 353 | 312 | 2.8 | 5.6 | 11.3 | 20.0 |
| 100 | 6,449 | 6,850 | 9,550 | 14,150 | 297 | 297 | 276 | 255 | 2.4 | 4.8 | 8.8 | 16.3 |
| 150 | 12,379 | 14,931 | 16,957 | 24,225 | 241 | 242 | 209 | 212 | 1.9 | 3.9 | 6.7 | 13.6 |
| 200 | 20,445 | 22,203 | 29,429 | 38,989 | 216 | 197 | 178 | 182 | 1.7 | 3.2 | 5.7 | 11.6 |

*The underlined numbers are results of NIDSs required in order to realize any rule of the five.

In this way, larger $k$ requires larger $m$ (more SMUs), and thus, degrades operating frequency because of longer critical paths in the feedback array. However, it can improve transmission speed significantly since the positive effect due to parallelization by $k$ is larger than its negative effect.

## VI. CONCLUSION AND COMMENTS

This paper proposed a programmable NIDS based on a multi-byte STNFA. By using multi-byte STNFAs, the proposed NIDS achieved more than 10 Gbps of network transmission speed that is difficult for existing programmable NIDSs to achieve. Thus, the proposed NIDS can avoid bottleneck even in latest Gigabit network. Since in the proposed NIDS, regular expression patterns can be set by just rewriting contents of memories and registers, the NIDS achieves both fast regular expression matching and quick pattern updating.

In the current design, increasing the number of SMUs makes critical paths in the feedback array longer, resulting in degradation of operating frequency. Thus, improving architecture of the feedback array is one of our future works. We will also study how to minimize the number of state transitions in order to reduce hardware cost.

## ACKNOWLEDGMENTS

This research is partly supported by the JSPS Grant-in-Aid for Scientific Research (C), (No. 26330691), 2015.

## REFERENCES

[1] J. Aoe, Computer Algorithms: String Pattern Matching Strategies, IEEE Computer Society Press, 1994.

[2] AV-TEST - The Independent IT-Security Institutie, http://www.av-test.org/.

[3] M. Becchi and P. Crowley, “An improved algorithm to accelerate regular expression evaluation,“ Proc. of the 3rd ACM/IEEE Symposium on Architecture for networking and communications systems, Dec. 2007, pp. 145–154.

[4] J. Bispo, I. Sourdis, J. M. P. Cardoso, and S. Vassiliadis, “Regular expression matching for reconfigurable packet inspection,” Proc. 2006 IEEE International Conference on Field Programmable Technology, 2006, pp. 119–126.

[5] J. Divyasree, H. Rajashekar, and K. Varghese, “Dynamically reconfigurable regular expression matching architecture,” Proc. International Conference on Application-Specific Systems, Architectures and Processors (ASAP 2008), July 2008, pp. 120–125.

[6] D. Ficara, et al., “An improved DFA for fast regular expression matching,“ ACM SIGCOMM Computer Communication Review, Vol. 38, No. 5, Oct. 2008, pp. 31–40.

[7] D. Ficara, et al., “Differential encoding of DFAs for fast regular expression matching,“ IEEE/ACM Transactions on Networking, Vol. 19, No. 3, June 2011, pp. 683–694.

[8] T. Ganegedara, Y.E. Yang, and V. K. Prasanna, “Automation framework for large-scale regular expression matching on FPGA,“ Proc. 2010 IEEE International Conference on Field Programmable Logic and Applications, 2010, pp. 50–55.

[9] J. E. Hopcroft, J. D. Ullman, and R. Motwani, Introduction to Automata, Theory, Languages and Computation, Second Edition, Addison-Wesley, 2000.

[10] B. L. Hutchings, R. Franklin, and D. Cover, “Assisting network intrusion detection with reconfigurable hardware,” Proc. 10th Annual IEEE Symposium on Field-Programmable Custom Computing Machines, 2002, pp. 111–120.

[11] Y. Kaneta, S. Yoshizawa, S. Minato, H. Arimura, and Y. Miyanaga, “Dynamic reconfigurable bit-parallel architecture for large-scale regular expression matching,” Proc. 2010 IEEE International Conference on Field Programmable Technology, Dec. 2010, pp. 21–28.

[12] S. Kumar, S. Dharmapurikar, F. Yu, P. Crowley, and J. Turner, “Algorithms to accelerate multiple regular expressions matching for deep packet inspection,” Proc. SIGCOMM’06, 2006, pp. 339–350.

[13] A. Mitra, W. Najjar, and L. Bhuyan, “Compiling PCRE to FPGA for accelerating SNORT IDS,” Proc. 2007 ACM/IEEE Symposium on Architecture for Networking and Communications Systems, Dec. 2007, pp. 127–136.

[14] G. Navarro and M. Raffinot, Flexible Pattern Matching in Strings, Cambridge University Press, 2002.

[15] H. C. Roan, W. J. Hwang, and C. T. Dan Lo, “Shift-or circuit for efficient network intrusion detection pattern matching,” Proc. International Conference on Field Programmable Logic and Applications, 2006, pp. 785–790.

[16] Y. SangKyun and L. KyuHee, “Optimization of regular expression pattern matching circuit using at-most two-hot encoding on FPGA,“ Proc. 2010 IEEE International Conference on Field Programmable Logic and Applications, Sept. 2010, pp. 40–43.

[17] R. Sidhu and V. K. Prasanna, “Fast regular expression matching using FPGAs,” Proc. 2001 IEEE International Symposium on Field-Programmable Custom Computing Machines, 2001, pp. 227–238.

[18] R. Smith, C. Estan, S. Jha, and S. Kong, “Deflating the big bang: fast and scalable deep packet inspection with extended finite automata,” Proc. SIGCOMM’08, 2008, pp. 207–218.

[19] Sourcefire Inc., “SNORT network intrusion detection system,” http://www.snort.org/.

 ISBN: 978-1-61208-496-1

[20] H. Takaguchi, Y. Wakaba, S. Wakabayashi, S. Nagayama, and M. Inagi, “An NFA-based programmable regular expression matching engine highly suitable for FPGA implementation,” 18th Workshop on Synthesis and System Integration of Mixed Information Technologies (SASIMI’13), 2013, pp. 231–236.

[21] K. Thompson, “Programming technique: regular expression search algorithm,” Communications of the ACM, Vol. 11, No. 6, June 1968, pp. 419–422.

[22] Y. Wakaba, M. Inagi, S. Wakabayashi, and S. Nagayama, “An efficient hardware matching engine for regular expression with nested Kleene operators,“ Proc. 2011 IEEE International Conference on Field Programmable Logic and Applications, 2011, pp. 157–161.

[23] N. Yamagaki, R. Sidhu, and S. Kamiya, “High-speed regular expression matching engine using multi-character NFA,” Proc. International Conference on Field Programmable Logic and Applications, Aug. 2008, pp. 131–136.

[24] Y.-H. E. Yang and V. Prasanna, “Automatic construction of large-scalce regular expression matching engines on FPGA,” Proc. 2008 International Conference on Reconfigurable Computing and FPGAs, 2008, pp. 73–78.

[25] Y.-H. E. Yang and V. K. Prasanna, “Space-time trade off in regular expression matching with semi-deterministic finite automata,” Proc. IEEE INFOCOM 2011, April 2011, pp. 1853–1861.

 ISBN: 978-1-61208-496-1 

# A Dynamically Reconfigurable NoC for Double-Precision Floating-Point FFT on FPGAs

Thanh Thi Thanh Bui, Braden Phillips, and Michael Liebelt

School of Electrical and Electronic Engineering
The University of Adelaide, Adelaide, Australia
Email: {thanh.bui, braden.phillips, michael.liebelt}@adelaide.edu.au

***Abstract*—This paper presents a dynamically partially reconfigurable network on chip (NoC) on a field-programmable gate array (FPGA) for double-precision floating-point Fast Fourier Transforms (FFTs). This is one of the first published examples of a practical system using a dynamically reconfigurable NoC that has been implemented in existing FPGA technology. Up to 16 parallel double-precision floating-point processing elements (PEs) can be implemented on the FPGA. Using dynamic partial reconfiguration, a user can change the number of running PEs to choose an optimal power-performance operating point. The design provides much better performance than i7-3.4GHz CPUs running Matlab and competitive performance with static-only FFT systems and the Xilinx FFT IP core, but it has the advantage of saving power and releasing hardware resources when maximum FFT performance is not required. With all 16 PEs running, the design can process an FFT of up to 131072 points and achieves its maximum throughput of of 33.5 FLOPs/cycle on a Xilinx Virtex-7 XC7VX485T FPGA.**

***Keywords*—Network-on-chip, partial reconfiguration, floating point, FFT, parallel architecture, FPGA.**

## I. INTRODUCTION

Dynamically partially reconfigurable FPGAs allow hardware modules to be placed and removed at runtime while other parts of the system keep working [1]. This permits a radical departure from the way application-specific hardware is usually designed. In a static system, there must be a fixed set of processing resources sufficient to meet performance requirements under worst-case load conditions. If the workload changes, processing resources sit idle. A system that moves through modes with distinctly different processing needs, should provide different resources for each mode. Dynamic partial reconfiguration can: reduce power consumption by removing resources not currently required; achieve better utilization by changing the mix of processing resources as the requirements of the system change; and deliver better performance by using heterogeneous processing resources optimized for particular stages in an algorithm, rather than making do with static generic processing resources that must serve all stages.

To exploit this new capability, there is a need for efficient, dynamically adaptive communication infrastructure that automatically adapts as modules are added to and removed from the system. Many network-on-chip architectures have been proposed in the last decade to exploit dynamic reconfiguration on FPGA technology. Examples include DyNoC (Dynamic Network on Chip) [2], CuNoC (Communication Unit Network on Chip) [3], CoNoChi (Configurable Network on Chip) [4], DRNoC (Dynamic Reconfigurable Network on Chip) [5] and OCEAN (On-Chip Efficiently Adaptive Network) [6]. However, most of these have been described in theory only or evaluated using general traffic models. Few have been realized in a practical application. To the best of our knowledge, only DyNoC has been applied to a traffic light controller and CoNoChi has been demonstrated in a dynamically reconfigurable network coprocessor called DynaCORE. However, they have not been fully realized and validated.

While dynamic reconfiguration offers clear benefits in theory, more application experiments are required to understand the benefits and limitations of dynamically reconfigurable NoCs and to guide their further development. The aim of this paper is to begin to address this research gap by using an adaptive NoC to connect parallel Processing Elements (PEs) in a dynamically reconfigurable implementation of the Fast Fourier Transform (FFT) on an FPGA. We have chosen to begin with the FFT because it is widely used in a diverse variety of applications in engineering, science and mathematics [7]. It is also commonly implemented using FPGAs, which can exploit parallel hardware to achieve power-efficient, high-speed performance.

The main contributions of this paper are:

1) To the best of our knowledge, this is the first publication that fully realizes a dynamically partially reconfigurable NoC in a specific application, a double-precision floating-point FFT. The system is implemented on an FPGA platform and evaluated for latency, area and power consumption. We compare it with static-only systems, software implementation on CPUs and Xilinx FFT IP core.
2) We show that using dynamic partial reconfiguration of the FPGA allows the user to efficiently change between power-performance operating points while still maintaining competitive performance with the static-only systems. In fact, the real-time performance of the system is among the fastest FFTs published so far.

The rest of the paper is organized as follows. In Section II, an overview of the FFT algorithm and its hardware implementation are presented. The proposed FFT architecture is detailed

 ISBN: 978-1-61208-496-1 

in Section III. Section IV presents the implementation results and design performance. Finally, the paper is concluded in Section V.

## II. FFT Algorithm and Hardware Implementation

In this paper, we use the radix-2 decimation in frequency variation of the FFT Cooley-Tukey algorithm [7]. This algorithm is composed of butterfly operations, as in Figure 1. The signals are represented by complex variables and for our implementation, they all use double-precision floating point representation. Radix-2 is chosen due to its simplicity and flexibility. Figure 2 shows an example of these butterfly operations arranged in a decimation in frequency architecture.

Many different hardware architectures for the FFT have been proposed. These include the parallel architecture [8], pipelined architecture [9]–[12], and combined parallel-pipelined architecture [13][14]. In this paper, the parallel architecture is chosen in preference to the more common pipelined architecture as it proves a good match with the dynamic NoC approach, achieving efficient hardware utilization without excessive memory bandwidth requirements.

## III. Dynamically Reconfigurable FFT Architecture

### A. Processing Elements

A PE performs a radix-2 butterfly operation. Each PE consists of: four floating-point multipliers and six floating-point adders as in Figure 3 for the full radix-2 butterfly operation; memory storage for input data, twiddle phase factors and intermediate results; a local controller; and a NoC interface. The local controller coordinates the read/write memory operation and defines which PE to communicate with at each stage. The NoC interface works as a bridge between the PE and the NoC. The floating-point multipliers and adders support IEEE 754 double-decision floating-point normal and abnormal numbers. They are deeply pipelined to improve speed.

The floating-point multiplier is based on the design in [15]. Its operation includes multiplying the mantissas, adding the exponents, calculating the result sign, normalizing and finally rounding the result according to the IEEE 754 double-decision floating-point standard. The multiplier consists of 17 pipeline stages.

The floating-point adder is also based on [15]. It consists of several steps as follows: ensure that the exponents of the two operands are equal by increasing the smaller one and shifting right its corresponding mantissa; add/subtract the mantissas if they have the same/opposite signs; normalize and finally round

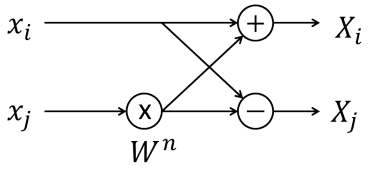



Figure. 1. Radix-2 butterfly datapath.

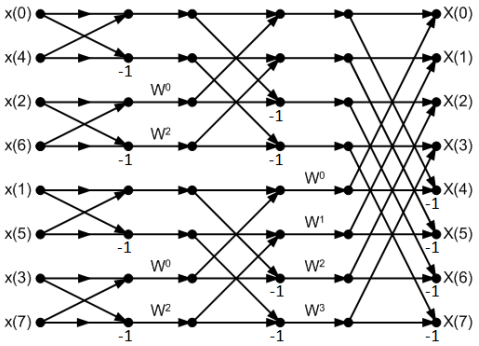



Figure. 2. Decimation in frequency diagram of 8-point radix-2 FFT.

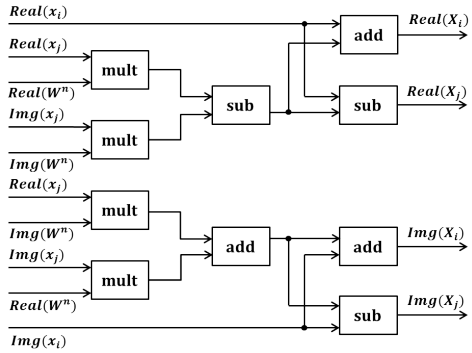



Figure. 3. Radix-2 butterfly implementation.

the result according to the IEEE 754 double-decision floating-point standard. The adder is implemented with 10 pipeline stages. The subtractor can be simply implemented using an adder by inverting the sign of the second operand.

The on-chip memory storage of each PE is determined according to the number of block RAMs available on the FPGA and the number of PEs implemented. A single PE synthesized for a Xilinx Virtex-7 XC7VX485T FPGA occupies 8739 register slices, 13022 LUT slices and 36 DSP48E1 slices. This accounts for 1%, 4% and 1% of the available registers, LUT and DSP slices respectively. The PE achieves a maximum operating frequency of 364 MHz. Up to 25 PEs could be implemented on the device but the number of PEs in a parallel architecture must be a power of two; hence only 16 PEs are implemented. There are a total of 1030 x 36Kb block RAMs; hence each PE can include up to 64 block RAMs. The total on-chip memory required for an $N$-point FFT is approximately $2N$ elements (including data elements and twiddle phase elements). With elements of 128 bits for double precision complex numbers (64 bits for the real part and 64 bits for the imaginary part), the maximum FFT size the design can process is $(1030 \times 36 \times 1024)/(2 \times 128) = 148320$ points. The actual maximum size is 131072 ($2^{17}$) points due to the use of a radix-2 design.

### B. Data Scheduling

Data scheduling is based on the decimation-in-frequency variation of the FFT Cooley-Tukey algorithm. The twiddle

phase factors are pre-calculated and stored in the external ROM. The input data and twiddle phase factors are loaded from the external memory to the on-chip memory of each PE by a global controller. To ensure the parallel operation of PEs, all input elements are transferred to the PEs before they start their operation. The number of elements transferred per cycle depends on the bandwidth of external memories. In this case, one element is transferred per cycle. If the system runs at 200 MHz, the transfer of one 128-bit element per cycle requires a memory bandwidth of 25.6 Gbps. This can be easily obtained by current memory technology.

With an $N$-point FFT and $P$ PEs, $\mathrm{PE}_i$ $(0 \leq i \leq P-1)$ is scheduled with the $\{x_j : Ni/P \leq j \leq N(i+1)/P-1\}$. With this scheduling strategy, there is no communication among the PEs in the first $(log_2(N/P)-1)$ stages; PEs feedback their own results to their inputs for the next stage. In the next $log_2 P$ stages, PEs transfer their results to each other as given in Table I. At each stage, the twiddle phase factors are read in a manner so that they can be reused. Communications among PEs and between PEs and external memory use the NoC.

### C. Network-on-chip Communication

The FFT implementation involves complex routing infrastructure for transferring input data and twiddle phase factors from external memory to the PEs, transferring intermediate results among PEs and outputting the final results from each PE to the external memory. These can all be efficiently implemented by a light-weight 2-D mesh NoC. The modularity and scalability of a NoC can reduce interconnect routing complexity and maintains sustainable performance. The NoC uses small routers, circuit switching and deterministic routing. Circuit switching is used for flow control because it does not require input buffers on the NoC routers. Deterministic routing is used to maximize the throughput of the NoC.

Each router has five input/output ports as in Figure 4. Four from the four cardinal directions (North, East, South and West) and one from the local PE. There are no input buffers at the input ports. Each output port consists of an arbiter and a crossbar switch. The arbiter determines which input port is selected to proceed in the next stage. Then, the crossbar switch matches the successful input port with the desired output port. Each router is connected to its neighbors and local PE through a bidirectional link with 130 bits for each direction. The synthesized single router occupies 758 register slices and 1442 LUT slices, which is relatively small compared with 8739 register slices and 13022 LUT slices for a PE. The router achieves a maximum operating frequency of 522 MHz, which is much higher than 364 MHz for the PE.

In circuit switching flow control, a link between source and destination PEs is set up before a transfer is performed and it is maintained throughout the transfer. Data transferred in the NoC is in 130-bit flit format. When a PE wants to communicate to another, it will assert a Send signal and send a set-up flit to its target PE through the NoC. If the target PE accepts the request, it will assert an Accept signal and a channel is constructed between the two PEs. Data is transferred and the channel is released when the last flit is received. There are 4 types of flits defined by the two most significant bits: set-up flits, data flits, tail flits and control flits.

In the deterministic routing algorithm, the route is defined by the source PE and stored in the set-up flit, beginning with the least significant bit. The routing information is a set of 2-bit directions, beginning with the source to the destination: 00 for North, 01 for South, 10 for East and 11 for West. When a router receives a set-up flit, it extracts the two least significant bits to find out the routing direction and then shifts the set-up flit two bits to the right before sending the flit to the next router. When the output direction is equal to the input direction, the flit is transferred to its attached PE. To ensure the highest throughput and no collision of the NoC, the position of each PE is given in Figure 5.

### D. The Dynamic Partial Reconfiguration System

To support dynamic partial reconfiguration, the system is divided into two areas: a static area with functionality unchanged during system operation; and a dynamic area. This arrangement is shown in Figure 5. The dynamic area is divided into partial reconfiguration regions which can be configured as a PE or a router. Each region must contain sufficient resources−such as slices, block RAMs, and DSP slices−to implement the modules assigned to it.

Based on the user's power-performance requirement, 1, 2, 4, 8 or 16 PEs and their attached routers can be implemented in the dynamic region. The maximum FFT size that each design variation can handle is given in Table II. With dynamic partial reconfiguration, time and power consumption for the reconfiguration process can be reduced significantly and the system can keep running during reconfiguration. For example, when the design is running with 4 PEs and the user wants to upgrade to 8 PEs, only 4 additional PEs are configured. This can save half of the time and power consumption required to

TABLE I
COMMUNICATIONS AMONG PES IN EACH STAGE

| Stage | Communications among PEs |
|---|---|
| $log_2(N/P)$ | $PE_i \rightleftharpoons PE_{i+1}$ $(i = 2j, 0 \leq j \leq 7)$ |
| $log_2(N/P)+1$ | $PE_i \rightleftharpoons PE_{i+2}$ $(i = 0, 1, 4, 5, 8, 9, 12, 13)$ |
| $log_2(N/P)+2$ | $PE_i \rightleftharpoons PE_{i+4}$ $(i = 0, 1, 2, 3, 8, 9, 10, 11)$ |
| $log_2(N/P)+3$ | $PE_i \rightleftharpoons PE_{i+8}$ $(0 \leq i \leq 7)$ |

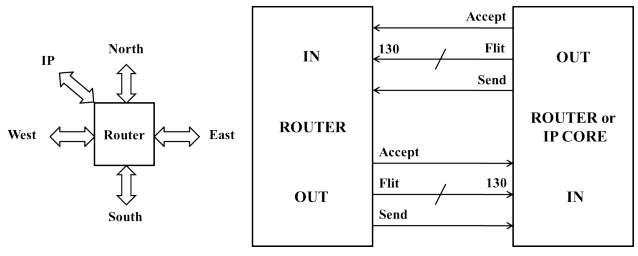



Figure. 4. Router datapath and interface.

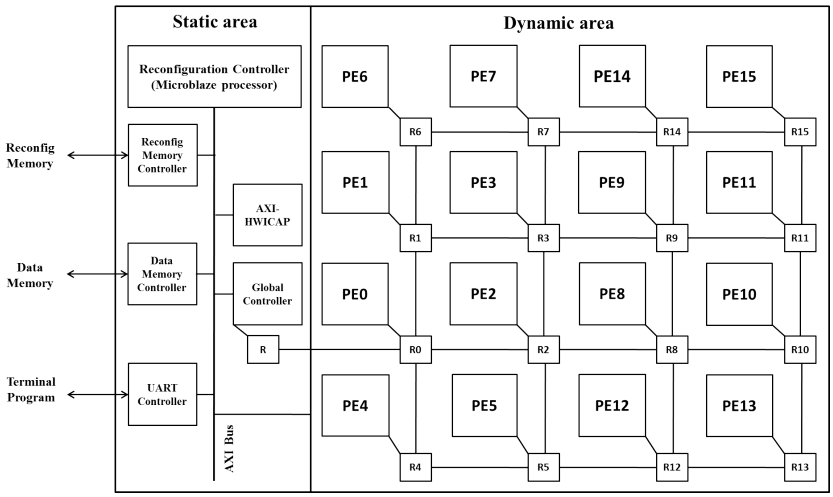



Figure. 5. FFT architecture.

fully configure a system with 8 PEs. In addition, the 4-PE design can keep running until the 8-PE design is completely configured. It is much simpler when the design is changed from a large number of PEs to a smaller one. In this case, the unused PEs are reconfigured with blanking bitstreams.

The static area contains a configuration controller, a configuration port, interfaces with external memories, a serial interface with an external host, and the global controller. The reconfiguration controller is responsible for loading the partial configuration bitstreams stored in the configuration memory through the configuration port. We chose to implement it using a MicroBlaze soft-core processor on the FPGA. The configuration port transmits the configuration data to the assigned region. The internal configuration access port (ICAP) primitive is used in this case since it provides the access to the configuration logic of the FPGA from within the FPGA fabric. The terminal program with the external host allows users to select the number of running PEs. The global controller is responsible for data scheduling and informing PEs when the design is changed.

## IV. Implementation Results and Design Performance

### A. Implementation Results

The design has been implemented in Verilog and verified using Modelsim. Synthesis and power analysis were performed with the Xilinx ISE Design Suite, targeting a Xilinx Virtex-7 XC7VX485T FPGA. Table III shows the results. Synthesis and power analysis of static-only FFT systems are also performed for comparison. Static-only FFT systems here are five design variations (1, 2, 4, 8, 16 PEs) based on point-to-point connections only. Figure 6 compares the power consumption of five differently sized static configurations with that of the new dynamically reconfigurable design with different numbers of active PEs.

Without dynamic partial reconfiguration, the user can choose between two options. The first one is using a fixed design of 16 PEs to perform all FFT sizes. This option is simple and does not require hardware reconfiguration but is not power-efficient since the biggest design is used for all FFT sizes. It can be seen from Figure 6 that the power consumption of the static 16-PE design is only smaller than the dynamic 16-PE design and much higher than the rest all four dynamic design variations.

The second option is using a static design appropriate for each FFT size. The system is fully reconfigured when changing between design variations. With this option, each design variation is smaller and more power-efficient than the corresponding variation with dynamic partial reconfiguration. However, the time and power consumption for reconfigura-

TABLE II
Maximum FFT Size of Each Design Variation

| Design Variation | Maximum FFT size |
|---|---|
| 1 PE | 8192 ($2^{13}$) |
| 2 PEs | 16384 ($2^{14}$) |
| 4 PEs | 32768 ($2^{15}$) |
| 8 PEs | 65536 ($2^{16}$) |
| 16 PEs | 131072 ($2^{17}$) |

TABLE III
Synthesis and Power Analysis Results

| | A single PE | A single router |
|---|---|---|
| Number of Slice Registers | 8739 | 758 |
| Number of Slice LUTs | 13022 | 1442 |
| Number of DSP48E1s slices | 36 | 0 |
| Number of 36Kb Block RAMs | 62 | 0 |
| Maximum Frequency | 364.458MHz | 522.575MHz |
| Estimated Power at 200MHz | 1.418W | 0.443W |

 ISBN: 978-1-61208-496-1

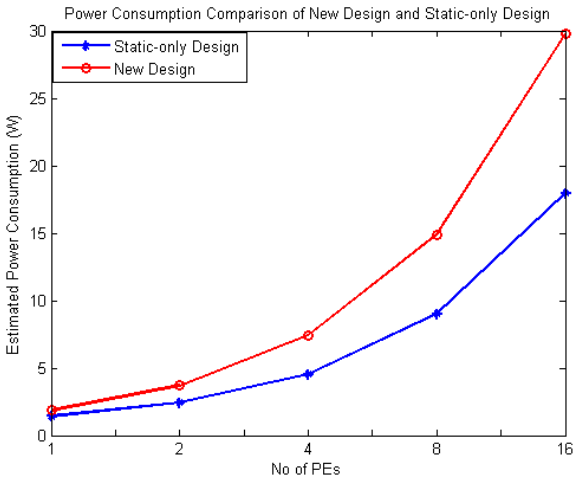



Figure. 6. Power comparison of new design and static-only design

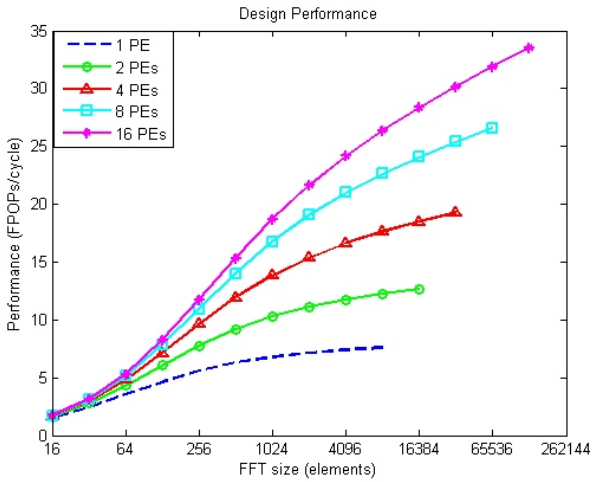



Figure. 7. Performance comparison of different design variations.

tion process is higher, which is a disadvantage for real-time applications.

Partial reconfiguration is performed using Xilinx PlanAhead. The partial bitstream of a single PE is 1212000 Bytes and that of a single router is 132512 Bytes. If the ICAP interface is 32-bit wide and clocked at 100 MHz, the configuration time of a single PE and a single router is 3.03 ms and 332 $\mu$s respectively. The partial reconfiguration time of $n$ PEs is 3.362$n$ ms.

### B. Design Performance

The total number calculation cycles of the dynamic system is slightly higher than the static system due to the NoC latency. The NoC latency (L) can be calculated as:

$$L = S + I + T \quad (1)$$

Here, $S$ is the link set-up time. It depends on the hop number ($h$) of the link. For each hop, it takes two cycles to process a Send signal and one cycle to process an Accept signal, hence $S = 3h$. The second term, $I$, is the initial latency, which is equivalent to the hop number. The last term, $T$, is the transfer time, which is the largest contribution to the NoC latency. The transfer time is equivalent to the number of flits in each packet. However, it is overlapped with the calculation time. Therefore, the NoC latency in this case is only $4h$, which is insignificant compared to the total number of calculation cycles.

The performance of the design in terms of floating point operations per cycle (FPOPs/cycle) with different numbers of PEs is as shown in Figure 7. Computing an $N$-point FFT takes $5Nlog_2N$ floating-point operations. With the number of calculation cycles ($C$) obtained from simulation, the performance of the design is $5Nlog_2N/C$ (FPOPs/cycle). It can be seen that the performance increases with an increase in the number of running PEs and the FFT size. The highest performance of 33.5 FPOPs/cycle is achieved by the 16-PE design with the 131072-point FFT. However, there is an insignificant difference in the performance of different designs when the FFT size is smaller than 256 points. This difference increases with larger FFT size. In other words, for small FFT size, a small number of PEs should be used. This will ensure power efficiency while obtaining satisfactory performance. For larger FFTs, users can select a design between different power-performance operating points.

A performance comparison in terms of calculation time between two design variations (1 PE and 16 PEs) running at 200 MHz and an Intel i7-3.4GHz CPU running Matlab R2014 is as shown in Figure 8. The calculation time of the Matlab program is determined using the 'tic' and 'toc' functions of Matlab. It can be seen that the calculation time of both design variations is much smaller that of the i7 CPU. The dynamically reconfigurable 16-PE design finishes the 131072-point FFT in 1.66 ms while the i7 CPU takes more than 6 ms. This proves that FPGAs can outperform software implementation on a general purpose processors for the FFT.

The performance of the design is also compared with the Xilinx FFT IP core [16] in terms of calculation cycles. The Xilinx FFT IP core supports the transformed size of $N = 2^m$ $(3 \leq m \leq 16)$, data precision of 8 to 34 bits, and fixed-point and block floating-point data format. The core provides four architectures: Radix-2, Radix-2 Lite, Radix-4 and Pipelined Radix-2. The Radix-2 Lite and Radix-4 architectures are not considered here because they have different butterfly structures, either lighter or heavier than the Radix-2.

The new 1-PE design has similar performance to that of the Xilinx Radix-2 FFT core. For example, the 1-PE design processes a 1024-point FFT in 7538 cycles; and the Xilinx Radix-2 FFT core needs 7367 cycles. The performance comparison between the dynamically reconfigurable 16-PE design and the Xilinx Pipelined Radix-2 FFT core is shown in Figure 9. However in making this comparison it should be noted that the number of PEs in the Xilinx Pipelined Radix-2 architecture in this case is smaller than 16. For example, only 13 PEs are used in the Pipelined Radix-2 architecture to perform the 8192-

 ISBN: 978-1-61208-496-1 

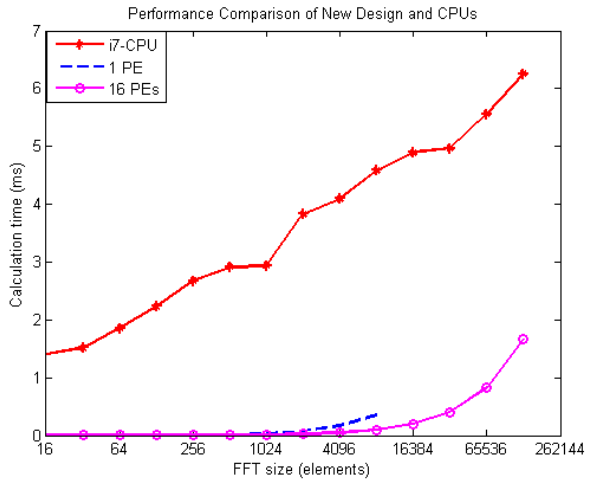



Figure. 8. Performance comparison of new design and CPUs.

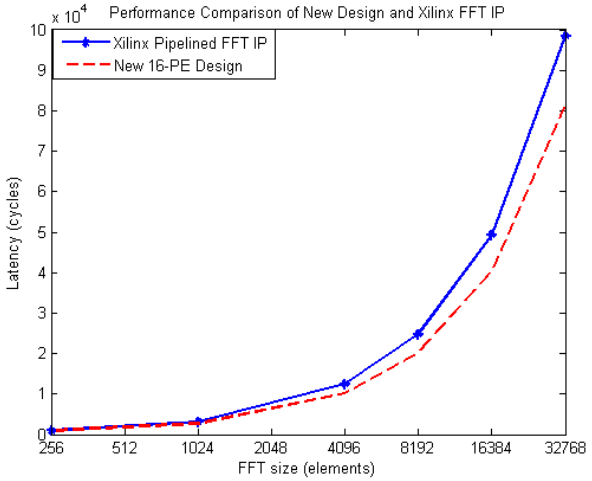



Figure. 9. Performance comparison of new design and Xilinx FFT IP.

point FFT. This comparison aims to illustrate the advantage of the parallel architecture over the pipelined architecture, in which more than $n$ PEs can be used to perform a $2^n$-point FFT to achieve smaller latency.

## V. Conclusion

In this paper, we have presented a practical system using a dynamically reconfigurable network on chip. Our implementation demonstrates that dynamically reconfigurable NoCs are feasible and can deliver power and performance benefits. In this case power is saved by only implementing sufficient hardware resources to meet current processing requirements. We have also shown how a dynamically reconfigurable NoC can be realized within the constraints of existing FPGA technology.

Our system, a double-precision floating-point FFT unit using network-on-chip communication and dynamic partial reconfiguration has been implemented on a Xilinx Virtex-7 XC7VX485T FPGA. The design provides much better performance than i7-3.4GHz CPUs running Matlab and competitive performance with static-only FFT systems and a Xilinx FFT IP core but with double-precision floating-point data format and the ability to adapt power-performance to suit the current workload. It can be concluded that the use of a dynamically partially reconfigurable NoC is a feasible and potentially beneficial solution for systems with all homogeneous PEs like the FFT system. This allows the system to scale performance at an acceptable cost of additional on-chip hardware. However, effectively exploiting dynamically reconfigurable systems requires a change in design practice. Designers conventionally craft algorithms to make good utilization of a fixed set of processing resources. Future work to demonstrate the benefit of dynamic reconfiguration will need to begin with algorithms redesigned for the new approach, which can take advantages of a changing set of heterogeneous processing elements.

## References

[1] Xilinx, "Partial Reconfiguration User Guide, UG702, V14.5," 2013.

[2] C. Bobda *et al.*, "DyNoC: A Dynamic Infrastructure for Communication in Dynamically Reconfugurable Devices," in International Conference on Field Programmable Logic and Applications 2005, pp. 153–158.

[3] S. Jovanovic, C. Tanougast, S. Weber, and C. Bobda, "CuNoC: A Scalable Dynamic NoC for Dynamically Reconfigurable FPGAs," in International Conference on Field Programmable Logic and Applications 2007, pp. 753–756.

[4] T. Pionteck, C. Albrecht, R. Koch, and E. Maehle, "Adaptive Communication Architecture for Runtime Reconfigurablle System-on-Chips," Parallel Processing Letters, vol. 18, no. 02, 2008, pp. 275–289.

[5] Y. E. Krasteva, E. de la Torre, and T. Riesgo, "Reconfigurable Networks on Chip: DRNoC Architecture," Journal of Systems Architecture, vol. 56, no. 7, 2010, pp. 293 – 302. Special Issue on HW/SW Co-Design: Systems and Networks on Chip.

[6] L. Devaux and S. Pillement, "OCEAN, A Flexible Adaptive Network-on-Chip for Dynamic Applications," Microprocessors and Microsystems, vol. 38, no. 4, 2014, pp. 337 – 357. Selected Papers from Euromicro Conference on Parallel, Distributed and Network-based Processing (PDP 2012).

[7] E. O. Brigham, The Fast Fourier Transform and Its Applications. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1988.

[8] J. Palmer and B. Nelson, "A Parallel FFT Architecture for FPGAs," in Field Programmable Logic and Application (J. Becker, M. Platzner, and S. Vernalde, eds.), vol. 3203 of Lecture Notes in Computer Science, 2004, pp. 948–953, Springer Berlin Heidelberg.

[9] S. He and M. Torkelson, "Design and Implementation of a 1024-Point Pipeline FFT Processor," in Proceedings of the IEEE Custom Integrated Circuits Conference 1998, pp. 131–134.

[10] Y. Jung, H. Yoon, and J. Kim, "New Efficient FFT Algorithm and Pipeline Implementation Results for OFDM/DMT Applications," IEEE Transactions on Consumer Electronics, vol. 49, no. 1, Feb 2003, pp. 14–20.

[11] M. Garrido, J. Grajal, M. Sanchez, and O. Gustafsson, "Pipelined Radix-$2^k$ Feedforward FFT Architectures," IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 21, no. 1, Jan 2013, pp. 23–32.

[12] Y. N. Chang and K. Parhi, "An Efficient Pipelined FFT Architecture," IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, vol. 50, no. 6, June 2003, pp. 322–325.

[13] M. Ayinala and K. Parhi, "Parallel Pipelined FFT Architectures with Reduced Number of Delays," in Proceedings of the Great Lakes Symposium on VLSI 2012, GLSVLSI '12, pp. 63–66, ACM.

[14] J. You and S. Wong, "Serial-Parallel FFT Array Processor," IEEE Transactions on Signal Processing, vol. 41, no. 3, Mar 1993, pp. 1472–1476.

[15] Z. Jovanovic and V. Milutinovic, "FPGA Accelerator for Floating-Point Matrix Multiplication," Computers Digital Techniques, IET, vol. 6, no. 4, July 2012, pp. 249–256.

[16] Xilinx, "LogiCORE IP Fast Fourier Transform v8.0," 2012.

 ISBN: 978-1-61208-496-1 