C O M M U N I C A T I O N S

FACULTY OF SCIENCES
UNIVERSITY OF ANKARA

DE LA FACULTE DES SCIENCES
DE L'UNIVERSITE D'ANKARA

Series A2-A3: Physical Sciences and Engineering

VOLUME: 61 **Number: 2** **YEAR: 2019**

Faculy of Sciences, Ankara University
06100 Beşevler, Ankara-Turkey
ISSN 1303-6009 E-ISSN 2618-6462

C O M M U N I C A T I O N S

FACULTY OF SCIENCES
UNIVERSITY OF ANKARA

DE LA FACULTE DES SCIENCES
DE L'UNIVERSITE D'ANKARA

Series A2-A3: Physical Sciences and Engineering

Volume 61 **Number : 2** **Year :2019**

This Journal is published two issues in a year by the Faculty of Sciences, University of Ankara. Articles and any other material published in this journal represent the opinions of the author(s) and should not be construed to reflect the opinions of the Editor(s) and the Publisher(s).

Correspondence Address:
COMMUNICATIONS
EDITORIAL OFFICE
Ankara University, Faculty of Sciences,
06100 Tandoğan, ANKARA – TURKEY
Tel: (90) 312-212 67 20 Fax: (90) 312-223 23 95
e-mail: commun@science.ankara.edu.tr
http://communications.science.ankara.edu.tr/index.php?series=A2A3

Print:
Ankara University Press
İncitaş Sokak No:10 06510 Beşevler
ANKARA – TURKEY

C O M M U N I C A T I O N S

FACULTY OF SCIENCES
UNIVERSITY OF ANKARA

DE LA FACULTE DES SCIENCES
DE L'UNIVERSITE D'ANKARA

Series A2-A3: Physical Sciences and Engineering

Volume 61 **Number : 2** **Year :2019**

Commun.Fac.Sci.Univ.Ank.Series A2-A3
Volume 61, Number 2, Pages 129-149 (2019)
DOI: 10.33769/aupse.529575
ISSN 1303-6009 E-ISSN 2618-6462
http://communications.science.ankara.edu.tr/index.php?series=A2A3

A COMPARISON OF DEEP LEARNING BASED ARCHITECTURE WITH A CONVENTIONAL APPROACH FOR FACE RECOGNITION PROBLEM

Fatima Zehra UNAL and Mehmet Serdar GUZEL

ABSTRACT. This paper addresses a new approach for face recognition problem based on deep learning strategy. In order to verify the performance of the proposed approach, it is compared with a conventional face recognition method by using various comprehensive datasets. The conventional approach employs Histogram of Gradient (HOG) algorithm to extract features and utilizes a multi-class Support Vector Machine (SVM) classifier to train and learn the classification. On the other hand, the proposed deep learning based approaches employ a Convolutional Neural Network (CNN) based architecture and also offer both a SVM and Softmax classifiers respectively for the classification phase. Results reveal that the proposed deep learning architecture using Softmax classifier outperform conventional method by a substantial margin. As well as, the deep learning architecture using Softmax classifier also outperform SVM in almost all cases.

1. INTRODUCTION

Face recognition, covering a large number of fields and disciplines such as safety and commercial applications, is an important research problem and gathers lots of attention from researchers. Deep Learning technology has dominated the field and made great progress in solving problems that have not been achieved with applications developed for a long time in artificial intelligence field. It is clear that deep learning and related technologies have improved the overall success performance of many classification problems in the field of computer vision. Especially, the adaptation of Convolutional Neural Networks (CNN) architecture to

Received by the editors: February 20, 2019; Accepted: June 15, 2019.
Key word and phrases: Convolutional neural network, deep learning, face recognition, fine tuning, softmax, svm.

computer vision problems has opened a new era at the field that achieved promising results in computer vision based applications.

Face recognition applications can be categorized into two main groups namely, security and commercial applications. A good example for security applications is real-time mapping according to video image sequences that employs previously recorded images to detect and recognize criminals and prevent unauthorized people entering restricted zones. Static mapping from credit card images, passports, driver's license and identity cards is a good example for commercial applications. It essentially provides real time transactions based on image or video sequences [1]. Face recognition applications have become enormously significant since they have offered successful results in the field of security. Essentially for such security systems, Machine learning algorithms play a crucial role, especially in recognition and verification tasks. Artificially recognizing people faces can be performed through supervised learning mechanisms by employing predefined features for training. However, this learning technique can only be successfully applied when the faces are captured in well-defined conditions. On the other hand, recognition becomes quite difficult when an irrepressible situation occurs, such as changes in the face expression or head directions, as well as lighting conditions is also crucial. The only way to overcome these problems is to employ a reliable feature extraction algorithm comprising consistent enhancement and restoration steps [2]. Deep learning based algorithms can be able to learn automatically to extract the needed properties to train a new classifier to be used to solve a different problem.

Deep Learning Technologies accomplished great progress in solving problems that have not been achieved with applications developed in the field of machine learning for many years. Deep learning technologies are able to explore the complex structures of high-dimensional data that has been applied in many areas from image and speech processing to classification and regression problems [3,4,5]. The processing power provided by the graphics processing unit (GPU) allows Deep Learning technologies to employ extensive amounts of data to train deeper or more advanced models with respect to the increased processing power.

Deep learning is essentially a multi-layer artificial neural network-based machine learning technique. The main advantage of deep learning is that layers of features are not obtained from conventional feature extractor algorithms, instead they are learned automatically from data using multilayer network hierarchy [3, 6]. The higher layers of a deep learning architecture strengthen the characteristic properties of the input given to the network and defeats the irrelevant properties. For example,

it is assumed that an image consists of pixel values and simple shapes is employed as input to the deep learning architecture. The extracted properties in the first network layer most probably signify the presence or absence of edges in certain directions and positions in the image. The second layer, on the other hand, identifies patterns by detecting special arrangements of edges by considering negligible minor changes in edge locations. The third layer can estimate larger combinations or more complex patterns of familiar objects, and essentially consequent layers may perceive concrete objects as combinations of these parts. Consequently, the network first learns the raw primitive edges, followed by learning more complex shapes based on the edges it has previously learned, and learns more advanced features using those shapes. This hierarchical structure allows the architecture to extract features in a systematic manner. Deep learning models have the ability to learn to focus on the right features automatically, and therefore require little guidance from the designer to intervene feature extraction process [3].

This paper, in essence, proposes a CNN architecture for a better understanding of deep learning based face recognition models. For the first architecture, the pertained AlexNet is used as feature extractor and supported by SVM classifier for face recognition problem. For the second architecture, pretrained AlexNet with fine-tuning is used for face recognition problem. In order to reveal the performance of the proposed architectures, those have been compared with a conventional face recognition system using HOG algorithm for feature extraction process and SVM classifier for data training step. Three comprehensive dataset are employed to evaluate those systems in a reliable manner. Overall, section 2 details the corresponding literature of the problem whereas section 3 details the proposed deep learning based architectures for face recognition problem. Section 4, on the other hand illustrates the experimental configuration and results. Finally, the paper is concluded at Section 5.

2. Literature review

Deep learning based architectures have recently dominated the field. This section includes some relevant studies that aims to help reader to follow the state of the art technologies. For instance, in a speech recognition study it is aimed to train large scale neural network-based speech models in large data sets. English Broadcast News has been trained on 400M symbols in this speech recognition task and the test results verify the overall accuracy the system within a small word based error rate

[7]. Another study proposes a natural processing system including, speech tagging, division, entity recognition, etc. with high speed and precision results. The critical issue within this multilayer network architecture that it does need optimized labelled data but employ unlabeled training data [8].

A deep learning based object classification system was presented in a contest that offers a deep convolutional neural network architecture called AlexNet. The systems employed more than 1 million high-resolution images and aimed to classify them into 1000 different categories and better results have been obtained from the previous technology. To reduce overfitting in fully connected layers, the recently developed method of normalization called "Dropout" has been used and proved to be very effective [9]. An outstanding study also employs deep learning approach for scene segmentation and labelling [10]. It mainly performs full scene labeling, also known as scene parsing that comprises labeling the category of the object that each pixel in the image belongs to. Once this is accomplished, every object is identified and labeled successfully. Markov Random Field model was integrated into the Deep Convolution Neural network architecture for the human exposure estimation system in molecular images [11].

Gaining the ability to machines to answer questions automatically is a crucial problem of artificial intelligence community. For this problem, an embedded system using deep learning technology was designed. This system is able to answers questions on a wide range of topics (5,810 question-answer pairs are used for training) from a knowledge base using a small number of engineering features [12].

DeepID network architecture have developed for predicting top-level facial features using a deep convolution neural network and including a formal class of 10,000 classes with the help of these features. These features have been shown to be effective in recognizing new faces that do not appear in face verification and training set. The network is trained to classify all faces in the training set according to their identities [13]. The DeepID network architecture utilizes 4 convolutional layers and a pooling layer, allowing hierarchical extraction of the features. For classification, The SoftMax output layer is involved. The developed system was trained by LFW dataset and it is claimed to have a success rate of 97.45% [13]. Alternatively, a face recognition system was developed based on DeepFace's deep learning technology so as to capture human-level performance in verification applications. The deep network contains more than 120 million parameters from standard convolutional layers. With the developed method, it is claimed to reach 97.35% performance ratio with LFW dataset. [14]. In 2015, a system called FaceNet using deep convolutional

neural network for face recognition and clustering applications was proposed. FaceNet learns a direct mapping from the facial image to the Euclidean space based on the direct facial similarity measure. After the Euclidean space has been obtained, primary tasks namely, face recognition, validation and clustering can be implemented using the techniques within the FaceNet system. The developed system was tested with LFW and Youtube Faces datasets and authors declared to reach success rates of 99.63% and 95.12% correspondingly [15]. One the major challenges for face recognition problem is to exctract effective features to reduce personal changes while increasing interpersonal differences. As well as, complexity and scalability of face recognition problem is also an important challenge, corresponding papers can be seen in [16, 17]. WebFace [18] called CASIA WebFace dataset which contains about 10,000 subjects and 500,000 face images is built by collecting a semi-automatically from internet. 11 layer CNN is used to learn discriminative representation and obtain accuracy on LFW and YTF based on WebFace. This study's aim is to create large scale public database for face recognition problems. VGG-Face [19] is a deep CNN model, consisting of 16 layers, was created with 2.6 million pictures of 2,600 people. [20] propose a novel deep architecture for person re-identification. They introduce two novel layers namely a cross-input neighborhood differences layer, and a subsequent layer. The architecture is conducted on CUHK03 data set and CUHK01 dataset. Their results comparable to the state of the art [20]. DeepID3[21] consist of two deeper neural architectures for face recognition. The networks achive the state of the art performance on LFW 99.53% for face verification accuracy and 96.0% for identification accuracy. Inception, a convolutional neural network is proposed by [22]. The main characteristic of this architecture is the improved utilization of the computing resources inside the network. GoogLeNet is the embodiment of Inception for ILSVRC14 contest and consist of 22-layer network for classification and detection. [22] Deep Pyramid Feature Learning (DPFL) model [23] is presented to extract multi-scale appearance features for person re-identification. Unlike the current methods, the proposed model is able to extract prominent scale-specific features by jointly learning multiple scales of person images by training CNN model. Model conducted on three databases namely Market-1501, CUHK03, and DukeMTMC-reID. Domain Guided algorithm [24] is proposed to improve the generic and robust feature learning procedure for person re-identification. Algorithm provides promising results.

3. METHODOLOGY

In this section conventional face recognition method and deep learning based face recognition architectures are detailed respectively and also, the convolution neural network layers are generally described for a better understanding of architectures. One of the main contribution of this paper to compare success rates of conventional machine learning and deep learning techniques for face recognition problem. Besides, performance of different classifiers, integrated into the deep learning based architecture, are compared using comprehensive face recognition dataset.

3.1 CONVENTIONAL FACE RECOGNITION METHOD

The general face recognition algorithm has the following logic: system consists of two parts, namely the registration and the recognition phases. During the registration process; the system is trained using thousands of face image data and a trained model is created. On the other hand, in recognition process, the facial features extracted from the test image are compared with facial features stored in the database to perform recognition.

Our conventional face recognition model consists of four main parts: pre-processing, face detection, feature extraction and classification. Viola & Jones algorithm [25], HOG [26] and SVM [27] are used for face detection, feature extraction and classification stages, respectively. The proposed system performs detection, alignment and recognition processes successfully. Figure 1 demonstrates the steps of processing the conventional face recognition model. Detailed explanations of the stages and methods used are made in the following subsections.

3.1.1 Pre-Processing and Face Detection Stage

The pre-processing step is important in order to remove the variation effects in the images such as illumination, expression, occlusion, the background of the image and the like. The face recognition performance will improve as the features not related to the facial image to be processed decrease. The face detection module can be considered as the most critical part of the overall framework, and it basically detects and extracts faces from the input images.

The presence of the face and the calculation of the corresponding region have been performed and images are subjected to pre-processing to come from the noise, lighting, pose / turn problems. For detecting phases, the conventional Viola & Jones algorithm is used. It is a commonly used real-time object detection method. The

algorithm has a very high rate of object detection, and instead of scaling the image itself, it scales the properties. It consists of four stages, namely haar features, integral image, adaboost training and cascading classifier. The details can be seen in [25]. The Haar feature classifier consists cascading trained strong classifier, which is based on the adaboost algorithm by Viola and Jones [28].

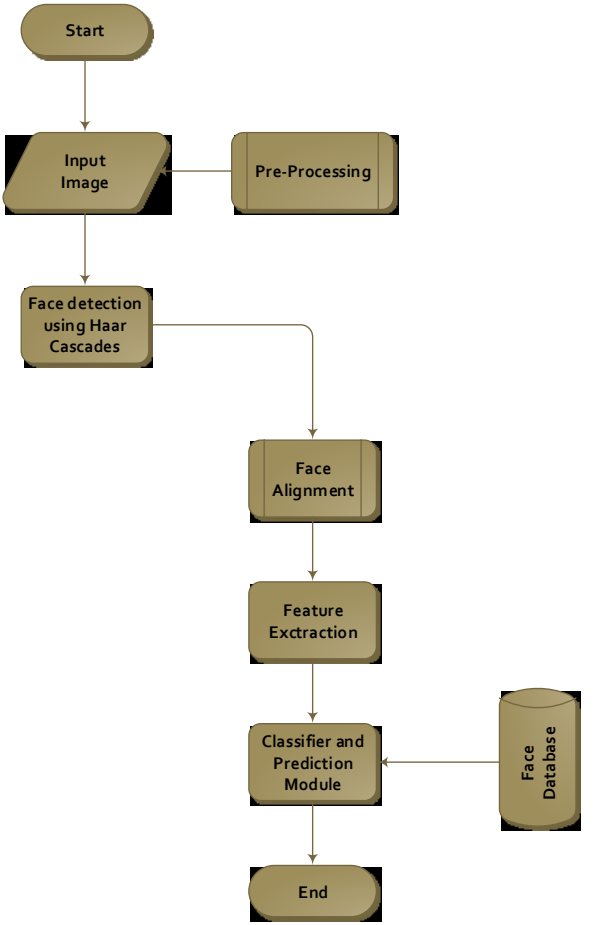


FIGURE 1. Conventional Face Recognition Architecture.

In our method; the original image is taken as input, the image is converted to gray color, and the face is detected using the Viola & Jones algorithm. The detected faces

are cropped and resized to 100 × 100 pixels. The images in the dataset are divided into 60% -40%, 70% -30%, 80% -20%, 90% -10% ratio rate for training and test data, respectively.

3.1.2 Face Alignment and Feature Extraction Stage

Face alignment is an intermediate process required before the face recognition module and is crucial for better recognition results. Extraction and localization of facial landmarks are the critical issues in this process. Facial feature extraction and localization problems have been addressed by algorithms combining shape and texture modelling. Correspondingly, in this study, a shape model-based approach was utilized [29]. The algorithm proposes a shape constraint technique basically employing a multi-stage algorithm to automatically locate facial features. The individual face detectors are combined and applied to the facial images in order to predict facial landmarks.

Feature extraction is the central process for the success of the face recognition algorithm [29]. Histogram of Gradients (HOG) is used in our conventional face recognition method as feature extractor. HOG algorithm which is a powerful descriptor for object recognition and particularly in face recognition was proposed by Dalal et al. [26] for goal of human detection. For feature extraction using HOG method; the images are divided into small bounded regions, called cells, and for each cell a histogram of edge orientations is obtained. The histogram counts are normalized to recover the illumination. The final HOG descriptor is represented by combination of these histograms [30]. For obtaining hog feature, gradient calculation, gradient vote calculation and normalization calculation are combining. The formulas necessary to extract the HOG features of an image are as follows.

$$f_x(x,y) = I(x+1,y) - I(x-1,y) \quad (1)$$

$$f_y(x,y) = I(x,y+1) - I(x,y-1) \quad (2)$$

For each pixel (x,y) horizontal and vertical gradients are calculated by equation (1) and (2).

$$|m| = \sqrt{f_x(x,y)^2 + f_y(x,y)^2} \quad (3)$$

$$\varphi = arctan\frac{f_x(x,y)}{f_y(x,y)} \quad (4)$$

For each pixel, gradient magnitude and gradient direction are calculated by equation (3) and (4). When the HOG algorithm is applied, it is possible to obtain more understandable results by grouping the orientations of the pixels in the generated histogram. This grouping is possible by drawing the angular values in the range 0-360 to a desired range [31]. In our method; 9 groups of 20 degrees were created for directions between 0-180 degrees (20 degrees in each zone). In the last step HOG feature vector is obtained for faces after normalization by equation (5).

$$f = \frac{v_k}{\sqrt{||v_k|| + \varepsilon^2}} \tag{5}$$

where $\boldsymbol{\varepsilon}$ refers constants, v_k is the normalized histogram vector obtained from a block, and f is the HOG feature vector. After these calculations, 4356 dimensional feature vector obtained from 100×100 pixel of size image.

3.1.3 Classification Stage

In our conventional model, multi-class SVM method, one vs all approach is used for classification of faces. Face recognition is k class problem that k is the number of each individual [32]. The input for the SVM is obtained as the result of the feature extraction process by HOG descriptor. Support Vector Machines are popular supervised learning models, developed for the solution of classification and regression analysis problems by Vapnik et al. [27]. The SVMs aim to find the best hyperplane to maximize the distance between support vectors of different classes. SVMs are able to solve multi-class classification problems [33]. There are two approaches used in SVMs to solve multi-class problems. In the proposed method, one vs all approach is employed that a number of SVMs are trained for a number of classes. Class is the number of each individual. Each SVM separates a single class from all remaining classes [34]. For instance, the data from the nth class is trained as a positive example with the n binary classifier, while the remaining (k-1) class is trained as a negative example. During testing, the class label is determined by the binary classifier giving the highest output value [35]. The mathematical equations of model is given below (6-8):

Consider a M class problem and N training samples: $\{x_1,y_1\},\dots,\{x_N,y_N\}$ $x_i \in \mathbb{R}^m$ is an m-dimensional feature vector representing the ith training sample and $y_i \in$

$\{1,2,\ldots,\mathrm{M}\}$ is the class label of x_i. A hyperplane in the feature space can be described by the equation:

$$f_i(x) = w_i^{\mathrm{T}}\Phi(x) + b_i w \tag{6}$$

Where w, is the weight, b is the bias and scalar value. Φ, is the feature vector in the multidimensional extension of the input vector x.

$$L\left(w_i,\xi_j^i\right) = \frac{1}{2}||w_i||^2 + C\sum_{j=1}^{N}\xi_j^i \tag{7}$$

Where ξ is slack variable which is relevant to the soft margin. The tuning parameter, C>0, which is applied to balance the weight of the margin and the training error. L, is the error penalty. Lagrange multipliers is used to solve the optimization problems which transform them to quadratic programming problems.

$$i^* = \underset{i=1,\ldots,M}{argmax} f_i(x) = \underset{i=1,\ldots,M}{argmax}(w_i^T\Phi(x) + b_i) \tag{8}$$

At the classification phase, x is classified for the maximum value i^*provided by f_i.The result of the classifier is the output of this argmax function.

3.2 DEEP LEARNING BASED FACE RECOGNITION MODEL

In this subsection, the CNN layer's working logics are generally described for a better understanding of proposed architectures of our study. Deep learning, in particular CNN's are generally trainable multi-layered architecture designed to learn the unchanging features of the neural network which is inspired by the biological neuron. CNN learns end-to-end training features in a hierarchical manner due to the characteristics of multi-layered architecture [6-10]. The architecture of convolutional neural networks consists of three basic layers such as convolution layer, a pooling layer, and a fully connected layer that follow the input layer.

A. Convolutional Layer

The convolution layer is the basis of the convolution neural network and its primary task is to extract the properties of the input image. In this layer; each filter (kernel or neuron) detects a different feature on the input. A different filter is applied to each

convolution layer and these results are combined for feature extraction. Essentially features from primitive to advance are obtained hierarchically. The output of a filter of the previous layer is the input of the filter in a next layer and contains features of the previous layer. The feature is learned by scanning the filter at a certain size that the network learns its values through filtering.

The original image pixel value and filter values are multiplied together to obtain a single value. This process is repeated for every position in the input volume. Finally, the values obtained are called feature map (activation map). In essence each feature map signifies a certain feature at the output layer. Each filter creates different features of the view. The mathematical description of convolution operation is illustrated in equation 9 as follows [6]:

$$y_j = b_j + \sum_{i}^{n} x_j * (k_{ij}) \tag{9}$$

where, y_j: Input activation map; x_j: Input activation map ; b_j: Bias parameter; k_{ij}:Trainable filter.

Convolutional layers are usually followed by a nonlinear activation function, activation layer. There are many activation functions like sigmoid, tanh, maxout, Relu (Rectified Linear Unit), leakly Relu, elu,etc. Despite the tanh and sigmoid functions are also preferred as activation functions, the ReLU has superiority over them due to its efficient calculations characteristics. This essentially accelerates the training period of the network and allows faster convergence process. An example ReLU function $y = \max(0, x)$ is illustrated in the Figure 2 below. In other words, the activation is simply thresholded at zero when x<0.

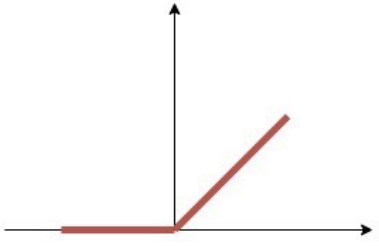


FIGURE 2. The ReLU function.

No parameters are learned in this layer. The aim of this layer is provide nonlinearity to the systems, applying linear operations through convolution layers [36].

B. Pooling Layer

Convolution and ReLU layers are usually followed by the Pooling layer. The main purpose of the pooling layer is to gradually reduce the spatial dimension of the input image which reduces the computational complexity of the model and therefore controls overfitting. A pool operator including max, average, sum is applied to the feature map obtained from the previous layer. The pooling operator returns a value for each filter [37].

C. Fully Connected Layer

The fully connected layer is usually employed at the end of the convolution and pooling layers. This layer resembles the conventional neural networks that each pixel is considered to be a separate neuron and contains as many neurons as the number of expectable classes. Output obtained as a result of convolution, ReLU and pooling layers; contains distributed features of the input image. The purpose of this layer is to use all of these properties to create properties with strong capabilities in the next stage. Accordingly, this layer classifies the input image using top-level properties that come from the previous layers. Besides, this layer allows to learn non-linear combinations of top-level features.

In the following subsections, CNN based face recognition architectures relying on two different classifiers are explained respectively.

3.2.1 CNN based Face Recognition Architecture using SVM

CNN based face recognition architecture “relying on a SVM classifier” has been employed to overcome face recognition problem. CNN, which can determine which parts of a face should be measured, is used in the step of extracting the distinguishing features from the images. For this purpose; a modified version of AlexNet convolutional neural network is used for feature extraction stage and a multi-class SVM is used classification stage. The proposed CNN based Face Recognition Architecture is illustrated in Figure 3.

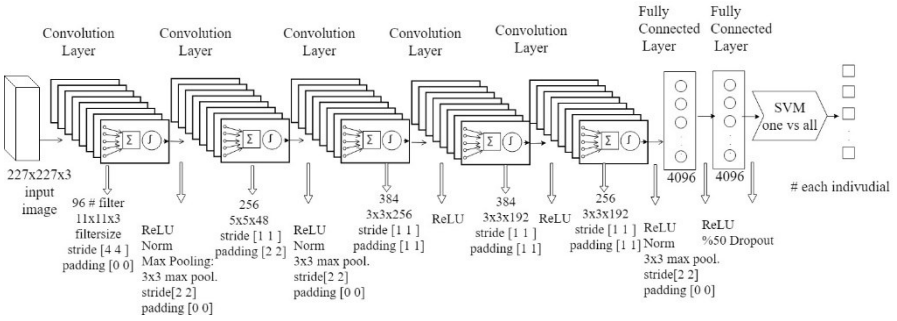


FIGURE 3. The CNN based face recognition architecture using SVM

AlexNet is a deep convolutional neural network was proposed by Krizhevsky et al. [9] to classify the 1.2 million labelled images in the ImageNet into the 1000 different classes. The architecture consists of five convolution layers with weights and decreasing filter size; followed by some of the pooling layers and 3 fully connected layers. One of the main characteristics of the AlexNet is the speed of downsampling of the intermediate representations through following convolutions and max-pooling layers. The final convolutional activation map is formed as a vector and send as an input to following two fully connected layers of 4096 units in size. The image descriptor produced by AlexNet is represented by the output of this layer [38].

The steps of the face recognition application performed by the deep learning method are detailed as follows:

- In the pre-processing stage; there are some modifications. Since the AlexNet network is trained on 227x227 pixel and colour images, all images are resized to 227x227 pixel. The network requires 3 channel input. The dataset with grey colour images (1 channel) are converted to RGB by the other two channels are simply copied and a three channel image is obtained [39].
- In the face detection stage; faces are detected by Viola & Jones algorithm.
- The images in the dataset are divided into 60% -40%, 70% -30%, 80% -20%, 90% -10% ratio for training and test data, respectively, and randomly selected.
- The features of the images in the training set are extracted by CNN, with pre-trained AlexNet architecture, and these properties are used to train and test the multi class SVM classifier as illustrated in Figure 3. The first layers

of the network extract basic image features such as edges, corners etc. These basic features are then processed by deeper layers of the network to produce higher-level image features. This higher-level feature is more suitable for classification. Because they combine all basic features with a richer image presentation. While each layer of a CNN generates a response to an input image only a few layers are suitable for feature extraction.

- For AlexNet architecture the final layer is for the classification problem. AlexNet has been trained to classify a 1000-class problem that is not suitable for the datasets have used in this study. Therefore, the classification layer of this network is not utilized, whereas the features (4096 dimension feature vector) obtained from the second fully connected layer of this network is used to train a multi-class SVM classifier (illustrated in figure3).
- The outputs of last fully connected layer are used as input for training the one vs all design SVM classifier (working logic is detailed in Section 3.1). Test images are classified by using trained SVM. The actual result is compared with the classifier's prediction. The operation is performed correctly if the two results are the same.

3.2.2 CNN based Face Recognition Architecture using Softmax

Training a deep network from scratch which means optimizing millions of parameters to learn weights requires a large amount of data, computational and memory resources for training stage [40,41]. Therefore, fine tuning a network with transfer learning is an alternative method and frequently used in deep learning applications. The second architecture is essentially tuning the pre-trained AlexNet network weights by backpropagation algorithm and the Softmax classifier, which is a probabilistic approach minimizes cross-entropy between the appraised class probabilities and the “true” distribution. The architecture of this model is also illustrated in Figure 4.

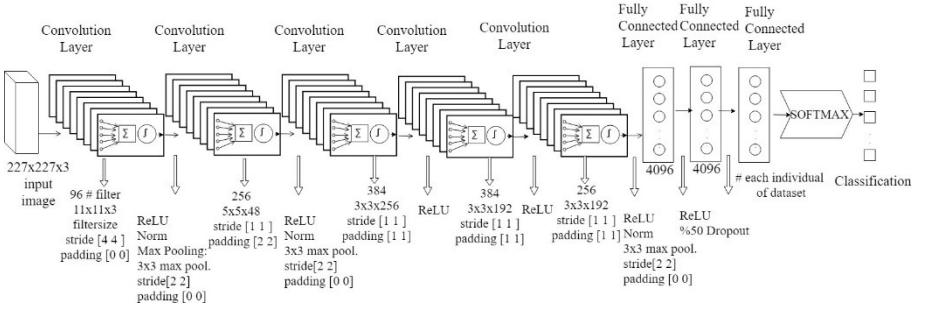


FIGURE 4. Fine tuning pre-trained network architecture

The steps of the face recognition application performed by a pre-trained model are detailed as follows:

- In the pre-processing and face detection steps are the same as the CNN based face recognition architecture. (see section 3.1)
- The images in the dataset are divided into 60% - 40%, 70% - 30%, 80% - 20%, 90% - 10% ratio for training and test data sets respectively, which are randomly selected as well.
- The last three layers of the architecture represent fully connected layers and also a softmax layer is added as the classifier layer, providing a probabilistic approach. And a classification layer of AlexNet are replaced for our task. For this purpose; earlier layers of AlexNet is fixed; last three layers are fine tuned for face recognition task. Previous layers of network is also fixed that those layers contain more general features. However following layers are more powerful in terms of extracting precise to the details of problem. The last fully connected layer's options are specified according to our datasets such as setting same class number of our dataset. Training options are specified as following: epoch:20; validation frequency: 3 iterations; minibatchsize:32; initial learning rate: 1e-4; hardware resource: single GPU; learning rate schedule: constant; learning rate: 0.0001.
- After obtaining features by fine tuning, the softmax classifier is trained using back propagation algorithm with our datasets.

The predicted class is:

$$i = \underset{i}{argmax} a_i \tag{11}$$

- Softmax classifier is the binary Logistic Regression classifier's generalization to multiple classes. It takes a vector which consists real-valued scores in given class and normalizes the values between zero and one that sum to one.
- Softmax classifier is calculated by equation 12.

$$P(c_r|x) = \frac{P(x|c_r)\,P(c_r)}{\sum_{j=1}^{k} P(x|c_j)\,P(c_j)} = \frac{e^{a_r}}{\sum_{j=1}^{k} e^{a_j}} \tag{12}$$

where $a_r = \ln P(x|c_r)\,P(c_r), P(x|c_r)$ is the conditional probability of the sample given class r, and $P(c_r)$ is the class prior probability.

4. Experimental configuration and results

In this section, data sets used in our experiments are introduced and the performances of deep and conventional face recognition methods are evaluated. All experiments are conducted on MATLAB 2018a and a desktop computer with the following specifications: Intel i7 7700K 4.20 Ghz CPU, Nvdia GeForce 1080 GPU, 16 GB RAM.

4.1 DATABASES

For the experimental section, 3 popular and comprehensive datasets, namely, AT&T [42], faces95 and faces96 [43] are employed to train and compare the performance of the conventional and deep learning based methods.

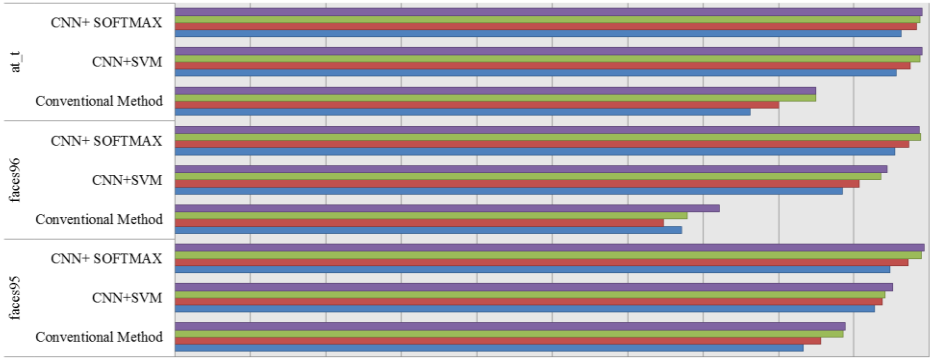
The AT&T face database contains 400 images of 40 different person. The size of each image is 92x112 pixels with 256 grey levels per pixel.

The faces95 dataset contains a total of 1440 colored images of 72 different individuals. The size of each image is 180x200 pixels.

The faces96 dataset contains a total of 3040 colored images of 152 different individuals. The size of each image is 196x196 pixels.

4.2 RESULTS

In this subsection, face recognition performance rate of conventional and deep models are examined and obtained results are illustrated. As it defines the efficiency of the algorithm, evaluating the performance of deep learning methods is as important as the algorithm itself [44].



	faces95			faces96			at_t		
	Conventional Method	CNN+SVM	CNN+ SOFTMAX	Conventional Method	CNN+SVM	CNN+ SOFTMAX	Conventional Method	CNN+SVM	CNN+ SOFTMAX
9/10 train set	88,89%	95,15%	99,31%	72,19%	94,37%	98,68%	85,00%	99,06%	99,06%
8/10 train set	88,54%	94,10%	98,96%	67,88%	93,54%	98,84%	85,00%	98,75%	98,75%
7/10 train set	85,64%	93,75%	97,22%	64,79%	90,73%	97,24%	80,00%	97,50%	98,33%
6/10 train set	83,30%	92,71%	94,79%	67,14%	88,49%	95,45%	76,25%	95,63%	96,25%

FIGURE 3. Conventional & Deep Learning based architectures comparison results.

Figure 5 illustrates the performance comparison of used architectures based on three different datasets with varying training dataset rate. It has been observed that the face recognition accuracy for all data sets increases correctly with the amount of data in the training data set.

Results reveal that Deep learning based architecture outperform the conventional one for all datasets even if the training dataset decreases. For all used datasets, deep learning based architecture achieves more than %90 success rate. In this context, it has been proven that self-learning ability, which distinguishes deep learning from

other methods in the process of feature extraction, is successful. As can be seen from the figure; face recognition performance of fine tuning pre trained AlexNet model is better than CNN+SVM deep model especially for faces95 and faces96 databases.

5. Conclusion

This paper proposes a deep learning based architectures for the face recognition problem. In order to reveal the efficiency and accuracy of the proposed system, it is compared with one of the most reliable conventional architecture. This architecture extracts feature using HOG algorithm whereas the classification is achieved by designing a one vs all multi-class SVM.

The proposed deep learning based architectures, on the other hand, utilizes a popular pre-trained Convolutional neural network architecture, called as AlexNet. This architecture is modified and adapted into the proposed face recognition system that an SVM classifier is integrated into the system for the classification phase.

As a second approach, the same architecture, utilizing fine tuning pre trained AlexNet model, integrated Softmax classifier instead of SVM for face recognition problem. Overall, these three architectures are compared by utilizing three standard datasets, namely, faces95, faces96 and at_t, designed for face recognition problem. Results verify the superiority of the deep learning based architecture over the conventional architecture. Also results indicate that fine-tuned model (CNN + Softmax) performs better than the pre-trained model relying on CNN and SVM classifier. As it is expected, results reveal that the performance of the deep learning architecture increases with respect to the size of the training data.

References

[1] Tolba, A. S., A. H. El-Baz, and A. A. El-Harby. "Face recognition: A literature review." International Journal of Signal Processing 2.2 (2006): 88-103.

[2] Sharif, Muhammad, Sajjad Mohsin, and Muhammad Younas Javed. "A survey: face recognition techniques." Research Journal of Applied Sciences, Engineering and Technology 4.23 (2012): 4979-4990.

[3] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." nature 521.7553 (2015): 436.

[4] Elgallad, Elaraby A., et al. "Human identity recognition using sparse auto encoder for texture information representation in palmprint images based on voting

technique." Computer Science and Information Technology (SCCSIT), 2017 Sudan Conference on. IEEE, 2017.

[5] Anar, Ali Canberk, Erkan Bostanci, and Mehmet Serdar Guzel. "Live Target Detection with Deep Learning Neural Network and Unmanned Aerial Vehicle on Android Mobile Device." arXiv preprint arXiv:1803.07015.2018.

[6] LeCun, Yann, Koray Kavukcuoglu, and Clément Farabet. "Convolutional networks and applications in vision." Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on. IEEE, 2010.

[7] Mikolov, Tomáš, et al. "Strategies for training large scale neural network language models." Automatic Speech Recognition and Understanding (ASRU), 2011 IEEE Workshop on. IEEE, 2011.

[8] Collobert, Ronan, et al. "Natural language processing (almost) from scratch." Journal of Machine Learning Research 12.Aug (2011): 2493-2537.

[9] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

[10] Farabet, Clement, et al. "Learning hierarchical features for scene labeling." IEEE transactions on pattern analysis and machine intelligence 35.8 (2013): 1915-1929.

[11] Tompson, Jonathan J., et al. "Joint training of a convolutional network and a graphical model for human pose estimation." Advances in neural information processing systems. 2014.

[12] Bordes, Antoine, Sumit Chopra, and Jason Weston. "Question answering with subgraph embeddings." arXiv preprint arXiv:1406.3676 (2014).

[13] Sun, Yi, Xiaogang Wang, and Xiaoou Tang. "Deep learning face representation from predicting 10,000 classes." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014.

[14] Taigman, Yaniv, et al. "Deepface: Closing the gap to human-level performance in face verification." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.

[15] Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[16] Setiowati, Sulis, Eka Legya Franita, and Igi Ardiyanto. "A review of optimization method in face recognition: Comparison deep learning and non-deep learning methods." Information Technology and Electrical Engineering (ICITEE), 2017 9th International Conference on. IEEE, 2017.

[17] Sun, Yi, et al. "Deep learning face representation by joint identification-verification." Advances in neural information processing systems. 2014.

[18] Yi, Dong, et al. "Learning face representation from scratch." arXiv preprint arXiv:1411.7923 (2014).
[19] Parkhi, Omkar M., Andrea Vedaldi, and Andrew Zisserman. "Deep Face Recognition." BMVC. Vol. 1. No. 3. 2015.
[20] Ahmed, Ejaz, Michael Jones, and Tim K. Marks. "An improved deep learning architecture for person re-identification." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.
[21] Sun, Yi, et al. "Deepid3: Face recognition with very deep neural networks." arXiv preprint arXiv:1502.00873 (2015).
[22] Szegedy, Christian, et al. "Going deeper with convolutions." Cvpr, 2015.
[23] Chen, Yanbei, Xiatian Zhu, and Shaogang Gong. "Person re-identification by deep learning multi-scale representations." (2018).
[24] Xiao, Tong, et al. "Learning deep feature representations with domain guided dropout for person re-identification." Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on. IEEE, 2016.
[25] Viola, Paul, and Michael J. Jones. "Robust real-time face detection." International journal of computer vision 57.2 (2004): 137-154.
[26] Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 1. IEEE, 2005.
[27] Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." Machine learning 20.3 (1995): 273-297.
[28] Li, Xiang-Yu, and Zhen-Xian Lin. "Face Recognition Based on HOG and Fast PCA Algorithm." The Euro-China Conference on Intelligent Data Analysis and Applications. Springer, Cham, 2017.
[29] Albiol, Alberto, et al. "Face recognition using HOG–EBGM." Pattern Recognition Letters 29.10 (2008): 1537-1543.
[30] Déniz, Oscar, et al. "Face recognition using histograms of oriented gradients." Pattern Recognition Letters 32.12 (2011): 1598-1603.
[31] Peker, Murat, Halis Altun, and Fuat Karakaya. "HOG Temelli Bir Yöntem ile Ölçek ve Yönden Bağımsız Gerçek Zamanlı Nesne Tanıma."
[32] Phillips, P. Jonathon. "Support vector machines applied to face recognition." Advances in Neural Information Processing Systems. 1999.
[33] Ayhan, Sevgi, and Şenol Erdoğmuş. "Destek vektör makineleriyle sınıflandırma problemlerinin çözümü için çekirdek fonksiyonu seçimi." Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi 9.1 (2014).
[34] Heisele, Bernd, Purdy Ho, and Tomaso Poggio. "Face recognition with support vector machines: Global versus component-based approach." Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on. Vol. 2. IEEE, 2001.

[35] Wang, Zhe, and Xiangyang Xue. "Multi-class support vector machine." Support Vector Machines Applications. Springer, Cham, 2014. 23-48.
[36] Wu, Jianxin. "Introduction to convolutional neural networks." National Key Lab for Novel Software Technology. Nanjing University. China (2017).
[37] O'Shea, Keiron, and Ryan Nash. "An introduction to convolutional neural networks." arXiv preprint arXiv:1511.08458 (2015).
[38] Grm, Klemen, et al. "Strengths and weaknesses of deep learning models for face recognition against image degradations." IET Biometrics 7.1 (2017): 81-89.
[39] Pilla Jr, Valfredo, et al. "Facial Expression Classification Using Convolutional Neural Network and Support Vector Machine."
[40] Ghazi, Mostafa Mehdipour, and Hazim Kemal Ekenel. "A comprehensive analysis of deep learning based representation for face recognition." arXiv preprint arXiv:1606.02894 (2016).
[41] Lu, Ze, Xudong Jiang, and Alex Kot. "Enhance deep learning performance in face recognition." Image, Vision and Computing (ICIVC), 2017 2nd International Conference on. IEEE, 2017.
[42] Samaria, Ferdinando S., and Andy C. Harter. "Parameterisation of a stochastic model for human face identification." Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on. IEEE, 1994.
[43] Hond, Darryl, and Libor Spacek. "Distinctive Descriptions for Face Processing." BMVC. No. 0.2. 1997.
[44] Bostanci, Betul, and Erkan Bostanci. "An evaluation of classification algorithms using Mc Nemar's test." Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012). Springer, India, 2013.

Current Address: FATIMA ZEHRA UNAL: Ankara University, Department of Computer Engineering, Ankara TURKEY
E-mail: fzkilic@ankara.edu.tr
ORCID: https://orcid.org/0000-0002-1789-0893
Current Address: MEHMET SERDAR GUZEL: Ankara University, Department of Computer Engineering, Ankara TURKEY
E-mail: mguzel@ankara.edu.tr
ORCID: http://orcid.org/0000-0002-3408-0083

Commun.Fac.Sci.Univ.Ank.Series A2-A3
Volume 61, Number 2, Pages 150-160 (2019)
DOI: 10.33769/aupse.532422
ISSN 1303-6009 E-ISSN 2618-6462
http://communications.science.ankara.edu.tr/index.php?series=A2A3

MAGNETOCALORIC EFFECT AROUND CURIE TEMPERATURE IN Ni_{50-x} $Cu_xMn_{38}Sn_{12}B_3$ SHAPE MEMORY RIBBONS

Olcay KIZILASLAN

ABSTRACT. The magnetocaloric effect in $Ni_{50-x}Cu_xMn_{38}Sn_{12}B_3$ ribbons depending on the Cu substitution (x= 0, 1, 3) was investigated around the Curie temperature. The purpose of the present study was to analyze the magnetocaloric effect around a second order phase transition (around the Curie temperature) which has a smaller thermal hysteresis compared to a first order phase transition (Martensitic transition). The Curie temperature of the ribbons shifted to higher temperatures with increasing Cu content. A conventional magnetocaloric effect (MCE) was observed around the Curie temperature when the ribbons are subjected to a magnetic field change of 5 T. The magnetic entropy changes were calculated based on the isothermal magnetization M(H) data using thermodynamic Maxwell equation. The highest magnetic entropy change and the refrigerant capacity was obtained for the x=1 ribbon.

1. INTRODUCTION

Ni-Mn-X (X = Sn, In and Sb) metamagnetic shape memory alloys (MSMAs) are of great interests due to their potential applications such as magnetic refrigeration materials, magnetic shape memory effect, magneto-resistance, magneto-thermal conductivity, elasta-caloric effect [1-11]. These alloys are one step ahead in practical applications when considering the cost-performance relationship [12]. They have also a large refrigeration capacity (RC) around martensitic-austenite phase transition [13] which is comparable to the compounds containing rare-earth element [14,15]. Strong conventional magnetocaloric effect (MCE) is caused by a magnetic transition since the magnetization strongly varies in a very narrow temperature range around transition temperature. The MCE can be tuned [16,17] by substituting or doping ferromagnetic [18-20] or non-ferromagnetic elements [21-23].

Received by the editors: February 26, 2019; Accepted: June 25, 2019.
Key word and phrases: Ni-Mn-Sn Heusler alloys, magnetocaloric effect, magnetic properties, substitution.

The total adiabatic entropy change of a magnetic material is the sum of magnetic ΔS_M, lattice ΔS_L, and electronic ΔS_E entropy given by $\Delta S_T = \Delta S_M + \Delta S_L + \Delta S_E$. The lattice and electronic entropy are independent of external magnetic field. The total entropy is constant if the process is adiabatic and reversible. When a magnetic field is applied on the material, the magnetic moments of the material are aligned along the magnetic field direction which implies a decrease in magnetic entropy. For the conservation of total entropy change, the decrease in the magnetic entropy is compensated by an increase in the lattice entropy. The increase in lattice entropy causes an increase of the material temperature.

This study investigates the effect of Cu substitution on magnetocaloric effect in the vicinity of Curie temperature (T_c) in $Ni_{50-x}Cu_xMn_{38}Sn_{12}B_3$ (x = 0, 1, 3) shape memory ribbons. The Curie temperatures for x = 0, 1, 3 were found to be 315, 321 and 319 K, respectively, which are very close to room temperature. This makes the investigation of the MCE around the Curie temperature important. On the other hand, the transition around the curie temperature is a second order phase transition and this will be discussed below. Such a transition provides a large usable temperature range (compared to a first order transition [24] observed in this compound at low temperature) as the transition temperature of the ribbon has to span the entire working region of the cooling device. The MCE and the effective refrigerant capacity RC_{eff} were calculated depending on the Cu substitution. The magnetic entropy change ΔS_M was determined on the basis of magnetization data $M(H)$. The highest RC value (96.44 J/kg) was found for the x=3.

This study investigates the effect of Cu substitution on magnetocaloric effect in the vicinity of Curie temperature (T_c) in $Ni_{50-x}Cu_xMn_{38}Sn_{12}B_3$ (x = 0, 1, 3) shape memory ribbons. The Curie temperatures for x = 0, 1, 3 were found to be 315, 321 and 319 K, respectively, which are very close to room temperature. This makes the investigation of the MCE around the Curie temperature important. On the other hand, the transition around the curie temperature is a second order phase transition and this will be discussed below. Such a transition provides a large usable temperature range (compared to a first order transition [24] observed in this compound at low temperature) as the transition temperature of the ribbon has to span the entire working region of the cooling device. The MCE and the effective refrigerant capacity RC_{eff} were calculated depending on the Cu substitution. The magnetic

entropy change ΔS_M was determined on the basis of magnetization data $M(H)$. The highest RC value (96.44 J/kg) was found for the x=3.

2. Experimental

The appropriate amount of Ni, Mn, Sn, B and Cu powders were mixed to fabricate $Ni_{50-x}Cu_xMn_{38}Sn_{12}B_3$ (x = 0, 1, 3) shape memory ribbons and the mixture was melted in an arc-melter in an argon atmosphere. Thus, the ignots of the polycrystalline $Ni_{50-x}Cu_xMn_{38}Sn_{12}B_3$ were produced. In order to make the ingots more homogeneous, the melting process was repeated several times. The obtained ingots were used to produce the ribbons by melt-spinning technique. The dimensions of the produced ribbons were about 5-6 mm in width, 20-25 mm in length and 15-20 μm in thickness, respectively. The ribbons were also heat treated in vacuumed quartz ampoules at 1173 K for 2 h and then quenched in ice-water.

Magnetic measurements were performed with Quantum Design Physical Property Measurement System (PPMS) - 9T. The system is able to resolve the magnetization changes of less than 10^{-6} emu at a data rate of 1 Hz. The sweep rate, which means how fast the magnetic field change between the field set points, was set to 100 Oe/sec for all the measurement.

3. Results and discussion

Figure 1 shows the magnetization-temperature (M-T) curves of the ribbons under a magnetic field of 1T. A magnetic transition (MT) from a ferromagnetic state to a paramagnetic state was observed. Curie temperatures T_c were obtained to be 315, 320 and 317 K for x= 0, 1 and 3, respectively. T_c values were determined by taking the first derivative of M(T) curves. The increase of the Curie temperature is attributed to the enhancement of ferromagnetic coupling.

Figure 2a shows the isothermal M-H curves of the x= 0 parent ribbon at the temperatures from 300 to 360 K in an interval of 5 K. The measurements were performed at the temperatures in the vicinity of the Curie temperature. A similar $M(H)$ characteristic was also observed for the x=1 and 3 ribbons.

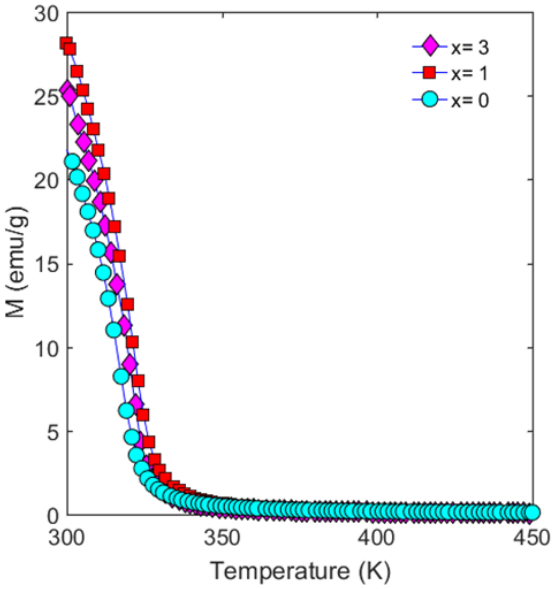


FIGURE 1. The magnetization curve of the samples as a function of temperature at the magnetic field of 1 T.

The Arrott plot technique was used to determine the nature of magnetic phase transition. Figure 2b shows M^2 versus H/M curves obtained from $M(H)$ data (figure 2a) at the temperatures around the Curie temperature. According to Banerjee's criterion [25], if the slope is positive, the material undergoes a second order phase transition (SOPT). Figure 2b clearly indicates that the ribbons used in this study undergo a second order phase transition around the Curie temperature. For the x=1 and 3 Cu substituted ribbons, a similar Arrott plot characteristic was observed. As the transition around the Curie temperature is a second order phase transition, there exists a small thermal hysteresis. A large thermal hysteresis generally accompanies to the first order magnetic phase transition and this strongly influences the refrigerant efficiency of the MCE. It should be emphasized that the thermal hysteresis increases the temperature range of refrigeration cycle, causing a reduce in the refrigerant efficiency [26]. On this purpose, in this study the MCE properties of Ni_{50-x} $Cu_xMn_{38}Sn_{12}B_3$ ribbons were investigated around the Curie temperature. The magnetocaloric efficiency around a transition with small thermal hysteresis is much higher and the corresponding MCE can be more effectively used in the magnetic refrigeration technology. It is worth to note that the MCE was generally investigated

in the literature around the martensitic transition which has very large thermal hysteresis. For the $Ni_{50-x}Cu_xMn_{38}Sn_{12}B_3$ ribbons there is no study reported in the literature on the investigation of the MCE around the Curie temperature.

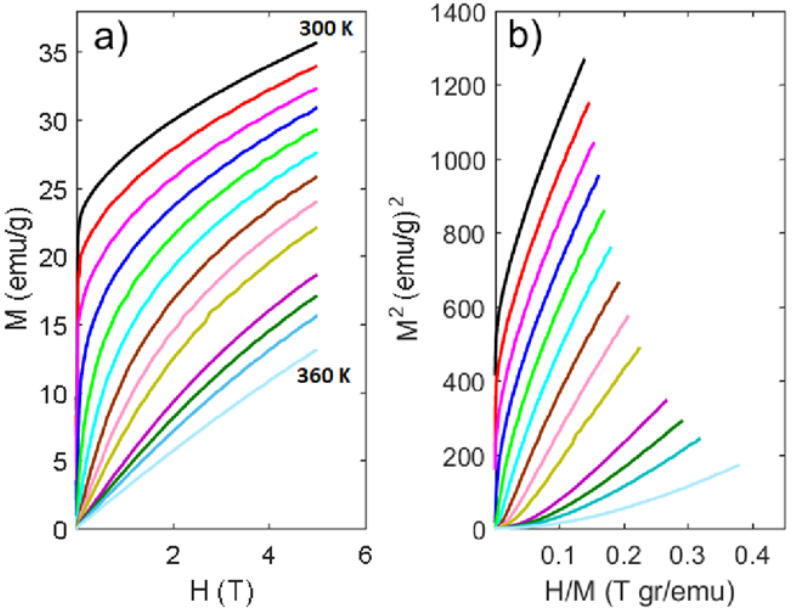


FIGURE 2. a) the M-H curves of the ribbon (x=0) up to 5T, at different temperatures between 300 K and 360K in an interval of 5K, b) Arrott (M^2 vs H/M) plots.

Magnetic entropy change can be calculated by using the following thermodynamic Maxwell equation;

$$\Delta S_M = \int_{H_O}^{H_F} \left(\frac{\partial M}{\partial T}\right)_H dH \qquad \text{Eq. 1}$$

where $H_0 = 0$, if the field is changed from 0 to H. $\frac{\partial M}{\partial T}$ can be calculated numerically with a simple formula given below, Eq. 2. The integral was numerically solved using the trapezoidal integration.

$$\frac{\partial M}{\partial T} = \sum_{i=1}^{i_{max}} \frac{M(T_{i+1}) - M(T_i)}{T_{i+1} - T_i} \qquad \text{Eq. 2}$$

If the number of experimental $M(H)$ curves is N (an integer number), the resulting $\Delta S_M - T$ data will have N-1 points. The reason for this loss is the numerical method used for the calculation of the derivative, see eq.2.

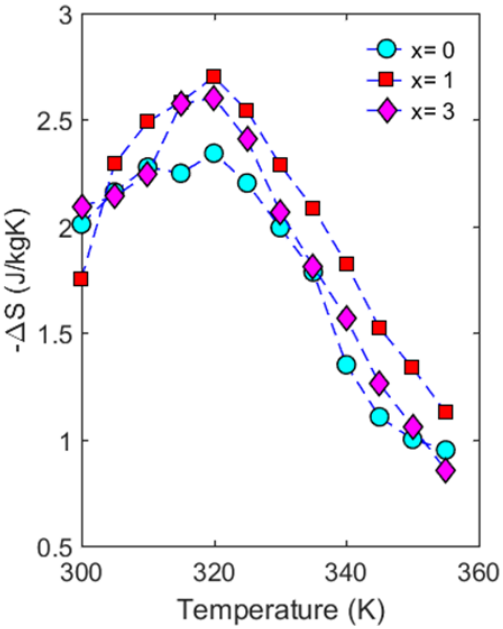


FIGURE 3. Temperature dependence of the isothermal magnetic entropy change change ΔS_M at different at different Cu substitution levels under a magnetic field change of 5 *T*.

Figure 3 shows ΔS_M as a function of temperature calculated by using Eq. 1 at different substitution levels (x= 0, 1 and 3). The measurements were performed at the temperatures below and above Curie temperature (T_c) with the temperature steps of ΔT=5 K. The ΔS_M^{max} values were calculated to be 2.34, 2.71 and 2.60 J/kgK for

the x=0, 1 and 3 ribbons, respectively, indicating an increase in ΔS_M^{max} of 15.8 % for the x= 1 ribbon compared to the x=0 ribbon.

The magnetization difference, ΔM, between two phases is responsible for magnetic entropy change [27]. If one wants to improve ΔS_M, the magnetization difference ΔM should be tuned. In the present study, the magnetization was increased by substituting Cu for Ni, see figure 1. Magnetic properties of Heusler alloys are very sensitive to the structural disorder and strongly depends on the exchange interaction between Mn-Mn atoms [28] because the contribution of the magnetic moments of Ni atoms to the total magnetic moment in Ni-Mn-X (X=Sn, Sb, In) alloys is quite low [29,30]. The exchange interaction determines whether the magnetic order is ferromagnetic or antiferromagnetic. The substitution of Cu for Ni causes a change of the distances between Mn-Mn atoms. These new interatomic distances are more favorable for the ferromagnetic order. Therefore, the Cu substitution leads to a strong ferromagnetic coupling. Figure 1 supports this idea and an increase in the magnetization difference ΔM was observed in the x=1 and 3 ribbons. The ΔM value is decreased in the x=3 ribbon compared to the x=1 ribbon but it was still bigger than the ΔM value of the x=0 parent ribbon.

A large refrigerant capacity value RC as well as large entropy change is also crucial parameter for the magnetic refrigeration applications. The area under $\Delta S_M - T$ curve in Figure 3 gives the refrigerant capacity $\left(RC = \int_{T_1}^{T_2} \Delta S_M \, dT\right)$, which is a measurement of heat transport between hot and cold reservoirs in an ideal refrigerator. T_1 and T_2 are the temperatures which correspond to half maximum value (ΔT_{FWHM}) in both side of ΔS_M peak. The area was calculated by trapezoidal integration method and the corresponding RC versus x is given in Figure 4. The highest RC value was found to be 88.89 J/kg for the x=1 which indicates an increase of RC by 13.7 %. The maximum RC obtained in this study is comparable with the other Heusler alloys [26,31-34].

Hysteresis loss (HL) must be taken into account for evaluating effective refrigerant capacity RC_{eff} during a thermodynamic cycle. The area between magnetization and demagnetization curves gives HL. The RC_{eff} can be calculated by subtracting hysteresis loss from RC, $RC_{eff} = RC - HL$ [27]. The calculated hysteresis areas at 335 K for the x=0, 1 and 3 were very small and found to be 0.59, 0.66 and 1.05 J/kg for the x=0 ,1 and 3, respectively. Then, the calculated RC_{eff} values are 77.58, 88.24 and 82.2 J/kg for the x= 0, 1 and 3, respectively. The obtained RC_{eff} values are comparable with the other Ni-Mn-Sn systems, see figure 8 in ref [27] and a promising value for the magnetocaloric applications in the future.

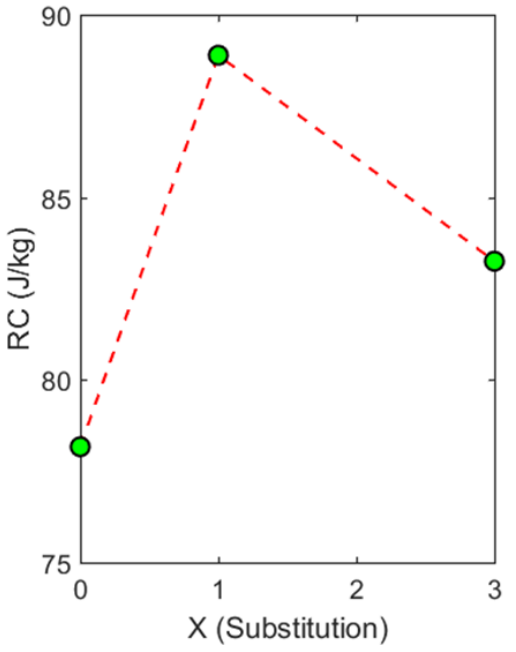


FIGURE 4. The calculated refrigerant capacity, RC, for different substitution levels.

4. CONCLUSION

In this study, the effect of Cu substitution on the magnetocaloric effect and Curie temperature was investigated. The Cu substitution helped to tune not only the magnetocaloric effect also the Curie temperature. An increase of 6 K was observed in the Curie temperature of the x=1 ribbon. However, a decrease in the Curie temperature of the x= 3 ribbon was observed, but it was still above the Curie temperature of the x= 0 parent ribbon. For the x=1 ribbon, a significant magnetic entropy change (ΔS_M^{max}= 2.71 J/kgK) was obtained under a magnetic field change of 5 T. The highest RC_{eff} value, which is a better criterion to evaluate the cooling efficiency, was obtained to be 88.89 J/kg for the x=1.

References

[1] Aydogdu, Y. , Turabi, A.S., Aydogdu, A., Kok, M., Yakinci, Z. D., Karaca, H. E.: The effects of boron addition on the magnetic and mechanical properties of NiMnSn shape memory alloys, J. Therm. Anal. Calorim. 126, (2016), 399–406.

[2] Zhang, B., Zhang, X., Yu, S. , Chen, J., Cao, Z. , Wu, G.: Giant magneto thermal conductivity in the Ni–Mn–In ferromagnetic shape memory alloys, Appl. Phys. Lett. 91, (2007), 012510.

[3] Castillo-Villa, P.O., Mañosa, L., Planes, A., Soto-Parra, D.E., Sánchez-Llamazares, J.L., Flores-Zúñiga, H., and Frontera, C.: Elastocaloric and magnetocaloric effects in Ni-Mn-Sn (Cu) shape-memory alloy, J. Appl. Phys. 113, (2013), 053506.

[4] Pramanick, S., Chatterjee, S., Giri, S., Majumdar, S., Koledov, V. V., Mashirov, A., Aliev, A. M., Batdalov, A. B., Hernando, B., Rosa, W. O., and Gonzlez-Legarreta, L.: Multiple magneto-functional properties of $Ni_{46}Mn_{41}In_{13}$ shape memory Alloy, J. Alloys Compd. 578, (2013), 157161.

[5] Samanta, T., Saleheen, A. U., Lepkowski, D. L., Shankar, A., Dubenko, I., Quetz, A., Khan, M., Ali, N. and Stadler, S.: Asymmetric switchinglike behavior in the magnetoresistance at low fields in bulk metamagnetic Heusler alloys, Phys. Rev. B. 90, (2014), 064412.

[6] Yu, S. Y., Liu, Z. H., Liu, G. D., Chen, J. L., Cao, Z. X., Wu, G. H., Zhang, B., and Zhang, X.: Large magnetoresistance in single-crystalline $Ni_{50}Mn_{50-x}In_x$ alloys, x= (14–16) upon martensitic transformation, Appl. Phys. Lett. 89, (2006), 162503.

[7] Sutou, Y., Imano, Y., Koeda, N., Omori, T., Kainuma, R., Ishida, K. and Oikawa, K.: Magnetic and martensitic transformations of NiMnX(X=In,Sn,Sb) ferromagnetic shape memory alloys, Appl. Phys. Lett. 85, (2004), 4358.

[8] Krenke, T., Acet, M., Wassermann, E. F., Moya, X., Manosa, L., and Planes, A.: Ferromagnetism in the austenitic and martensitic states of Ni-Mn-In Alloys, Phys. Rev. B. 73, (2006), 174413.

[9] Manosa, L., Alonso, D. G., Planes, A., Bonnot, E., Barrio, M., Tamarit, J. L., Aksoy, S. and Acet, M.: Giant solid-state barocaloric effect in the Ni-Mn-In magnetic shape-memory alloy, Nat. Mat. 9, (2010), 478–481.

[10] Kirat G, Kizilaslan O, Aksan M. A.: Magnetoresistance properties of magnetic Ni-Mn-Sn-B shape memory ribbons and magnetic field sensor aspects operating at room temperature. J. Magn. Magn. Mater. 477, (2019), 366-371.

[11] Chattopadhyay, M. K., Manekar, M. A., Sharma, V. K., Arora, P., Tiwari, P., Tiwari, M. K. and Roy, S. B.: Contrasting magnetic behavior of $Ni_{50}Mn_{35}In_{15}$ and $Ni_{50}Mn_{34.5}In_{15.5}$ Alloys, J. Appl. Phys. 108, (2010), 073909.

[12] Chen, F., Liu, W.L., Shi, Y.G., Müllner, P.: Influence of annealing on martensitic transformation and magnetic entropy change in $Ni_{37.7}Co_{12.7}Mn_{40.8}Sn_{8.8}$ magnetic shape memory alloy ribbon, J. Magn. Magn. Mater. 377, (2015), 137–44.

[13] Huang, L., Cong, D. Y., Suo, H. L. and Wang, Y. D.: Giant magnetic refrigeration capacity near room temperature in $Ni_{40}Co_{10}Mn_{40}Sn_{10}$ multifunctional alloy. Appl. Phys. Lett. 104, (2014),132407.

[14] Gschneidner K. A., and Pecharsky, V. K.: Magnetocaloric Materials, Annual Rev. Mater. Sci. 30, (2000), 387–429.
[15] Khattak, K. S., Aslani, A., Nwokoye, C. A., Siddique, A., Bennett, L. H., and Torre, E. D.: Magnetocaloric properties of metallic nanostructures, Cogent Engineering 2, (2015), 1050324.
[16] Liu, J., Scheerbaum, N., Lyubina, J. and Gutfleisch, O.: Reversibility of magnetostructural transition and associated magnetocaloric effect in Ni-Mn-In-Co., Appl. Phys. Lett. 93, (2008), 102512.
[17] Zhang, Y. et al.: Large magnetic entropy change and enhanced mechanical properties of Ni-Mn-Sn-C alloys., Scripta Mater. 75, (2014), 26–29.
[18] Tan, C., Tai, Z., Zhang, K., Tian, X. and Cai, W.: Simultaneous enhancement of magnetic and mechanical properties in Ni-Mn-Sn alloy by Fe doping., Sci. Rep. 7, (2017), 43387.
[19] Tan, C. L., Feng Z. C., Zhang, K., Wu, M. Y., Tian, Guo E. J.: Microstructure, martensitic transformation and mechanical properties of Ni-Mn-Sn alloys by substituting Fe for Ni, Trans. Nonferrous Met. Soc. China. 27, (2017), 2234–2238.
[20] Zhang, H. H., Zhang, X., Qian, M., Wei, L., Xing, D., Sun, J., Geng, L.: Enhanced magnetocaloric effects of Ni-Fe-Mn-Sn alloys involving strong metamagnetic behavior. J. Alloy. Compd. 715, (2017), 206–213.
[21] Qu, Y.H., Cong, D.Y., Sun X.M., Nie Z.H., Gui, W.Y., Li R.G., Ren Y., Wang, Y.D..: Giant and reversible room-temperature magnetocaloric effect in Ti-doped Ni-Co-Mn-Sn magnetic shape memory alloys. Acta Mater. 134, (2017), 236–248.
[22] Cong, D. Y., Huang, L., Hardy, V., Bourgault, D., Sun, X. M., Nie, Z.H., Wang, M.G., Ren, Y., Entel, P., Wang, Y. D.: Low-field-actuated giant magnetocaloric effect and excellent mechanical properties in a NiMn-based multiferroic alloy. Acta Mater. 146, (2018), 142–151.
[23] Liu, C., Li, Z., Zhang, Y., Liu, Y., Sun., J., Huang., Y., Kang, B., Xu, K., Deng, D., Jing, D.: Martensitic transition, inverse magnetocaloric effect and shape memory characteristics in $Mn_{48-x}Cu_xNi_{42}Sn_{10}$ Heusler alloys. Phys. Rev. B Condens. Matter. 508, (2017), 118–123.
[24] Kizilaslan, O., Thermal hysteresis dependent magnetocaloric effect properties of $Ni_{50-x}Cu_xMn_{38}Sn_{12}B_3$ shape memory ribbons, Intermetallics 109, (2019), 135–138.
[25] B.K.Banerjee, On a generalised approach to first and second order magnetic transitions, Phys. Lett. 12, (1964),16.
[26] Tian, F. Zeng, Y, Xu, M., Yang, S., Lu, T, Wang, J., Chang, T, Adil, M., Zhang, Y, Zhou, C. and Song, X.: A magnetocaloric effect arising from a ferromagnetic transition in the martensitic state in Heusler alloy of $Ni_{50}Mn_{36}Sb_8Ga_6$, Appl. Phys. Lett. 107, 012406 (2015).
[27] Zhang, X., Zhang, H., Qian, M. and Geng, L.: Enhanced magnetocaloric effect in Ni-Mn-Sn-Co alloys with two successive magnetostructural transformations Sci. Rep. 8, (2018), 8235.

[28] Hernando, B., Sanchez Llamazares, J.L., Santos, J.D., Sanchez, M.L., Escoda, Ll., Sunol, J.J., Varga, R., Garcia, C., Gonzalez, J.: Grain oriented NiMnSn and NiMnIn Heusler alloys ribbons produced by melt spinning: Martensitic transformation and magnetic properties, J. Magn. Magn. Mater. 321, (2009), 763–768.
[29] Luo, H., Meng, F., Jiang, Q., Liu, H., Liu, E., Wu G. and Wang, Y.: Effect of boron on the martensitic transformation and magnetic properties of $Ni_{50}Mn_{36.5}Sb_{13.5-x}B_x$ alloys, Scripta Mater. 63 (2010), 569–572.
[30] Kübler, J., William, A. R., and Sommers, C. B.: Formation and coupling of magnetic moments in Heusler alloys, Phys. Rev. B. 28, (1983), 1745.
[31] Pecharsky, V. K. Gschncidner K. A. Jr.: Giant Magnetocaloric Effect in $Gd_5(Si_2Ge_2)$, Phys. Rev. Lett. 78, (1997) 4494.
[32] Khan, M., Ali, N., Stadler, S.: Inverse magnetocaloric effect in ferromagnetic $Ni_{50}Mn_{37+x}Sb_{13-x}$ Heusler alloys J. Appl. Phys. 101 (2007), 053919.
[33] Varzaneh, A. G., Kameli, P., Amiri, T., Ramachandran, K.K., Mar, A., Sarsari, I. A., Luo, J. L., Etsell, T. H., Salamati, H.: Effect of Cu substitution on magnetocaloric and critical behavior in $Ni_{47}Mn_{40}Sn_{13-x}Cu_x$ alloys, J. Alloy. Compd., 708, (2017) 34-42.
[34] Zhang, Y., Zheng, Q., Xia, W., Zhang, J., Dua, J., and Yana, A.: Enhanced large magnetic entropy change and adiabatic temperature change of $Ni_{43}Mn_{46}Sn_{11}$ alloys by a rapid solidification method, Scripta Mater. 104, (2015) 41–44.

Current Address: Olcay KIZILASLAN: Inonu University, Department of Biomedical Engineering, Faculty of Engineering 44280, Malatya, Turkey
E-mail: olcay.kizilaslan@inonu.edu.tr, olcaykizilaslan@gmail.com
Orcid ID: https://orcid.org/0000-0003-2528-433X

Commun.Fac.Sci.Univ.Ank.Series A2-A3
Volume 61, Number 2, Pages 161-171 (2019)
DOI: 10.33769/aupse.633838
ISSN 1303-6009 E-ISSN 2618-6462
http://communications.science.ankara.edu.tr/index.php?series=A2A3

PREPROCESSING STEPS IN fMRI: SMOOTHING

Hacer DAŞGIN, Ali YAMAN and Yılmaz AKDİ

ABSTRACT. Functional magnetic resonance imaging is a technique with a primary and dominant effect in the investigation of the cognitive functions of the brain since it has a complex structure. In this study, data obtained from single subject was examined. First statistical parametric mapping results were obtained after applying the standard preprocessing steps with including smoothness. Spatial smoothing was performed using a 3 mm Gaussian kernel which is twice of the voxel size. Second, statistical parametric mapping results were obtained with applying standard preprocessing steps without smoothing. The effects of these two applications on the mapping results were compared for selected slices and locations in terms of statistical and pattern.

1. INTRODUCTION

During neuronal activity there is an increase in blood flow to the brain, resulting in increased oxygen consumption and an expansion of blood vessels [1],[2]. The fMRI method developed to measure this change [3] is used as an important technique to understand brain functions and connectivity in the brain. Prior to performing statistical analysis and statistical parametric mapping in functional magnetic resonance imaging (fMRI) some preprocessing steps applied to the data. The main purpose of this process; remove unrelated variable from the data and prepare the data for statistical analysis. In this study, the effects of smoothing on mapping results was investigated.

Statistical parametric mapping results were obtained by applying the standard preprocessing steps with including smoothness. Spatial smoothing was performed using a 3 mm gaussian kernel. Secondly, statistical parametric mapping results were obtained with applying standard preprocessing steps without smoothing to investigate the effects of smoothing process.

Received by the editors: June 16, 2019; Accepted: August 15, 2019.
Key word and phrases: Statistical parametric mapping (SPM), functional magnetic resonance imaging (fMRI), Gaussian kernel, smoothing.

The change in signal obtained in fMRI is directly dependent on the blood oxygenation. Because the level of blood oxygenation changes rapidly following the activity of neurons in the brain region. Thus, the positions of the active brain regions are determined by means of the fMRI [4]. What changes in the brain of a neural stimulus has recently been followed predominantly by fMRI techniques. Processing and mapping of the signal obtained after stimulation is a necessity that requires statistical analysis. In the fMRI, activation is determined in the relevant region of the brain using various statistical test methods for signal exchange at any firing neurons in response to a stimulus. Since this change is very small, statistical methods are used to determine the activation. Before performing statistical mapping, some preprocessing steps should be applied to remove unrelated variable from the data. These steps include; realignment (eliminates effects of motion), coregistration (to show activation results on high-resolution structural images), segmentation, normalization (that spatial transformations into standard anatomical space) and smoothing.

2. Imaging Method and Analysis

Data were collected using a Siemens Magnetom 7 T (CMRR, USA) system. The block design consists of 10 data sets with a 12 seconds rest as 12 seconds task (figure 1) and the slice thickness is 1.5 mm. The motor task in the experiment consisted of pressing the button with the left or right hand and a left-facing arrow for left-handed movement and a right-facing arrow for right-handed movement. In this experiment data were collected from a single subject with has no neurological dissorder. EPI pulse sequence was used for the scans (TR: 2700 ms, TE: 15, slice thickness of 1.5 mm).

This study is concentrated on how smooth and not smooth process affects the results of activation mapping for left hand movement.

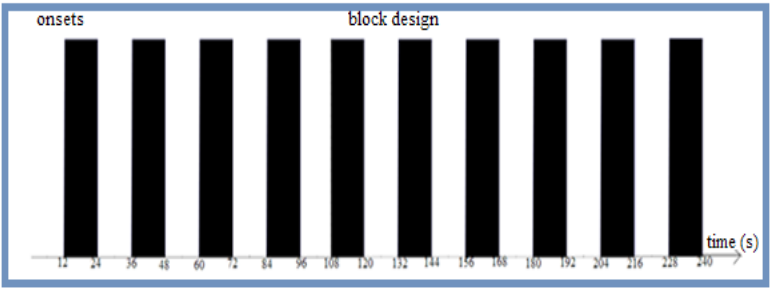


Figure 1. Block design with a 12 seconds rest (white block) as 12 seconds task (black block-left hand movement)

The experimental design as shown figure 1, defines the nature of the hypothesis testing to be applied. In this study a block design was used and 90 functional images were taken with this design, then the data was preprocessed and analyzed before the statistical analysis preprocessing steps applied to the data.
Figure 2 shows realignment results. This process is intended to minimize the effects of movement.

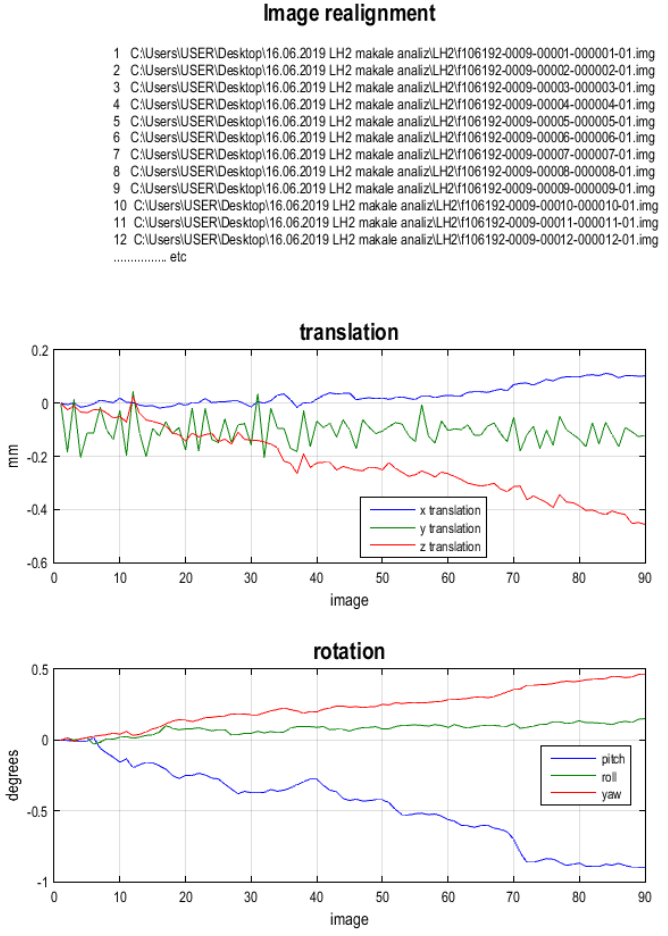


FIGURE 2. Realignment results for left hand movement data. Rotation and translation amounts are shown for 90 images

One of the most important preprocessing steps in fMRI is realignment. Since head motion is an huge problem while analyzing the data this helps to reduce the effect of movements for finding brain activations. After realignment, data normalized and smoothed for analyzing.

The General Linear Model (GLM) was used for statistical analyzing. GLM is the most often used model for functional magnetic resonance imaging [5]. Statistical parametric mapping is performed by placing the activation information on the structural MR images. In order to perform statistical parametric mapping, linear models are used which vary according to one or more explanatory variables. Test methods such as t-test, f-test, analysis of variance (ANOVA), correlation calculation are frequently used for parameter estimation in the General Linear Model.

The dependent variable $\boldsymbol{Y_i}$ in the GLM model has been predicted via the explanatory variables $\boldsymbol{X_i}$ and given by

$$\boldsymbol{Y_i} = \beta_{\mathbf{0}} + \boldsymbol{X_i}\beta + \varepsilon_i, \quad i = 1, 2, \ldots, n. \tag{2.1}$$

Here $\boldsymbol{Y_i}$ is the dependent variable such that $E(Y_i) = \beta_{\mathbf{0}} + \boldsymbol{X_i}\beta_{\mathbf{1}}$ and $Var(\boldsymbol{Y_i}) = Var(\varepsilon_i) = \sigma^2$ and $\beta_{\mathbf{0}}$ and $\beta_{\mathbf{1}}$ are the model parameters to be estimated. Estimated time series values $\hat{Y}$, including the linear combination of explanatory variables are calculated as $\boldsymbol{\hat{Y}=X\hat{\beta}}$, the residual values

$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} \tag{2.2}$$

Because error values contain both positive and negative values, the GLM procedure does not estimate beta values that minimize the sum of the error values, but finds these beta values by minimizing the sum of the square error values. Thus the sum of error differences for all time points given in eq. (2.3)

$$\sum_{i=1}^{n}(\mathrm{Yi} - \hat{\mathrm{Y}}\mathrm{i})^2 = \sum_{i=1}^{n}\varepsilon_i^2 \tag{2.3}$$

Least squares estimator of $\boldsymbol{\beta}$ parameter vector ($\hat{\boldsymbol{\beta}}$) given by eq.(2.4)

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X^T X}\right)^{-1}\boldsymbol{X^T Y} \tag{2.4}$$

The main assumption at the model has been done on the error terms ε_i. In order to estimate the parameters $\beta_{\mathbf{0}}$ and $\beta_{\mathbf{1}}$ it's assumed that ε_i's are independent and

identicaly distributed random variable with mean zero and constant variance σ^2. Moreover, for any statistical inference it's also assumed that the error terms are normally distiributed random variables for this an AR(1) model was performed during parameter estimation [6], [7], [8]. GLM parameters were estimated classical (Restricted Maximum Likelihood, ReML) parameter estimation.

The data analyzed with *Statistical Parametric Mapping8* (SPM8) programme [9] which is freely available for brain researchers. Single subject data analysis and parameters estimations were performed with the t test and for $p < 0.05$ active brain regions were marked.

3. Results and Conclusion

After perform preprocessing steps, data would be ready for the statistical analysis. In our work, images acquired with 1.5 x 1.5 x 1.5 mm resolution, the level of smoothing can be estimated as FWHM=3 mm.

In this section activation maps results were performed with and without smoothing. After 3 mm Gaussian kernel was used in smoothing stage, model specification and parameters estimations were done respectively. T test evaluates the effect of a parameter also to distinguish the contrasts of rest and task conditions used in the experiment. In order to identify the active voxels, the significance value of the test was determined as 0.05. In fMRI T contrast is one dimensional and can be expressed as

$$C^T = [1\ 0\ \dots\ ] \tag{3.1}$$

The contrasts can be specified to examine for conditions in the model. The sum of the contrasts of the conditions must be equal to 1. Main effects for active condition given as "1", (active > rest) and "-1" for rest condition (rest > active). High resolution anatomical images are used to show activation results after estimation (Figure 3).

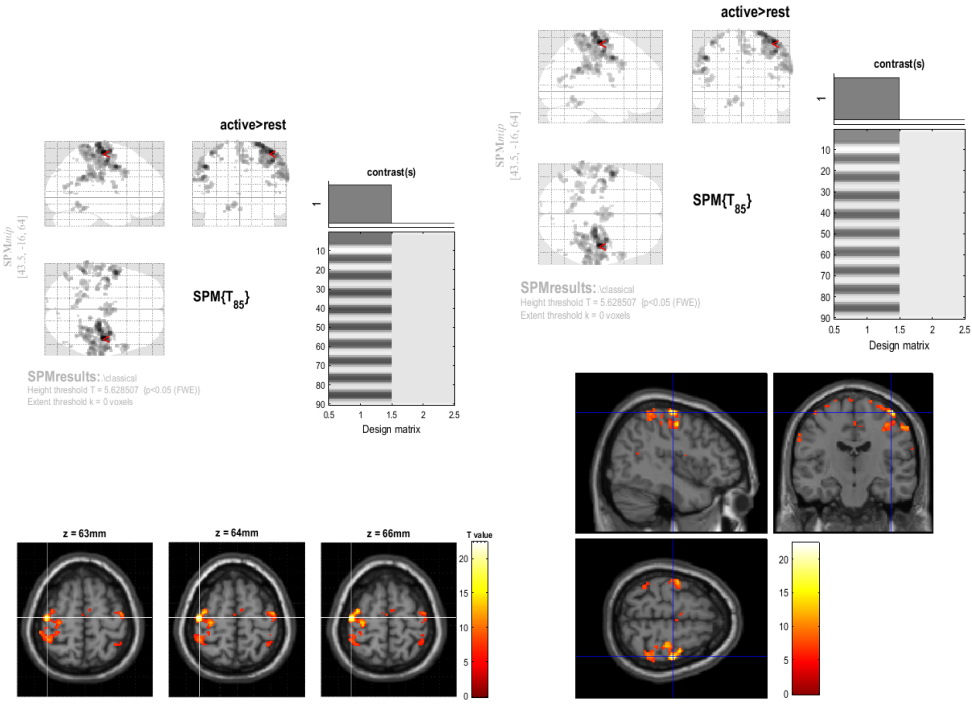


FIGURE 3. Statistical parametric mapping results using 3 mm smoothing and a significance threshold of $p < 0.05$

Left Hand Movement Activation Results with 3 mm Smoothing

The effects of the smoothing process can be applied differently on single or group fMRI data [8]. The goal is to improve the normality of the data with spatial smoothness. In general the smoothing process increases the signal to noise ratio [10], [11].

Statistical parametric maps shown in figure 3 was obtained in such a way that the activations would be on the right side of the motor cortex for the left hand movement.

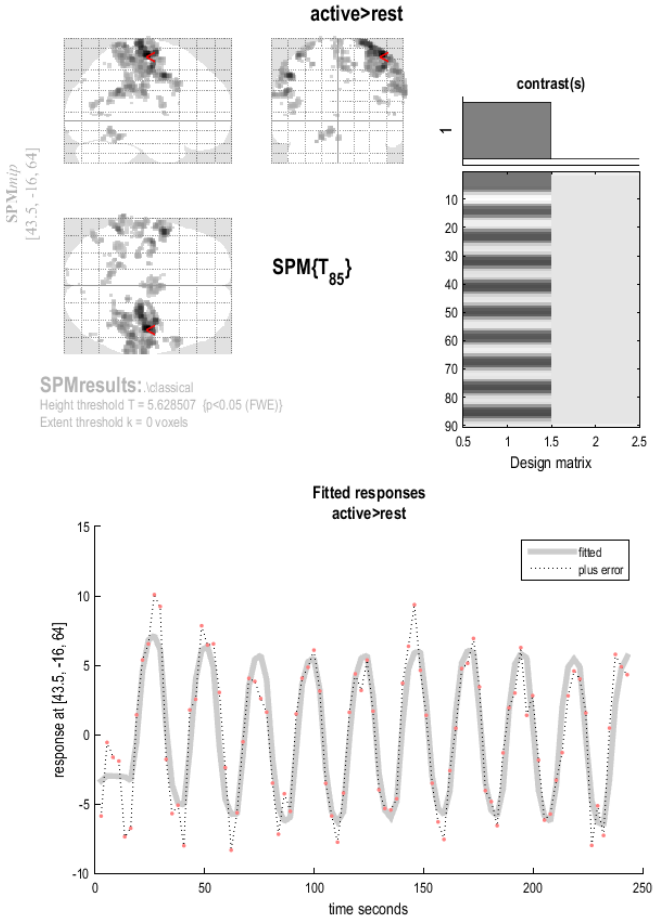
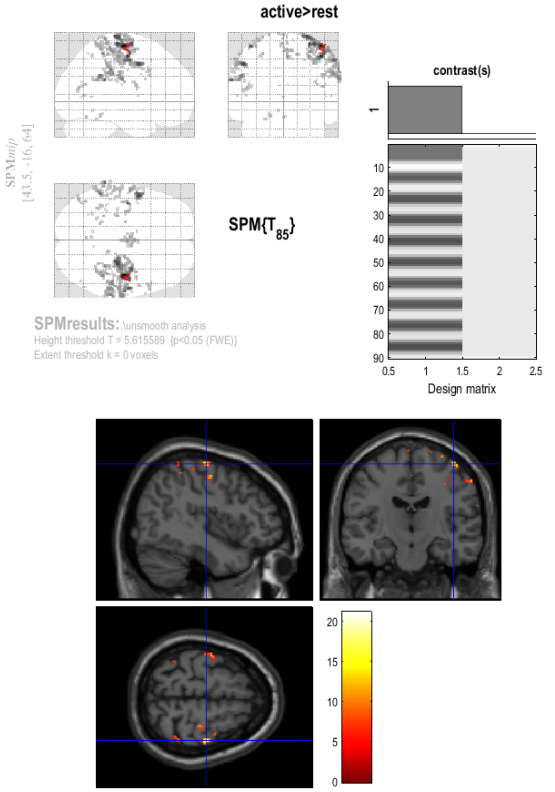


FIGURE 4. Time series for a selected active voxel

In the time series obtained for a selected active voxels, it is seen that the model and the results are compatible with each other. In the presence of the stimulus, a peak is observed in the signal, while in the case of rest, the signal shows a depression. In the graph, the time series is plotted with the addition of errors predicted with GLM (Figure 4).

Activation Results Without smoothing

All preprocessing steps were used without smoothing for same data. Statistical parametric mapps shows active brain regions were marked in figure 5 without smoothing.



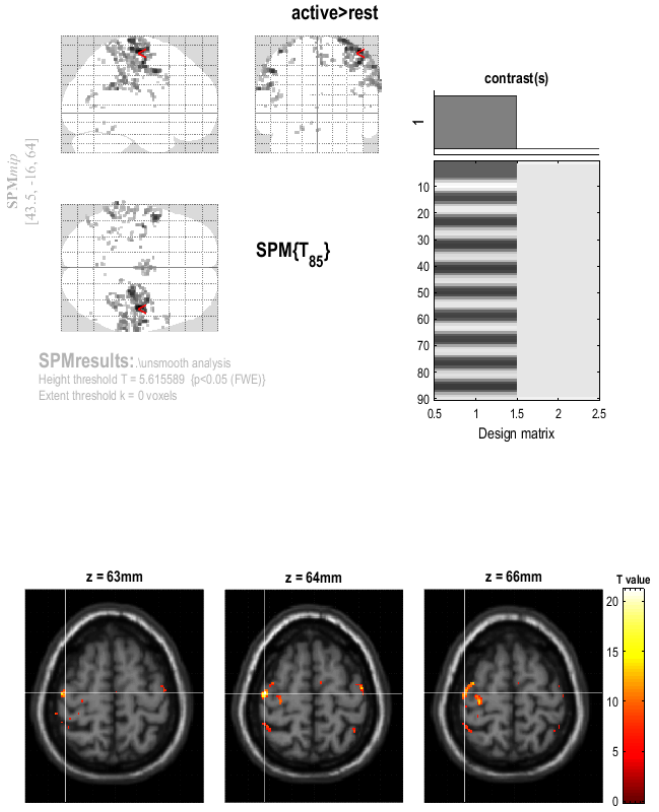


Figure 5: Statistical parametric mapping results using a significance threshold of p< 0.05 without smoothing

Results show that active voxel count decreasing without smoothing. The activation map results obtained with our data that is not applied smoothing in the pre-processing steps results less active voxels. For left hand movement task activation was achieved on the motor corteks right side similarly with results figure 3.

If the preprocessing steps are done correctly, this increases the functional resolution of fMRI experiments. Spatial smoothing is for applying a small blurring kernel across the images and for averaging densities from neighboring voxels together. Using a 3 mm smooth application allows us to obtain more active brain regions.

This process can be applied both the single-subject and the group analysis. fMRI data has a lot of noise in it, but studies have shown that most of the spatial noise is mostly Gaussian and is independent from voxel to voxel, and almost centered around zero.

Smothing makes the data more normal as statisticaly and also increases the signal to noise ratio (SNR). Thus smoothing process gives more reliability for single subject data analysis. Smoothing should be used in Neuroimaging to determine the active voxels correctly. In addition, these results can be used an alternative for studies on identifying specialized brain regions [12].

References

[1] Ogawa, S.T., Lee, M., Kay, A.R. and Tank, D.W., Brain Magnetic Resonance Imaging with Contrast Dependent on Blood Oxygenation. Proc. NatI. Acad. Sci., Biophysics (87), (1990), 9868-9872.

[2] Ugurbil, K., Garwood, M., Hendrich, K., Hinke, R., Hu, X., Menon, R.S., Merkle, H., Ogawa, S. and Salmi, R., Imaging at High Magnetic Fields; Initial Experiences at 4 Tesla. Magn Reson Quarterly, (9), (1993), 259 – 277.

[3] Kwong, K.K., Belliveau, J.W., Chesler, D.A., Goldberg, I.E., Weisskoff, R.M., Poncelet, B.P., Kennedy, D.N., Hoppel, B.E., Cohen, M.S. and Turner, R., Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proc. Nadl. Acad. Sci., (89), (1992), 5675-5679.

[4] Huettel, S.A., Song, A.W. and McCarthy, G., Functional Magnetic Resonance Imaging, Second Edition. Sinauer Associates Inc, Sunderland, Massachusetts USA, 2008.

[5] Friston, K.J., Holmes, A.P., Poline, J.B., Grasby, P.J., Williams, S.C., Frackowiak, R.S. et al. Analysis of fMRI time-series revisited, Neuroimage, 2(1), 1995, 45–53.

[6] Worsley, K.J., Marrett, S., Neelin P, Vandal, A.C., Friston KJ, Evans AC. A unified statistical approach for determining significant signals in images of cerebral activation. Hum Brain Mapp, 4(1), (1996), 58–73.

[7] Filippi, M., fMRI Techniques and Protocols. Humana Press, 830, New York, 2009.

[8] Worsley, K. J. and Friston, K.J., Analysis of fMRI Time-Series Revisited Again, Academic Press Inc., Neuroimage 2, (1995), 173-181.
[9] Anonymous, SPM By members & collaborators of the Wellcome Trust Centre for Neuroimaging, http://www.fil.ion.ucl.ac.uk/spm/.
[10] Mikla, M., Mareček, R., Hluštíkc, P., Pavlicová, M., Drasticha, A., Chlebus, P., Brázdil, M., Krupaf, P., Effects of spatial smoothing on fMRI group inferences. Magnetic Resonance Imaging 26, (2008), 490–503.
[11] Friston, K. J., Josephs, O., Zarahn E and Poline, P., To Smooth or Not to Smooth? Bias and Efficiency in fMRI Time-Series Analysis, NeuroImage 12, (2000), 196–208.
[12] Wilke, M. and Lidzba, K., LI-tool: A new toolbox to assess lateralization in functional MR-data. Journal of Neuroscience of Methods, (163), (2007), 128–136.

Current Address: Hacer DAŞGIN (Corresponding author): Ankara University Faculty of Science Department of Physics, Ankara, Turkey
E-mail: hdasgin@ankara.edu.tr
Orcid ID: https://orcid.org/0000-0003-4372-1134

Current Address: Ali YAMAN: Ankara University Faculty of Science Department of Physics, Ankara, Turkey
E-mail: yamana@ankara.edu.tr
Orcid ID: https://orcid.org/ 0000-0002-6859-762X

Current Address: Yılmaz AKDİ: Ankara University Faculty of Science Department of Statistics, Ankara, Turkey
E-mail: akdi@ankara.edu.tr
Orcid ID: https://orcid.org/ 0000-0003-0188-0970

Commun.Fac.Sci.Univ.Ank.Series A2-A3
Volume 61, Number 2, Pages 172-180 (2019)
DOI: 10.33769/aupse.593754
ISSN 1303-6009 E-ISSN 2618-6462
http://communications.science.ankara.edu.tr/index.php?series=A2A3

ROBUST REFRACTIVE INDEX FIBER SENSOR BASED ON TWO UP-TAPERS PLACED IN DOWN-TAPER

Isa NAVRUZ and Mustafa BİLSEL

ABSTRACT. In this study, a novel tapered optical fiber sensor based on a Mach–Zehnder interferometer (MZI) for refractive index measurement is proposed. Our sensor is constructed with two up-tapers symmetrically placed in a down-taper single mode fiber. Although its waist diameter is as thick as 55 μm, the sensor can measure the refractive index. Simulation results demonstrate RI sensitivities of -94 nm/RIU and -125 nm/RIU at the lower and higher wavelength of the spectrum. The sensor is extremely robust, it can be easily manufactured and it can be used not only for RI but also for simultaneous strain and temperature measurements.

1. INTRODUCTION

Fiber optical sensors offer some unique advantages compared to traditional electrical sensors such as small size, immunity to electromagnetic interference, high sensitivity, ease of fabrication, low cost, ability to operate in harsh environments, remote sensing, simultaneous measurement applications and so on. Refractive index [1-2], pressure [3-4], temperature [5], strain [6], chemical and biological sensing [7-8] are widely measured parameters that can be sensed by fiber sensors. Recently, combinations of down-taper fiber (DTF) and up-taper fiber (UTF) have been widely researched to obtain a more robust sensor. The most common methods to fabricate DTF sensors are electrical arc discharge [9], flame-brush technique [10], CO_2 laser drawing [11], chemical etching [12] and fusion splicer drawing [13]. UTF sensors are fabricated by using some of the commercial fusion splicers via cleaving and splicing two fibers by built-in mode through setting the overlap parameters without changing the other splicing parameters.

Received by the editors: July 18, 2019; Accepted: October 11, 2019.
Key word and phrases: Up taper, down taper, tapered fiber, refractive index

Previously, some sensor structures have been proposed for measurement of different parameters such as refractive index, strain, temperature, humidity, force etc. Zhao et al. [14] proposed a photonic crystal fiber interferometer which was formed by splicing a photonic crystal fiber between two single mode fibers having UTF joints. The sensitivity of this sensor was up to 252 nm/RIU. However, the sensing structure requires a special type of fiber so its cost is slightly high. Xiong et al. [15] fabricated an MZI based on concatenated two UTFs and a long period grating (LPG). UTF parameters were as follows: overlap was 150 μm, the diameter and length was 165 μm and 278 μm, respectively. LPG with a period of 550 μm and the length of 30 mm was fabricated by high frequency laser pulses. The corresponding sensitivity to measure refractive index in the range of 1.338-1.363 RIU was -108.16 nm/RIU but the fabrication of LPG is complicated and expensive. Furthermore Fu et al. [16] demonstrated a Michelson interferometer based on a UTF which is formed by pushing a single-mode and a multimode fiber while splicing. UTF excites several high order modes due to core mismatch and both core and high order cladding modes propagate in multimode fiber until it is reflected back from the end surface. Reflected modes recombine in UTF to interfere with each other. Sensitivity of this sensor reached -178.424 dB/RIU in the range between 1.351 and 1.4027 RIU. In a recent study, Zhang et al. [17] presented a UDF-DTF-UTF structure to measure refractive index, strain and temperature simultaneously. The length of the DTF was 250 μm and diameters of UTFs were 160 μm. The sensitivities of three dips in transmission spectrum were -131.93 nm/RIU, -22.875 nm/RIU and 0 nm/RIU in the range of 1.3211-1.3527 RIU. However, extinction ratio of the dip with highest RI sensitivity is too low. In another recent study, Ahsani et al. [18] proposed an MZI based optical fiber refractive index (RI) sensor constructed by uniformly tapering standard single mode fiber. The sensor with a cladding diameter of 35.5 μm and length of 20 mm exhibits RI sensitivity of 415 nm/RIU for RI range between 1.332 and 1.384. However, this sensor is not robust due to the thick cladding diameter.

Among the fabrication methods of DTF, electrical arc discharge come into prominence because it enables low cost, high reproducibility, disuse of hazardous chemicals and flexibility to configure desired sensor parameters. In this method, fabrication of tapered fiber requires stretching a fiber while heating it up to its softening temperature, typically 1500°C to compose a structure of diminishing both core and cladding diameter. In electrical arc discharge method, on the other hand, effective heat zone of electrodes has a direct unfavorable influence on adiabaticity. To overcome this problem DTF sensors are required to establish with small

diameters less than about 35 μm but, in this way, more robust sensor could not be achieved.

In this study, a more robust sensor design is proposed by concatenating a UTF and DTF with higher diameters more than 35 μm to enhance the robustness. The proposed sensor is shown in Figure 1. When light is penetrated into UTF region, high-order cladding modes are excited because of the mismatch of the mode field diameter and it causes light to split into two parts, one propagating in the core and the other in the cladding. Down tapering provides evanescent field to access the surrounding environment because the light is confined by the boundary between the taper and surrounding environment. Changing the parameters of the environment stimulates different types of modes supported by the taper, which results in a shift of wavelength spectrum as a function of surrounding parameter.

2. Sensing Principles

The sensor structure consists of three parts. The first is the long part, which is ordinary DTF. The second and third are short identical sections placed in the first part, which are UTFs. The schematic diagram of the sensor structure is shown in Figure 1.a. Two standard single mode fibers are spliced to sensor ends to connect to the light source and optical spectrum analyzer.

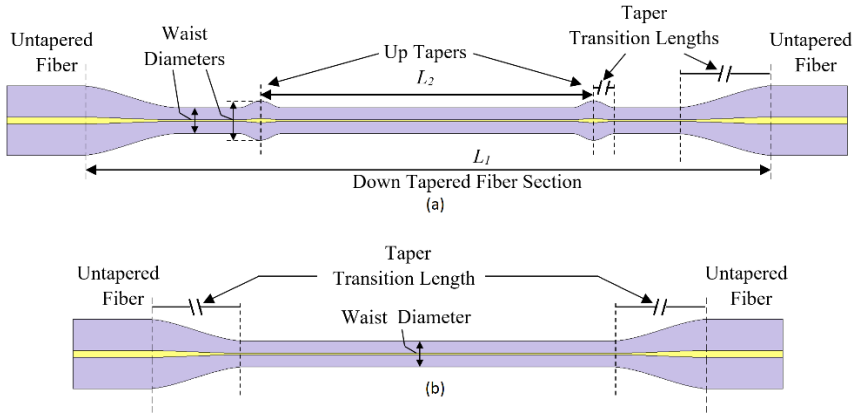


Figure 1. The schematic sensor structure, a) our sensor, b) Conventional DTF sensor.

Conventional DTF sensor structure given in Figure 1.b, similar to the long part of our sensor design, is fabricated by sufficiently reducing the fiber cladding diameter by using one of the methods mentioned in Introduction part. The cladding diameter in the thinnest region of the DTF is called waist diameter. The waist diameter can be reduced to less than 30-35 μm to obtain interference of forward guiding modes which cause resonant dips in wavelength spectrum. The changes in the measurands such as refractive index of the medium, temperature and stress will change the spectrum of light at the fiber output. However, measurable spectral changes can be achieved with thinner tapered SMF sensors. If the tapered fiber is further thinned, the optical power is preferably compressed into two modes leading to powerful oscillations in the spectrum. When the light is injected into tapered part of fiber, the transition section will excite the higher order modes. The interference of the optical modes in the waist region, which is thin enough, is formed according to not only the fiber guide but also by surrounding environment condition such as RI. The relative phase difference between two interfering modes can be expressed as;

$$\Delta\emptyset = \frac{2\pi}{\lambda}(\Delta n)L$$

where L is the length of the sensor section, λ is the wavelength of the light source and Δn is the effective refractive index difference. For the down tapered sensor with a sufficiently thin waist diameter, the fiber cladding acts as a core and surrounding medium of fiber replaces the cladding. In this case, the index difference can be expressed as $\Delta n = n_{eff}^{co} - n_{eff}^{sur,m}$. Here, n_{eff}^{co} and $n_{eff}^{sur,m}$ are the RI of core mode and mth-order surrounding medium mode, respectively. However, in the case of down taper single mode fiber with thick waist diameter, e.g. larger than 35-40 μm, propagating modes weakly form an interference.

The electric field amplitude distributions of two DTF sensors with different waist size are shown in Figure 2.a and 2.b obtained from beam propagation simulation results. Beam propagation simulation is a well-known technique in the literature and there are many commercial software. As the waist diameter increases, the amplitude of evanescent field decreases and thus the interference intensity is reduced as shown in Figure 2.b. This results in weak dips in the transmission spectrum and poor sensitivity. Therefore, the transmission spectrum of the sensor exhibits weak dips corresponding to the resonance wavelengths as given in Figure 3 where the waist size is reduced from 55 μm to 8 μm. Since poor interference leads to low resolution, the dynamic spectrum is desirable to obtain high resolution measurement. This can be achieved by the powerful interference of the fiber modes that propagate along sensor section.

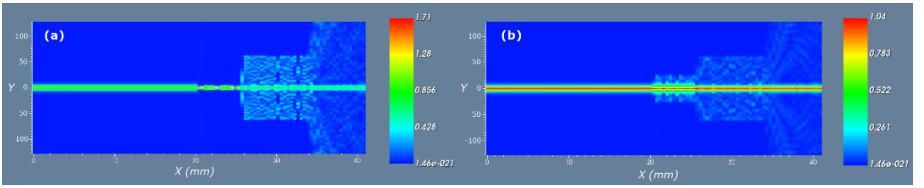


FIGURE 2. The electric field amplitude distributions of DTF sensor for different waist size, (a) 8 µm (b) 55 µm.

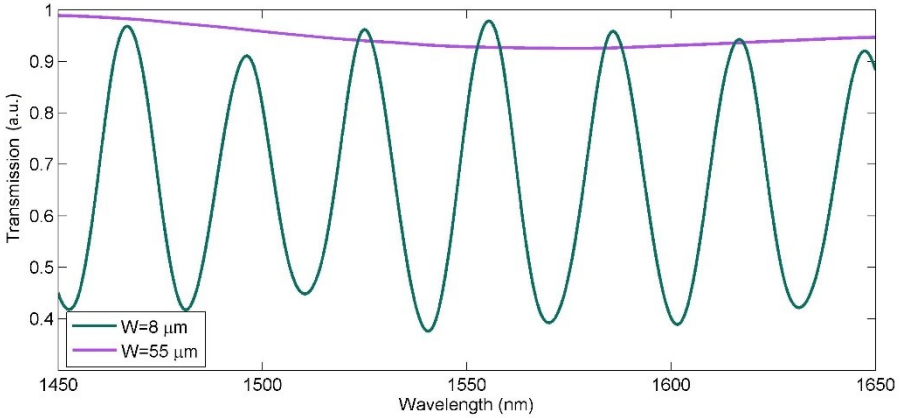


FIGURE 3. The transmission spectra of DTF sensor for waist size of 8 µm and 55 µm.

In this study, we propose to place two up-tapered sections in the down-tapered fiber with a length of 30 mm as shown in Figure 1. DTF region excites a few leaky modes, but these modes will be quite poor due to the thicker waist diameter so the interference spectrum originating from DTF itself will not be strong enough to sense the environment. The placement of UTF sections in the DTF region can strengthen leakly modes and the sensor designed as a MZI can be activated. The next section presents simulation results and discussion to test the RI sensitivity of the proposed sensor.

3. Results and Discussions

In this section, we carried out simulations of proposed fiber sensor by using beam propagation method in two dimensional structure. UTFs were positioned symmetrically with respect to the axis of symmetry in the center of the DTF section. The waist diameters UTFs and DTF were set to 85 µm and 55 µm, respectively. The

length of the DTF section, L_1 is 32 mm and the distance between the UTF tapers is 22 mm. The taper transition lengths of UTF and DTF sections were set to 300 and 1000 µm, respectively.

Figure 4 shows the transmission spectra of our sensor for different refractive indexes of environment. The spectrum was obtained for a wide wavelength range between 1.3 µm and 1.8 µm. Two dip points were observed in the spectrum where the transmission amplitude decreased to about 0.6 and less. As the refractive index increases the spectrum shifts to lower wavelengths called as blue-shift in contrast to conventional DTF sensors that experience red-shift.

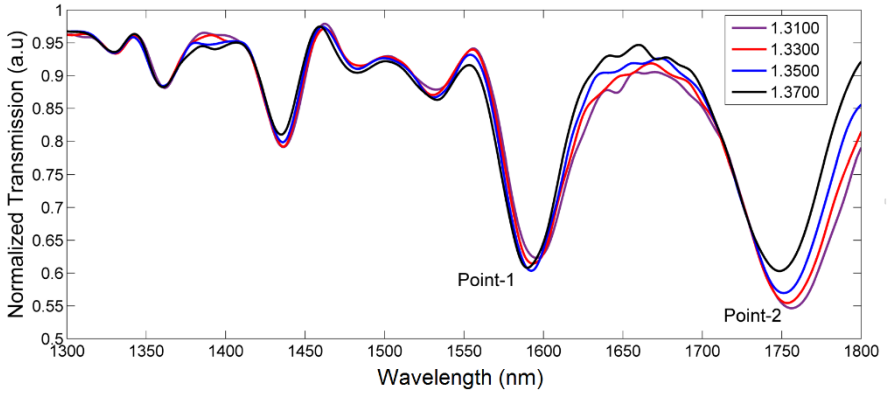


FIGURE 4. Normalized transmission spectrum obtained for RI range of 1.31-1.37.

In order to test RI sensitivity characteristic of sensor, four simulation data were obtained by changing the RI in the range of 1.31-1.37. The sensitivity can be defined as $\Delta\lambda/\Delta RI$ which is ratio of wavelength change in the transmission spectrum to RI change in the surrounding medium. Figure 5 presents wavelength shifts obtained from two dip points, namely Point-1 and Point-2.

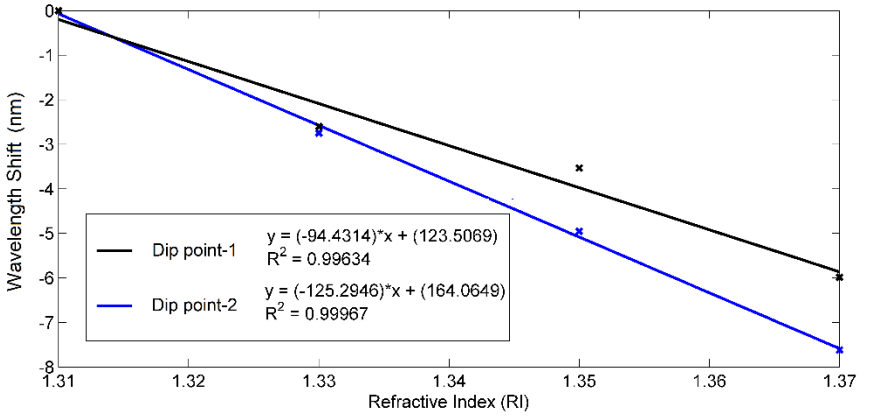


FIGURE 5. Sensor sensitivities calculated from point-1 and point-2.

As shown in the Figure 5, calculated sensitivities were about -94 and -125 nm/RIU at point-1 and point-2 corresponding to wavelengths of about 1590 nm and 1755 nm, respectively. It is clear that the dip point close to higher wavelengths exhibits higher sensitivity than the dip point at lower wavelengths. On the other hand, relation between RI and wavelength shift can be evaluated by a well-known parameter called linear correlation coefficient, R^2. Figure 5 indicates that our sensor has responded linearly to RI with a high linear correlation coefficient greater than 0.99.

4. CONCLUSION

In conclusion, a novel type of MZI fiber sensor based on placed two UTFs into the waist region of a DTF for measurement of refractive index is proposed and investigated theoretically in this paper. Simulation results show that first and last UTFs act as mode splitter and combiner, respectively, while the DTF is responsible for evanescent field to interact with surrounding environment. Furthermore, the results show that such an optical fiber sensor can work over a wide refractive index range of 1.31-1.37 with a sensitivity up to -125.29 nm/RI. Such a sensor structure including a UTF can be a good choice to satisfy the adiabaticity criteria without requiring thin waist diameters for DTF so that a robust sensors can be realized. Robustness and simple fabrication make it a good candidate for not only refractive

index measurement but also strain and temperature measurements in a wide range of potential applications such as chemistry, biology, biomedical or photonics.

References

[1] Ji, W. B., Liu, H. H., Tjin, S. C., Chow, K. K. and Lim, A., Ultrahigh sensitivity refractive index sensor based on optical microfiber, *IEEE Phot. Tech. Lett.*, 24 (2012) 1872-1874.

[2] Lacroix, S., Gonthier, F., Black R. J., and Bures, J., Tapered-fiber interferometric wavelength response: the achromatic fringe, *Opt. Lett.*, 13 (1988) 395-397.

[3] Bariáin, C., Matías, I.R., Arregui, F. J. and López-Amo, M., Tapered optical-fiber-based pressure sensor, *Opt. Eng.*, 39 (2000) 2241-2247.

[4] Massaro, A., Spano, F., Missori, M., Malvindi, M. A., Cazzato, P., Cingolani, R. and Athanassiou, A., Flexible nanocomposites with all-optical tactile sensing capability, *RSC Adv.*, 4, (2014) 2820-2825.

[5] Lu, P., Men, L., Sooley, K. and Chen, Q., Tapered fiber Mach–Zehnder interferometer for simultaneous measurement of refractive index and temperature, *Appl. Phys. Lett.*, 94, (2009) 131110-131110-(3).

[6] Shao, L-Y., Zhao, J., Dong, X., Tam, H. Y., Lu, C. and He, S., Long-period grating fabricated by periodically tapering standard single-mode fiber, *App. Optics*, 47 (2007) 1549-1552.

[7] Bariáin, C., Matías, I. R., Fernández-Valdivielso, C., Arregui, F. J., Rodríguez-Méndez, M. L. and Saja, J. A., Optical fiber sensor based on lutetium bisphthalocyanine for the detection of gases using standard telecommunication wavelengths, *Sens. Act. B Chem.*, 93 (2003) 153-158.

[8] Corres, J. M., Matías, I. R., Bravo, J. and Arregui, F. J., Tapered optical fiber biosensor for the detection of anti-gliadin antibodies, *Sens. Act. B, Chem.*, 135 (2008) 166-171.

[9] Yang L. Jiang, J., Wang, S., Li, B., Wang, M., Xiao, H., Lu, Y. and Ysai, H., High sensitivity of taper-based Mach-Zehnder interferometer embedded in a thinned optical fiber for refractive index sensing, *Appl. Opt.*, 50 (2011) 5503–5507.

[10] Layeghi, A., Latifi, H. and Frazao, O., Magnetic field sensor based on nonadiabatic tapered optical fiber with magnetic fluid, *IEEE Phot. Tech. Lett.*, 26 (2014) 1904-1907.

[11] Davis, D. D., Gaylord, T. K., Glytsis, E. N., Kosinski, S. G., Mettler, S. C. and Vengsarkar, A. M., Long-period fibre grating fabrication with focused CO2 laser pulses, *Elect. Lett.*, 34, (1998) 302-303.

[12] Brophy, T. J. and Shankar, P. M., Formation and measurement of tapers in optical fibers, *Rev. Sci. Instr.,* 64 (1993) 2650-2654.

[13] Shi, J., Xiao, S., Yi, L. and Bi, M., A Sensitivity-Enhanced Refractive Index Sensor Using a Single-Mode Thin-Core Fiber Incorporating an Abrupt Taper, *Sensors*, 12, (2012) 4697-4705.

[14] Zhao, Y., Li, X., Cai, L. and Yang, Y., Refractive index sensing based on photonic crystal fiber interferometer structure with up-tapered joints, *Sens. And Act. B:Chemical*, 221 (2015) 406-410.

[15] Xiong, M., Gong, H., Wang, Z., Zhao, C-L. and Dong, X., Simultaneous refractive index and temperature measurement based on Mach-Zehnder interferometer concatenating two bi-tapers and a long-period grating, *IEEE Sens. J.*, 16 (2016) 4295-4299.

[16] Fu, H., Zhao, N., Shao, M., Xu, Y., Huidong, L., Liu, Q., Gao, H., Yinggang, L. and Qiao, X., In-fiber Quasi-Michelson interferometer based on waist-enlarged fiber taper for refractive index sensing, *IEEE Sens. J.*, 15 (2015) 6869-6874.

[17] Zhang, N., Xu, W., You, S., Yu, C., Yu, C., Dong, B. and Li, K., Simultaneous measurement of refractive index, strain and temperature using a tapered structure based on SMF, *Opt. Comm.*, 410 (2018) 70-74.

[18] Ahsani, V., Ahmed, F., Jun, M. B. G. and Bradley, C., Tapered Fiber-Optic Mach-Zehnder Interferometer for Ultra-High Sensitivity Measurement of Refractive Index, *Sensors*, 19 (2019) 1-10.

Current Address: Isa NAVRUZ: Ankara University, Faculty of Engineering, Department of Electrical and Electronics Engineering Ankara TURKEY
E-mail: inavruz@ankara.edu.tr
Orcid ID: https://orcid.org/0000-0003-2976-076X

Current Address: Mustafa BİLSEL: Ankara University, Faculty of Engineering, Department of Electrical and Electronics Engineering Ankara TURKEY
E-mail: mfbilsel@gmail.com
Orcid ID: https://orcid.org/0000-0001-8252-3734

Commun.Fac.Sci.Univ.Ank.Series A2-A3
Volume 61, Number 2, Pages 181-196 (2019)
DOI: 10.33769/aupse.557951
ISSN 1303-6009 E-ISSN 2618-6462
http://communications.science.ankara.edu.tr/index.php?series=A2A3

OPERATION TESTS of the 260 MHz 1500 W SOLID STATE RF AMPLIFIER at TARLA FACILITY

Ozlem KARSLI and Evrim COLAK

ABSTRACT. Turkish Accelerator and Radiation Laboratory (TARLA) will be the first accelerator-based user facility in Turkey. The facility is under construction at the Institute of Accelerator Technologies of Ankara University. Based on the state-of-art superconducting technology, TARLA accelerator offers a multi-experimental facility providing a variety of accelerator-based radiation sources for users coming from various fields like chemistry, physics, biology, material sciences, medicine and nanotechnology. TARLA consists of two acceleration lines: the first one is the injector that provides high current continuous wave (CW) electron beam at 250 keV energy, and the second one is the main accelerator that comprises of two superconducting (SC) cryomodules separated by a bunch compressor in order to accelerate the electron beam up to 40 MeV energy. Two normal conducting accelerators, so called subharmonic (SHB) and fundamental (FB) buncher cavities whose operation frequencies are 260 and 1300 MHz, respectively, are used to compress the electron bunches from ~600 ps to ~10 ps. SHB cavity is powered by a 1500 W Radio frequency (RF) power amplifier. Currently, the electron gun training, and superconducting modules acceptance tests, personal safety system, and helium cryogenic system commissioning tests are performed simultaneously. In this study, we present the operation tests of the 1500 W RF amplifier in the scope of the commissioning tests of injector line which showed phase drift coefficients of ~0.5 deg/C^o and ~0.67 deg/C^o in repeated tests. Moreover, the importance of constancy of the water pressure in the water-cooling line for phase constancy of the delivered power has become evident as a result of current observations.

1. INTRODUCTION

Turkish Accelerator and Radiation Laboratory (TARLA) has been proposed as an oscillator mode infrared (IR) free electron laser (FEL) facility in the scope of Turkish

Received by the editors: April 26, 2019; Accepted: December 11, 2019.
Key word and phrases: Radio frequency (RF), solid state amplifier, accelerator, bunch length, phase stability, high power, operation test.

Accelerator Center (TAC) Project [1] and it is under construction since 2012 [2]. After submitting the conclusion reports of TAC project, the installation of the facility has been continuing as the TARLA project since 2016 [3].

TARLA will be the first accelerator based research center in Turkey that produces electrons up to 40 MeV with a tunable infrared free electron laser (FEL) source in 3-350 µm range with high peak (~5 MW) and high average power (~100 W), and a Bremsstrahlung radiation between 5-30 MeV energy [3]. The facility will serve four FEL experimental stations for researchers to study material science, infrared spectroscopy, biomedical, chemistry, nanotechnology and semiconductor. Additionally, researchers will have an opportunity to perform fixed target experiments with Bremsstrahlung laboratory. The schematic view of TARLA is illustrated in Fig. 1[2].

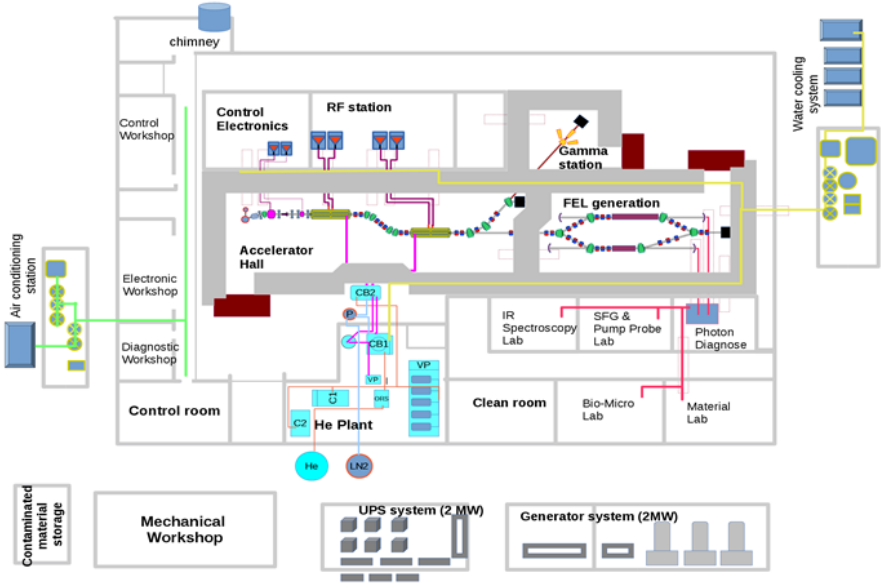


FIGURE 1. General view of TARLA facility [2].

Currently, acceptance tests of superconductive accelerator modules, commissioning tests of personal security system, and helium cooling system, electron gun training tests are carried out, simultaneously.

TARLA injector line consists of an electron gun, subharmonic buncher (SHB) and fundamental buncher (FB) cavities, a macro-pulser, and solenoid and steering magnets, in basic. Electron gun is a thermionic type of DC gun and is composed of a tungsten dispenser cathode and a grid providing a ~600 ps pulsed electron beam in 13 MHz repetition rate [4]. Electron beam should be formed into bunches in the

injector line, before entering the superconducting modules. A buncher cavity is a normal conducting accelerating cavity. It allows the particles to be either accelerated or decelerated depending on the phase difference. It also provides velocity modulation which is called as ballistic bunching [5]. TARLA injector line has a SHB and a FB that operates at frequency of 260 MHz and 1300 MHz, respectively. The longitudinal waist of electrons are formed into from ~600 ps to 10 ps by ballistic bunching in buncher cavities. A magnetic chicane, so called bunch compressor, will be located between two superconducting modules. Electron beam energy will be increased up to 40 MeV after the second cryomodule [3].

Buncher cavities match the longitudinal characteristics of electron beam before entering the superconducting modules operating at 1300 MHz [4]. SHB is powered by a 1500 W Radio Frequency (RF) power amplifier operating at 260 MHz. Basic parameters of the RF amplifier to feed the SHB are listed in Table 1 [6, 7].

TABLE 1. Basic parameters of RF amplifier feeding the SHB [6, 7].

Parameter	**Value**
Amplifier	Class AB push-pull
Operation frequency	260 MHz
Bandwidth (-3 dB)	<±5 MHz
Output power (@1 dB compression) pulse or CW	1300 W@1dB and 1500W saturated
Input / Output VSWR	<1.5 /<1.5 with measurement results
Linear gain	>61 dB (min)
Pulse length	10 us to CW
Pulse repetition rate	1 Hz to CW
Rise / Fall time	<60 ns
Phase drift coefficient	0.35 deg / ºC
Gain drift coefficient	1.2 % / K typ. Max.
Output phase change from min to max power	±10 deg (max)
System efficiency	> 55 % at 1500 W output

Three types of RF amplifiers have been used in acceleration science area: klystrons, inductive output tubes (IOT), and solid-state power amplifiers (SSA). Simple start-up and low-cost maintenance procedures, the high modularity with associated redundancy and flexibility, removal of a high-voltage and high-power circulator are the features that make solid-state RF amplifiers superior to others [8]. The solid-

state amplifiers exhibit extremely fast pulse rise and fall times, very tight pulse-to-pulse repeatability and excellent ruggedness [9].

Class B amplifiers are commonly used due to their simplicity and high efficiency. It is possible to operate in a broad range of frequency, if the wideband transformers are used. However, this type of amplifiers cause significant amount of distortion and harmonics due to the unmatched transistors and crossover distortion. Transistor mismatch can be avoided by using transistors in uniform characteristics. The crossover distortion can be reduced by both transistors being normally biased at a level that is slightly above cutoff (forward bias). The forward bias causes the circuit to operate in Class-AB mode, so both transistors are slightly 'On' during crossover. Under these conditions, the Class B amplifier is typically referred to as the Class AB amplifier [10].

Therefore, Class AB push-pull amplifiers have been chosen for TARLA as the high power RF sources. The parameters specified in Table 1 are designated by taking beam dynamic calculations and the characteristics of SHB into consideration. Gain, the logarithmic ratio of the output and the input power, is ~61 dB at minimum with 1 mW input power to feed the SHB. The electron bunch length which will be delivered from the SHB, heat load capacity of the antenna and the cooling parameters of the SHB determine the output power of the RF amplifier. The frequency and bandwidth of the RF amplifier should be the same with those of the SHB. Cooling parameters of SHB affects the operation temperature and cooling requirements of the RF amplifier. The rise and fall times in the order of nanoseconds are two of the important criteria for selecting solid-state RF amplifiers. Reduced rise and fall times of the amplifiers will increase the interaction between electron beam and RF wave in the SHB. Maximum output Voltage Standing Wave Ratio (VSWR) is found as 1.81 with the assumptions of 100 W reverse and 1.2 kW forward power. Regarding system efficiency, 55 % is a typical value for SSAs. These parameters were discussed with the manufacturers to find out whether such a SSA was producible or not. This amplifier module was manufactured by SigmaPhi [11] with a spare one and was delivered in 2017.

Phase and gain drift coefficients of RF amplifiers are of vital importance since the amplifiers are to operate in steady state without interruption for a long time period and they are expected not to be affected by ambient temperature changes.

In this study, we measure phase and magnitudes of the scattering parameters (S-parameters), namely, S_{11} and S_{21} for the 1500 W RF amplifier for long-time operation to observe the phase drift. The effect of variation in the temperature and the pressure of the cooling water system on the measured S-parameters is

investigated. Utilizing a home-made test routine written in NI LabVIEW Software [12] to automate the phase and magnitude measurements, the tests are repeated twice for 50 hours and ~180 hours and the results are compared with each other.

Section I is devoted to literature and basic description of the accelerator facility. Mathematical relationship between the phase of the RF signal and the parameters of the beam which is accelerated is discussed in Section II. We explain the measurement setup and data acquisition & controlling method in Section III. Section IV is dedicated to the presentation of the measurement data. The comparison of the results and the discussion of the effects of external factors such as cooling temperature and pressure level are discussed in Section V, i.e., the Conclusion Section.

2. Theoretical Background

Let us consider work done (ΔE) on a moving particle with charge q in a cavity, by the electrical field $\vec{\mathcal{E}}$ which applies force ($\vec{F}_e = q\vec{\mathcal{E}}$) along a path from $z = -d/2$ to $z = d/2$. Within the cavity, the acceleration field is applied along the z-direction, therefore, E_x and E_y, x- and y-components of the electric field are negligible. Given operation frequency ω and the electrical field phasor $\vec{\mathcal{E}} = E_z(z)\mathrm{e}^{\mathrm{j}\omega\mathrm{t}}$, the beam is accelerated only along the propagation direction (z-axis of the cavity). Such a particle entering the cavity at $z = -d/2$ will encounter the phase of the RF field. Then, the energy gain, i.e., the distribution in phase space, for the particle is given as the integral of applied force along the path [13]:

$$\Delta E = q\Re e\left\{\int_{-d/2}^{d/2} E_z(z)e^{j\omega t}dz\right\} \tag{1.1}$$

with

$$\omega t = \omega\frac{z}{v} + \psi_p \tag{1.2}$$

where ψ_p is the phase relative to RF and v is the speed of the particle when entering the gap, respectively. Hence,

$$\Delta E = q\Re e\left\{e^{-j\psi_p}\int_{-d/2}^{d/2} E_z(z)e^{j\omega\frac{z}{v}}dz\right\} \tag{1.3}$$

By introducing $\phi = \psi_p - \psi_i$, where ψ_i is the phase relative to initial position, one finally gets:

$$\Delta E = q\left|\int_{-d/2}^{d/2} E_z(z)e^{j\omega\frac{z}{v}}dz\right| cos\phi \tag{1.4}$$

where ϕ appears as the phase of the particle referred as the particular phase, which would yield the maximum energy for $\phi = 0$.

Thus, one can deduce the following: the energy that is transferred to the particle will decrease if $\phi > 0$, i.e., the phase difference between the RF wave powering the cavity and the particle, is greater than zero while the particle travels across the cavity. Since the beam emittance is affected by the phase change and the energy gain, $\phi > 0$ condition results in the deviation of the beam during its transport through the entire beamline as well as energy spread and emittance growth [4, 14].

ψ_p is optimized to minimize the energy spread, where $\frac{\Delta E}{E}$ is the energy spread of the particles [15]. The residual energy spread after compensation is found from the convolution of the bunch with the longitudinal wake-field and the acceleration RF [15],

$$\frac{\Delta E}{E} \approx \frac{1.25}{Gcos\psi_p}\left[KGsin\psi_p\right](FWHM) \tag{1.5}$$

where *K* is a variable depending to the electron beam and RF, and $G = qV_{RF}/m_e c^2$, where m_e is electron mass and c is the speed of light [15]. TARLA has set a design criterion for resolution in energy of the exit beam such that $\frac{\Delta E}{E} < 5x10^{-4}$.

To understand the stability of the output signal phase of the RF amplifier (RF-PA), S-parameters which give information about the phase and amplitude of the electromagnetic wave are to be examined [16]. It is defined as $S_{ij} = {V_i^-}/{V_j^+}$ for an *N*-port system ($\{N, i, j\} \in \mathbb{N}^+$) where V_i^- is the signal leaving port-*i* whereas V_j^+ is

the signal entering port-*j*, Specifically, the reflection coefficient (S_{11}) and transmission coefficient (S_{21}) give information about the amplitude and phase of the reflected power from port-1 and transmitted power (gain) from port-1 to port-2 of the vector network analyzer (VNA), respectively. The signal measured by VNA is directly proportional to electric and magnetic fields in the transmission line.

3. Experimental Setup and Data Acquisition

The layout of the experimental set-up is illustrated in Fig. 2(a). The experimental setup utilizes an Anritsu MS4640A VectorStar™ VNA which is controlled by a personal computer (PC). The VNA's port-1 signal feeds the 1500 W high power RF-PA which operates at the center frequency of 260 MHz. The detailed description of the setup is given in Table 2.

The power delivered to the load is sampled through the DC coupler (with attenuation 40 dB) which is in series with another 30 dB attenuator. Thus, total attenuation for the RF output power before being measured by the VNA's port-2 is 70 dB. The phase and amplitude of S_{11} (reflected signal) and S_{21} (transmitted signal) are recorded by the PC for each sample taken periodically during the experiment. The picture of the experimental setup is given in Fig. 2(b). The number of samples, sampling period, and the data acquisition (including the initialization and remote control of the VNA, adjustment of the power level, type of the output format i.e., Smith Chart as well as S_{11} and S_{21} measurements in log/linear scale selection etc.) are conditioned using testing software (TS) written in NI LabVIEW Software [12] (Fig. 3). TS also enables the instantaneous plots of the measured quantity on the available 4 windows in the user interface. Thus, in our experiments, the phase and amplitude of S_{11} and S_{21} for the most current sample can be observed by the operator. At the same time, these raw data obtained for each sample are recorded as a .txt file being automatically named by the TS. Finally, once all the samples are obtained, the acquired data are read from the recorded .txt files and are analyzed using MATLAB software [17].

4. Measurement Results

Once the test is run and all the samples are saved to the individual .txt files, the raw data are read by a prepared MATLAB script file. On the other hand, TARLA has a water cooling system controlled by a SCADA system which is also used to follow the water parameters (temperature, pressure with time stamp) on-line. The water

temperature data are kept in this system in weekly periods. The MATLAB script is also used to extract the time stamps of the recorded .txt files which include the phase and amplitude of S_{11} and S_{21}. Then, variation of S-parameters, temperature and pressure as a function of time can be observed together.

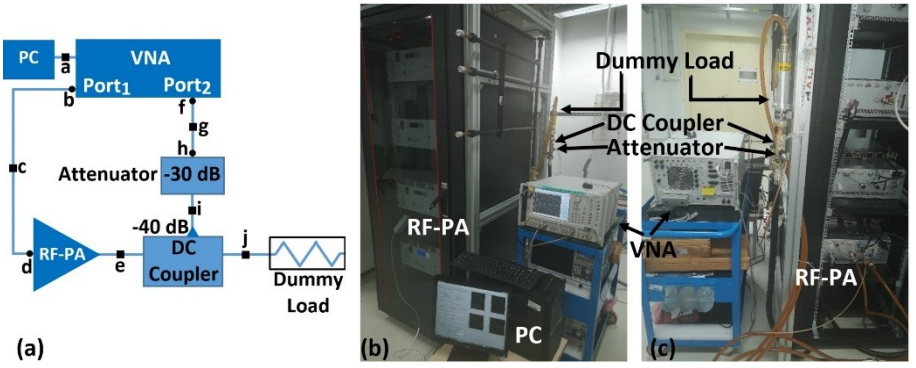


FIGURE 2. a) Block diagram of the experimental setup (ES), b) Picture of ES (Front side view), (c) Picture of ES (Back side view).

To avoid the spontaneous power shut down at the output of RF-PA which occurs due to the interlock as a result of fluctuation of the excitation signal, output signal of VNA from port-1 is set deliberately to relatively a lower value (-2 dBm) less than the maximum input signal capability of the RF-PA, which is 0 dBm. This setting yields ~1200 W output power from the RF-PA. One more reason for the selection of 1200 W is that the power which will be sent to SHB during real operation is also 1200 W. In this configuration, the reflected power shown on the display of the RF-PA is ~5 W. Leaks from connection points and cable ends, bending of high power RF cable for aligning it to its operation position act as both reflection and leakage points. These can cause increase in the reflected power and decrease in the delivered power.

In the first run of the experiment, we took 2794 many samples which resulted in a ~50-hour measurement. The reflection and transmission data obtained in this experiment are plotted in Fig. 4 thru Fig. 6 together with the plot of water temperature (on the right-hand side axis) as a function of time to examine the effect of water temperature on the response of the experimental setup.

Time information of all activities such as maintenance and system upgrade, etc. that interact with the water cooling line was recorded manually. The vertical red dash-

dotted lines, from 1 to 7, in Figs. 4 and 5 show such recorded interferences with the water cooling line at different time instants during the test. For instance, the chillers were being commissioned in the vicinity of time instants that are pointed out by lines 1 and 2. In the time interval in which line 3 resides, the pressure of force manifold varied and line 4 is to mark the time instant that the pressure of lift manifold was increased.

TABLE 2. Apparatus description for the experimental setup.

Label	Type	Description
a	cable	Standart USB cable.
b	port	Port-1 of VNA; N-type female.
c	cable	RG142 RF, N-type male connectors on both ends; 2 m long; with an N-type female adapter on the VNA side.
d	port	Input port of RF-PA, N-type female.
e	cable	LCF78-50JA low loss, phase stable, high power RF cable, 7/8" male connectors on both ends with 7/16" female connector at RF-PA side and 15/8" EIA at DC Coupler side.
f	port	Port-2 of VNA; N-type female.
g	cable	RG142 RF N-type male connectors on both ends; 10 m long; with an N-type female adapter on the VNA side.
h	port	Output port of 30 dB attenuator; N-type male.
i	connection	Direct series connection between the 30 dB attenuator and the sampling ouput port (N-type female) of the DC coupler.
j	connection	Direct series connection (1-5/8" EIA) between the ouput port of the DC coupler and the dummy load which is cooled down with water that is run through the water cooling system.

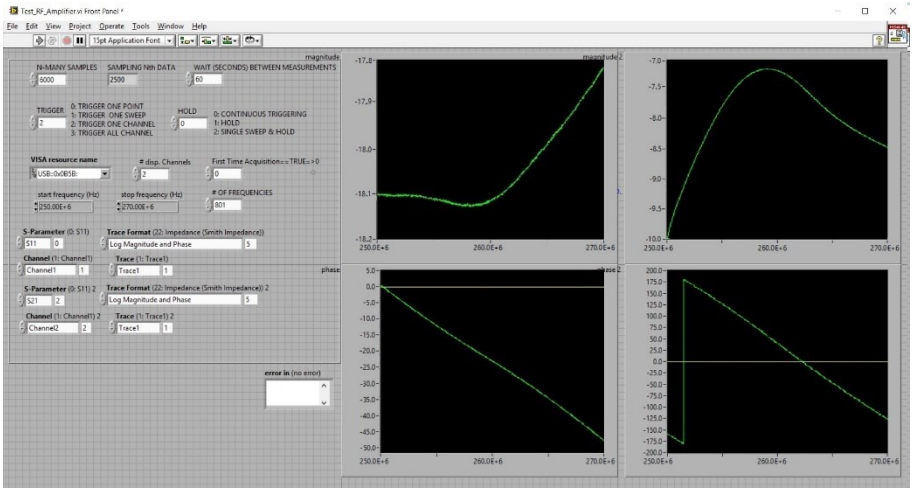


FIGURE 3. NI LabVIEW user interface which can control the VNA remotely, initialize the test, and save test data.

Line 5 indicates the pressure fluctuation in the water cooling line. The pressure level drop of the water cooling line becomes evident at the time instant that line 6 indicates. Maintenance was being performed around the time instant 7.

1200 W (~+60.79 dBm) output power of RF-PA is also verified with another digital power meter (R&S®NRP2 model). After attenuating this output power by 70 dB to protect the VNA, the power which is sampled by the VNA's port-2 becomes ~-9.21 dBm. Remembering that the input signal is -2 dBm, $|S_{21}|$ for the RF-PA should be measured as ~-7.21 dB (~-9.21 dBm-(-2 dBm)). This is in accordance with the experimental result given in Fig. 5(a) where $|S_{21}|$ is measured as ~-7.16 dB at steady state.

Large ripples in the magnitude and phase of S_{11} are evident in Fig. 4(a) and Fig. 4(b), respectively. These fluctuations are also observed in the magnitude and phase plots of S_{21} at the same time instants (Fig. 5(a) and Fig. 5(b)). Accordingly, it can be deduced that all pressure fluctuations at the input through the water cooling line will reflect as large ripples in the magnitude and phase of the output power. The water-cooling operation temperature of the amplifier is seen to slightly vary around 19.4°C in Fig. 5. In addition to the effect of water line pressure, the phase graph of S_{21} in Fig. 5(b) illustrates that all temperature variations in water cooling line triggers the phase change at the output.

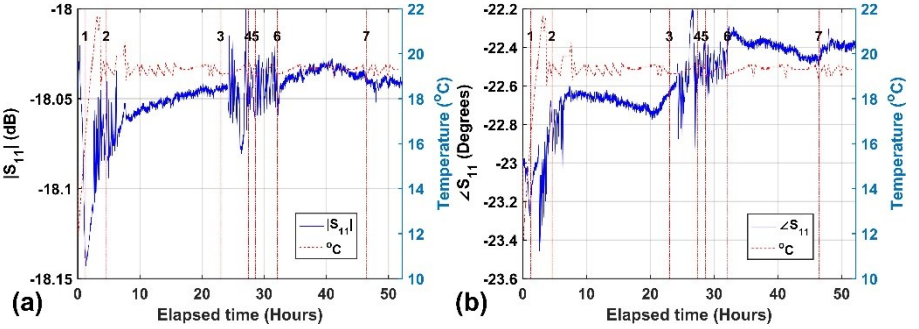


FIGURE 4. Measurement results of S_{11}. a) magnitude ($|S_{11}|$): (solid blue) b) phase ($\angle S_{11}$) (solid blue) together with water cooling line temperature (dashed red) on the right-hand side axis as a function of elapsed time in the experiment. Dash-dotted vertical lines are to show the selected time instants described in the text above.

As a result of Eqn. (1-5), it is desired to have minimum variation of phase in the output power. Under optimum conditions, i.e., with minimal variation in the water line temperature and pressure, one can see in Fig. 5(b) that there is a minimal phase variation between hours 15.5 and 21. Fig. 6 provides with a closer look at the mentioned situation in Fig. 5(b), presenting the phase variation of output power in the mentioned time interval between hour 15.5 thru hour 21. It is seen that the phase variation in this period is about 0.1 deg/0.2 °C ≈ 0.5 deg/°C.

These measurements were repeated several times and similar results were obtained. A representative of the repeated measurements is depicted in Fig. 7 and Fig. 8 using the results of another ~180-hour long measurement. In Fig. 8(a), again, $|S_{21}|$ converges to ~-7.17 dBm which is similar to the steady state value observed in Fig. 5(a). A closer view of S_{21} for the phase measurement in Fig.8 (b) is given in Fig. 9. The water cooling temperature plot is in line with the phase variation plot of S_{21}. Figure indicates a closer view for understanding this relationship.

The phase variation of output power around hour 95, which is considered to correspond to the most stable operation from Fig. 9, in the second measurement is about 0.4 deg/0.6 °C ≈ 0.67 deg/°C.

The variation in the output power's phase has made the following necessity evident for us as a result of the mentioned measurements above: requirement of integrating a phase shifter for the feeding circuit of the SHB to control the phase variation.

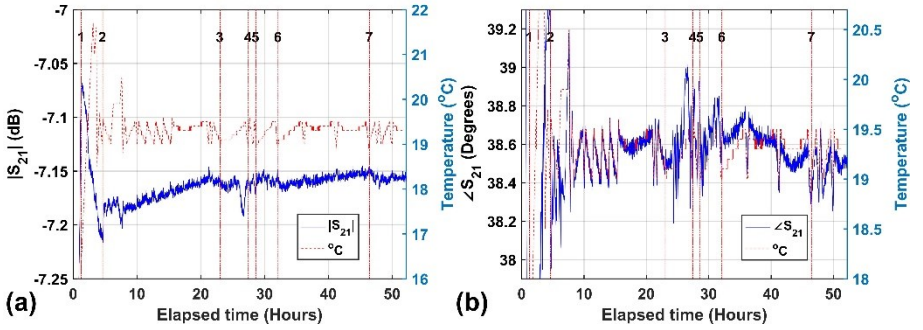


FIGURE 5. Measurement results of S_{21}. a) magnitude ($|S_{21}|$): (solid blue) b) phase ($\angle S_{21}$) (solid blue) together with water cooling line temperature (dashed red) on the right-hand side axis as a function of elapsed time in the experiment. Dash-dotted vertical lines are to show the selected time instants described in the text above.

5. CONCLUSION

The operation studies of SHB within the commissioning of injector line studies have been continuing and it is in accordance with the TARLA injector line commissioning scope.

This study shows that not only the temperature variations [18-20], but also the pressure stabilization of cooling (water) line is of importance for minimizing the phase variation of the output power of the amplifier. The effect of variation in the cooling line temperature on the magnitude and phase of S_{11} (i.e., the reflected power) as well as the magnitude of S_{21} (i.e., the transmitted power) is evident through Figs. 4-5(a). On the other hand, the effect of variation in the phase of S_{21} is more dominant in the performance of the overall system. Special attention to the variation in the phase of S_{21} should be paid. Based on the observations in this study, it is concluded that besides the necessity of stable temperature, keeping the pressure level of the cooling line constant has turned out to be important for phase stabilization in the output power for a 260 MHz RF-PA.

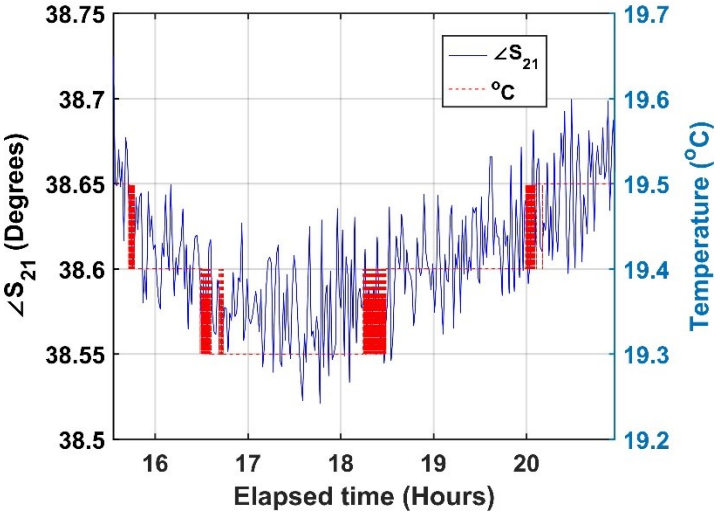


FIGURE 6. Phase variation of S_{21} ($\angle S_{21}$) (solid blue) and measured water cooling line temperature (dashed red) on the right-hand side axis, with a closer look for the situation in Fig. 5(b) within the time interval in between hour 15.5 and hour 21.

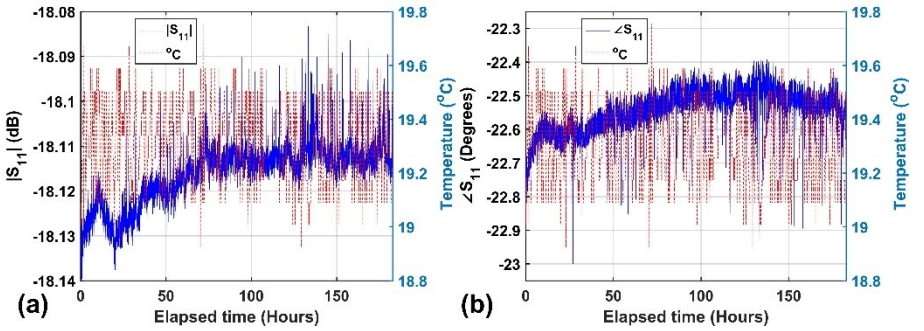


FIGURE 7. Measurement results of S_{11} a) magnitude ($|S_{11}|$) (solid blue) b) phase ($\angle S_{11}$) (solid blue) together with water cooling line temperature (dashed red) on the right-hand side axis as a function of elapsed time in the experiment.

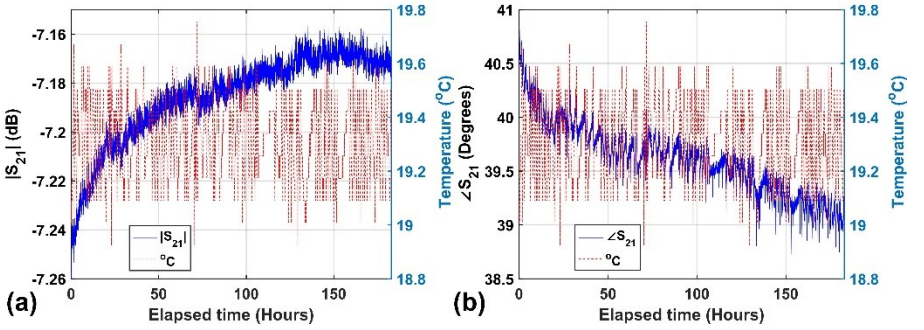


FIGURE 8. Measurement results of S_{21} a) magnitude ($|S_{21}|$) (solid blue) b) phase ($\angle S_{21}$) (solid blue) together with water cooling line temperature (dashed red) on the right-hand side axis as a function of elapsed time in the experiment.

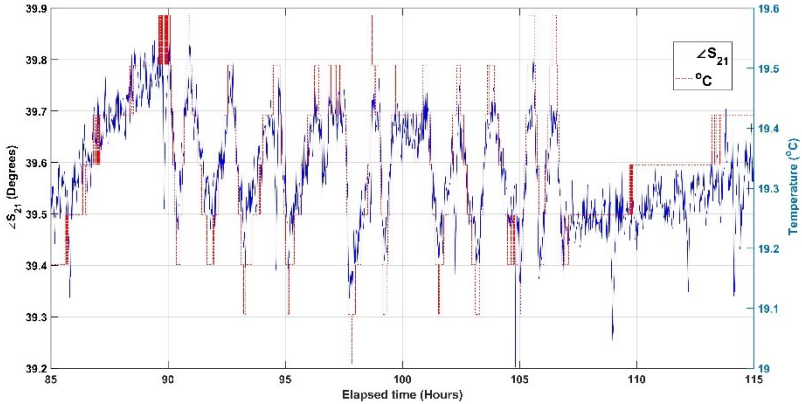


FIGURE 9. A closer view of S_{21} phase ($\angle S_{21}$) measurement borrowed from Fig. 8(b) in between hours 85 and 115: phase (solid blue) together with water cooling line temperature (dashed red) on the right-hand side axis as a function of elapsed time in the experiment.

The beam energy spread depends on RF phase and amplitude variations. Cooling temperature changes the beam energy indirectly. The long term beam energy drift and phase drift should be reduced by using accurate cooling control system. Furthermore, RF distributions for the injector system should be designed for maximal phase stability relative to the bunches in the linear accelerator [21].

In order to obtain the desired amplitude and phase stability in the accelerator structure in the presence of large perturbations such as beam loading or frequency shifts, it is necessary to use feedback control [22]. For this reason, the next step in the commissioning of the amplifier is the realization of the long-term operation of the complete system with the phase shifter and SHB, including the RF controller.

These tests will be performed in two stages: the first step will include the manual tuning of the phase shifter and in the second step, the long-term operation process will be carried out. The utilization of phase shifter will enable us to decrease the reflected power to the order of mWs, which is measured as ~5Ws in this study. After completion of the tests for the overall integrated system including RF-PA, SHB and RF controller, we aim to achieve the phase stability value of 0.35 deg/ºC as given in Table 1. This target seems feasible since the phase variation results which are the measured as 0.5 deg/ºC and 0.67 deg/ºC, are low enough to be corrected by RF controller which will be incorporated in near future.

Acknowledgments. This work is supported by Presidency Strategy and Budget Department under Grant No: 2006K-120470. The authors would like to thank to the TARLA team for their devoted work.

References

[1] Aksoy, A., Karslı, O. and Yavas, O., The Turkish accelerator complex IR FEL project, *Infrared Phys. Technol.,* 51/5 (2008) 378-81.

[2] Aksoy, A. and Karsli, O. (Eds.), The technical design report of Turkish Accelerator and Radiation Laboratory in Ankara, Technical Report (Ankara University, 2015).

[3] Aksoy, A., Karsli, O., Aydin, A., Kaya, C., Ketenoglu, B., Ketenoglu, D. and Yavas, O., Current status of Turkish Accelerator and Radiation Laboratory in Ankara: the TARLA facility, *Can. J. Phys.,* 96/7 (2018) 837-42.

[4] Karsli, O., Aksoy, A., Kaya, C., Koc, B., Dogan, M., Elcim, O.F. and Bozdogan, M., High power RF operations studies at TARLA facility, *Can. J. Phys.,* Accepted: https://doi.org/10.1139/cjp-2018-0778.

[5] Wangler, T.P., RF Linear accelerators, John Wiley & Sons, 2008.

[6] Karsli, O., Yavas, O. and Dogan, M., Design of L Band 20 kW High Power Solid State Amplifier for TARLA/TAC Project, *Известия высших учебных заведений. Физика.*, 55/10-3 (2012) 154-9.

[7] Karsli, O. and Yavas, O., A design study on high power RF system for the TARLA facility of TAC, *Nucl. Instrum. Methods Phys. Res. A*, 693 (2012) 215-9.

[8] Wang, F., Liu, K., Feng, L., Lin, L., Zhang, B., Hao, J. and Quan, S., Using a 1.3 GHz 20 kW Solid State Amplifier as RF Power Supply for DC-SRF Photo-injector, *6th Workshop on ERL,* New York, USA 2015.

[9] Dillon, S., Schach, C. and Nobel, B., Design of a high speed pulsed 324 MHz solid-state amplifier for use in a beam chopper, *IPAC2012*, Louisiana, USA, (2012) 2242-4.

[10] Frenzel, L.E., RF power for industrial applications, Pearson Prentice Hall 2004.

[11] SigmaPhi Accelerator Technologies, https://www.sigmaphi.fr/en/produits/amplificateurs-radiofrequence, Accessed: 2019-03-20.

[12] National Instruments, http://www.ni.com/en-tr/shop/labview/labview-details.html, Accessed: 22.03.2019.

[13] Wurlich, A., CERN Accelerator School, CERN-94-01, 1994.
[14] Wiedemann, H., Particle accelerator physics, (Vol. 314), Springer, Berlin, 2007.
[15] Chao, A.W., Handbook of Accelerator Physics and Engineering, 2nd Printing, World Scientific Publishing Co. Pte. Ltd., pp.256, 1998.
[16] Pozar, D. M., Microwave engineering, 4th Edition, John Wiley & Sons, 174 Chapter 4, 2011.
[17] MathWorks, https://www.mathworks.com/products/matlab.html, Accessed: 22.03. 2019.
[18] Akre, R., Temperature Stability of RF Components, (2005). http://slac. stanford.edu/grp/lcls/controls/global/subsystems/llrf/26sep2005Review/Temperature%20Stability%20of%20RF%20Components%20Apr%2005.pdf Accessed: 13.04.2019.
[19] Akre, R., Emma, P. and Krejcik, P., Measurements on SLAC LINAC RF system for LCLS Operation *PACS2001, Proceedings of the 2001 Particle Accelerator Conference (Cat. No. 01CH37268)*, 2 (2001) 1453-1455.
[20] Decker, F.J., Akre, R., Byrne, M., Farkas, Z.D., Jarvis, H., Jobe, K., Koontz, R., Mitchell, M., Pennacchi, R., Ross, M. and Smith, H., Effects of temperature variation on the SLC LINAC RF system, *IEEE Proceedings of Particle Accelerator Conference*, 3 (1995) 1821-1823.
[21] MAX IV Detailed Design Report, https://www.maxiv.lu.se/accelerators-beamlines/accelerators/accelerator-documentation/max-iv-ddr Accessed: 13.04. 2019.
[22] Suelzle, L.R., RF amplitude and phase stabilization for a superconducting linear accelerator by feedback stabilization techniques, High Energy Physics Lab, (1968), https://www.bnl.gov/magnets/staff/gupta/Summer1968/0067.pdf Accessed: 22.03. 2019.

Current Address: Ozlem Karsli: Institute of Accelerator Technologies, Ankara University 06830, Gölbaşı, Ankara, TURKEY
E-mail: okarsli@ankara.edu.tr,
ORCID: https://orcid.org/0000-0002-1466-4989

Current Address: Evrim Colak (corresponding author): Institute of Accelerator Technologies, Ankara University, 06830, Gölbaşı, Ankara, TURKEY
Electrical and Electronics Engineering Department, Ankara University, TURKEY
E-mail: evrim.colak@ankara.edu.tr
ORCID: https://orcid.org/0000-0002-4961-5060

INSTRUCTIONS TO CONTRIBUTORS

Communications Faculty of Sciences University of Ankara Series A2-A3: Physical Sciences and Engineering is a peer reviewed journal which has been published since 1948 by Ankara University, accepts original research articles written in English in the fields of Physics, Engineering Physics, Electronics/Computer Engineering, Astronomy and Geophysics. Review articles written by eminent scientists can also be invited by the Editor.
Manuscripts should be submitted as a single PDF file attached to an e-mail with a covering letter. In the covering letter, authors should nominate three potential reviewers and e-mailed the file to the most appropriate Area Editor of the research. The editorial office may not use these nominations, but this may help to speed up the selection of appropriate reviewers.
Manuscripts should be typeset using the LATEX typesetting system. Authors should prepare the article using the COMMUNICATIONS style before submission by e-mail. Manuscripts written in DOC form are also acceptable. A template of manuscript can be reviewed in **https://dergipark.org.tr/communc.** After the acceptance of manuscripts for publication, we will ask you to submit the **TeX** form of the manuscript prepared in accordance with the style of the Journal. Authors are required to submit their Open Researcher and Contributor ID (**ORCID**) 's which can be obtained from http://orcid.org as their URL address in the format http://orcid.org/xxxx-xxxx-xxxx-xxxx. Acknowledgements should be given as short as possible at the end of the text. Formulas should be numbered consecutively in parentheses ( ). Footnotes should be avoided if possible, but when necessary, should be short and never contain any important part of the work and should be numbered consecutively by superscripts. All illustrations not including tables (photographs and other films, drawings, graphs, etc) must be labeled as "Figure". The proper position of each table and figure must be clearly indicated in the paper.

All tables and figures must have a number (Table 1, Figure 1) and a caption or legend. References including comments must be numbered consecutively in order of first appearance in the text. The reference number should be put in brackets [] where referred to in the text. References should be listed at the end of the manuscript in the numbered order in which they appear in the text as follows:

[1] Bairamov, E, Ozalp N., Uniform convergence and numerical computation of the Hubbell radiation rectangular source integral, *Radiation Physics and Chemistry*, 80 (2011) 1312–1315.
[2] Kelley, J. L., General Topology, Van Nostrand, 1970.

It is a fundamental condition that articles submitted to COMMUNICATIONS have not been previously published and will not be simultaneously submitted or published elsewhere. After the manuscript has been accepted for publication, the author will not be permitted to make any new additions to the manuscript.

Before publication, the galley proof is sent to the author for correction. Thus, it is solely the author's responsibility for any typographical mistakes occur in their article as it appears in the Journal. The contents of the manuscript published in the COMMUNICATIONS are the sole responsibility of the authors.

The PDF copies of accepted papers are free of charges, but hard copies of the paper, if required, are due to be charged for the amount of which is determined by the administration each year.

Editor in Chief
Commun. Fac. Sci. Univ. Ank. Ser. A2-A3.
Ankara University, Faculty of Sciences
06100, Besevler - ANKARA TURKEY

C O M M U N I C A T I O N S

FACULTY OF SCIENCES
UNIVERSITY OF ANKARA

DE LA FACULTE DES SCIENCES
DE L'UNIVERSITE D'ANKARA

Series A2-A3: Physical Sciences and Engineering

Volume 61 **Number : 2** **Year :2019**