

# Tiny-VBF: Resource-Efficient Vision Transformer based Lightweight Beamformer for Ultrasound Single-Angle Plane Wave Imaging

Abdul Rahoof<sup>1</sup>, Vivek Chaturvedi<sup>1</sup>, Mahesh Raveendranatha Panicker<sup>1</sup>, Muhammad Shafique<sup>2</sup>

<sup>1</sup>Indian Institute of Technology Palakkad, India; <sup>2</sup>New York University, Abu Dhabi, UAE

**Abstract**—Accelerating compute intensive non-real-time beamforming algorithms in ultrasound imaging using deep learning architectures has been gaining momentum in the recent past. Nonetheless, the complexity of the state-of-the-art deep learning techniques poses challenges for deployment on resource-constrained edge devices. In this work, we propose a novel vision transformer based tiny beamformer (Tiny-VBF), which works on the raw radio-frequency channel data acquired through single-angle plane wave insonification. The output of our Tiny-VBF provides fast envelope detection requiring very low frame rate, i.e. 0.34 GOPs/Frame for a frame size of 368 x 128 in comparison to the state-of-the-art deep learning models. It also exhibited an 8% increase in contrast and gains of 5% and 33% in axial and lateral resolution respectively when compared to Tiny-CNN on in-vitro dataset. Additionally, our model showed a 4.2% increase in contrast and gains of 4% and 20% in axial and lateral resolution respectively when compared against conventional Delay-and-Sum (DAS) beamformer. We further propose an accelerator architecture and implement our Tiny-VBF model on a Zynq UltraScale+ MPSoC ZCU104 FPGA using a hybrid quantization scheme with 50% less resource consumption compared to the floating-point implementation, while preserving the image quality.

**Index Terms**—Vision Transformer, Beamforming, FPGA, Image Reconstruction, Ultrasound Imaging

## I. INTRODUCTION

Ultrasound (US) imaging is an important diagnostic method which is used extensively due to its cost-effectiveness, convenience and absence of ionizing radiation. Conventional ultrasound imaging employs a focused transmit scheme that systematically reconstructs the region of interest line by line. However, this method's frame rate is restricted, making it unsuitable for scenarios involving significant motion in the area being scanned, as seen in echocardiography, elastography, and doppler applications [1]. On the other hand, plane wave (unfocused) transmit schemes are popular for their higher frame rates, typically exceeding a few 1000 Frames/sec [1], but faces a significant challenge in achieving satisfactory image quality, particularly resolution and contrast, due to its complex beamforming process during image reconstruction [1].

Ultrasound imaging systems commonly adopt the Delay-and-Sum (DAS) [2] beamforming technique due to its lower computational demands. However, when combined with single plane wave imaging, DAS leads to reduced resolution and contrast due to the data-independent apodization process. Coherent Plane-Wave Compounding offers an adept solution to mitigate the reduction in image quality [3]. However, this

solution comes with the drawback of decreased frame rate, resulting in a trade-off between image quality and frame rate. A significant advancement in this realm is the Minimum Variance Distortionless Response (MVDR) [4]. However, the computational complexity of MVDR, particularly the matrix inversions poses challenges for real-time implementation having a complexity of  $O(n^3)$ , where  $n$  is the number of channels. For instance, it demands approximately 98.78 GOPs/Frame for a frame size of 368 x 128 [5].

Recently, several works have shown significant performance improvement in image reconstruction by using deep learning models for beamforming process [6]–[10]. The integration of deep learning models enhances the quality of the reconstructed images with single angle plain wave insonification. However, the complexity of these models such as [8] (50 GOPs/Frame) and [9] (199 GOPs/Frame) makes them non-trivial to implement on resource-constrained devices.

Evidently, compact ultrasound imaging system holds the potential to enhance diagnostic confidence, improve patient care and offer the advantages of portability and accessibility in medical settings. These compact systems or edge devices have stringent resource constraints and achieving high performance on such systems is non-trivial. Employing Field Programmable Gate Arrays (FPGAs) in beamforming emerges as a compelling choice due to their flexibility and parallel processing capabilities, making them a more efficient solution compared to Graphics Processing Units (GPUs) and processors like Media Processors (MP) and Digital Signal Processors (DSP) for enhancing ultrasound imaging systems, particularly in portable and resource-constrained environments [11].

Towards the above mentioned objective, we propose a vision transformer (ViT) [12] based ultrasound image reconstruction beamformer Tiny-VBF which is accelerated on a FPGA based hardware evaluation board. Recently, ViTs have gained immense popularity, revolutionizing the accuracy of pivotal computer vision tasks. The calculation pattern employed in ViTs differs from the convolutional units used in convolutional neural networks (CNNs) for extracting partial information. ViTs prioritize the comparison of pixel channels rather than focusing solely on feature values within the receptive field. It employs stacked self-attention and point-wise fully connected layers in both the encoder and decoder. The self-attention layers in ViTs are global in nature and establish comprehensive dependencies between input and output. Currently, ViT models

are rapidly surpassing CNN models in various computer vision tasks [13]. Next, we discuss the major contributions of our work.

### Our Novel Contributions:

- We propose a ViT-based image reconstruction model, Tiny-VBF, which improves the quality of ultrasound images, achieves more accurate visualization of structures and abnormalities compared to conventional DAS beamformer. Moreover, Tiny-VBF can effectively detect edges compared to DAS and Tiny-CNN [7] beamformers, as illustrated in the reconstructed images presented in Fig. 1(a).
- Our Tiny-VBF is computationally light and uses fewer FLOPs (0.34 GOPs/Frame) compared to the state-of-the-art tinyML models Tiny-CNN (11.7 GOPs/Frame) and FCNN [6] (1.4GOPs/Frame).
- We performed an extensive quality evaluation in terms of contrast and resolution on two different datasets, i.e. in-vitro dataset and in-silico dataset.
- For in-vitro data, the Tiny-VBF showed an 8% increase in contrast, while also achieving gains of 5% and 33% in axial and lateral resolution compared to the Tiny-CNN, and a 4.2% increase in contrast with gains of 4% and 20% in axial and lateral resolution compared to the DAS.
- Additionally, for in-silico data, the Tiny-VBF showed a 10.7% increase in contrast, while also achieving gains of 17.7% and 25% in axial and lateral resolution compared to Tiny-CNN, and an 8% increase in contrast with gains of 16.8% and 25% in axial and lateral resolution compared to DAS.
- We implemented our Tiny-VBF on the ZCU104 FPGA board (with clock frequency of 100 MHz) and conducted a comprehensive assessment of resource utilization and image quality across quantization, including our hybrid quantization. The hybrid quantization strategy led to a remarkable 50% reduction in resource consumption compared to the floating-point implementation, while maintaining image quality, as illustrated in Fig. 1(b).

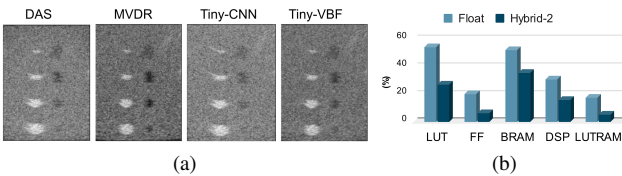


Fig. 1: (a) The B-mode image of cyst evaluation experiment data; and (b) Comparison of resource consumption of Float and Hybrid-quantized Tiny-VBF model on FPGA

To the best of our knowledge, we are the first to evaluate the applicability of beamforming through a vision transformer network and deploy a deep learning-based beamformer on resource-constrained edge devices. In Fig. 2, we present an overview of our innovative approach.

**Paper Organization:** Section II provides an overview of related works, followed by Section III which outlines the proposed methodology. Section IV details the experimental results, and lastly, Section V presents the conclusion.

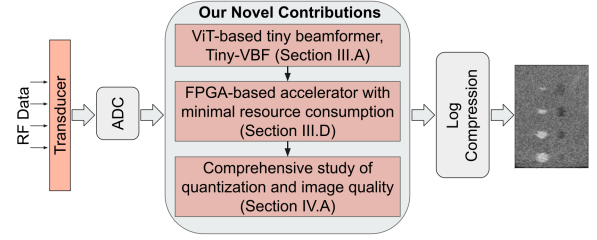


Fig. 2: Overview of our novel contributions

## II. RELATED WORKS

There are several machine and deep learning approaches to improve various signal processing tasks in ultrasound imaging with emphasis on beamforming. The work [14] proposed a preprocessing technique for the detection and elimination of reflection artifacts within photoacoustic channel data. Similarly, the work in [15] proposed a preprocessing technique for suppressing offaxis scattering in the received channel data. In the works of [16] and [17], proposed a deep learning-based denoising method for ultrasound single angle plane wave imaging.

The work described in [6] designed a fully connected neural network (FCNN) to perform beamforming on a pixel-by-pixel basis. This work improved the resolution and contrast compared to DAS and showed faster compared to MVDR beamformer. However, FCNN only captures the local features in the input data to estimate the apodization weights. The architecture of the Tiny-CNN beamformer [7] receives data of a region with dimensions  $(x, y, c_h)$ , where  $x, y$ , and  $c_h$  represent the axial and lateral coordinates and the number of transducers (channels), respectively. The product of the output layer (the apodization weights) and the input layer (the ToFC data) is summed along the channel axis to produce the beamformed image data. The resulting image is then processed with the Hilbert Transform to obtain the final B-mode image. Compared to DAS and FCNN, this work improved the resolution and contrast in fewer angles of transmission, which enhanced reconstruction rate. However, the quality of images produced through single-angle plane-wave (PW) imaging is not satisfactory, closely resembling the performance achieved by the DAS method. The work by [8] introduced an extensive U-Net model with wavelet decomposition and reconstruction. Notably, this approach exhibited enhanced resolution and contrast in single-angle images when tested on simulation data, showing improvements over the DAS method. Specifically, the axial resolution experienced a remarkable 37.5% enhancement, while the lateral resolution improved by 65.9% compared to DAS. Moreover, the Contrast-to-Noise Ratio (CNR) and Generalized Contrast-to-Noise Ratio (GCNR) were notably elevated by 160% and 33.8%, respectively, in comparison to DAS. However, challenges emerged when applying the model to in-vivo data, as non-linear attenuation with depth led to a loss of contrast and vital medium-related information. Additionally, the method exhibited high computational complexity, exceeding 50 GOPs/Frame for a frame size of  $368 \times 128$ . Lastly, the approach is non-trivial to implement on resource-constrained edge devices. Similarly, [9] introduced

an innovative approach involving a fusion of the GoogLeNet and U-Net models for beamforming purposes. This method achieved notable improvements in resolution and contrast for both in-vitro and in-vivo data when contrasted with the DAS method. Moreover, the method showed substantial reduction in sidelobes and artifacts, effectively outperforming DAS. Point-reflector in-vitro data revealed a 23.6% enhancement in both axial and lateral resolutions compared to DAS. The Contrast-to-Noise Ratio (CNR) for anechoic in-vitro data achieved 158% enhancement when compared to DAS. However, it is essential to note the high computational complexity of this method, reaching approximately 199 GOPs/Frame for a frame size of 384 x 256. To the best of our knowledge, there have been limited attempts to implement the beamforming accelerators using neural networks on edge devices. In this paper, we have tried to attempt the same using vision transformers.

### III. PROPOSED METHODOLOGY

In this section, we describe the implementation of the proposed Tiny-VBF model and its deployment on FPGA. Fig. 3 illustrates an overview of the proposed methodology. Our Tiny-VBF Model is crafted through an iterative optimization process, adjusting the encoder-decoder layers to enhance both throughput and resulting image quality. Following this optimization, the model is quantized and an extensive comparison of reconstructed image quality is conducted. Subsequently, an FPGA-based accelerator is designed for deploying the quantized Tiny-VBF model.

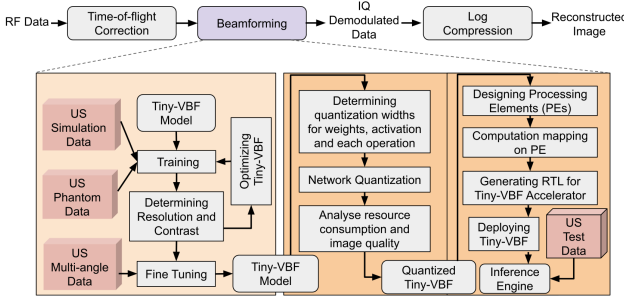


Fig. 3: Overview of the proposed methodology

#### A. Tiny-VBF architecture

The architecture of our Tiny-VBF is shown in Fig. 4. The single-angle plain wave data acquired from the ultrasound imaging system undergoes initial time-of-flight correction before being fed into the Tiny-VBF model. The output of the model is an IQ demodulated beamformed image. The ground truth for learning is the MVDR beamformed IQ demodulated data. The data points were normalized to the interval of  $[-1, 1]$ . The architecture includes an encoder block including two transformer blocks and a decoder block. The encoder block is designed such as to obtain the most important features. It contains dense layers to map the channel data to lower dimensions. The transformer block includes a normalization layer, a Multi-Head Attention Layer (MHAL), two skip connectors and two dense layers. The decoder block contains dense layers to

reconstruct the IQ demodulated beamformed image from the encoded features.

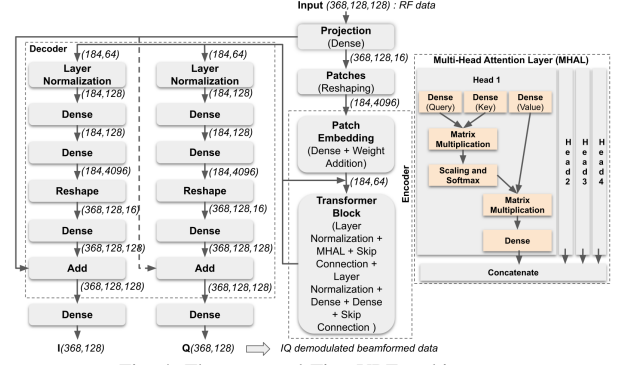


Fig. 4: The proposed Tiny-VBF architecture

The proposed Tiny-VBF model is implemented using TensorFlow platform (version 2.40) in Python (version 3.8) and was trained and tested on a system AMD® Ryzen 7 4700g with radeon graphics x 16 with Ubuntu 20.04. The model has a total of 15,07,922 weights. The model reduced the artifacts in the reconstructed image and improved the resolution and contrast of the image compared to standard DAS beamformer.

#### B. Dataset

The training data was collected from the Verasonics research ultrasound machine, specifically the Verasonics Vantage 128 and L11-5v transducer models (Verasonics Inc., Redmond, WA) [19], [20]. The ultrasound machine operates at a center frequency of 7.6 MHz and a sampling frequency of 31.25 MHz. The details of the acquisition are described in [7], [19], [20]. The data collection process followed the guidelines outlined in the Helsinki Declaration of 1975, which were revised in 2000. Then, the model was fine tuned using the multi-angled images obtained from the CUBDL dataset with 10 angles of transmissions [21].

The evaluation of performance is carried out on in-silico and in-vitro datasets provided by the plane-wave imaging challenge for medical ultrasound (PICMUS) of the 2016 IEEE International Ultrasonics Symposium [18]. The in-silico data for PICMUS is generated using Field II, while the in-silico dataset is acquired using a Verasonics Vantage 256 research ultrasound scanner (Verasonics Inc., Redmond, WA).

#### C. Parameter Setting

The proposed technique involves several hyperparameters that are crucial for its implementation. These include the batch size, optimizer, learning rate and loss function. The batch size was set to 10, and the optimization was performed using the Adam optimizer. The learning rate was not constant but was subjected to a polynomial decay schedule with cyclic changes, which allowed for more flexible parameter updates. Specifically, the initial and final learning rates were set to  $10^{-4}$  and  $10^{-6}$ , respectively. During each learning stage, the number of epochs was set to 1000. The loss function for this technique is Mean Squared Error (MSE) applied to the IQ demodulated beamformed image prior to log compression.

#### D. Tiny-VBF Accelerator

The accelerator architecture of the Encoder block is shown in Fig. 5. The architecture involves following non-linear operations ReLU, softmax, division and sqrt. It has 4 Processing Elements (PEs), each of which operates 16 element-wise multiplications followed by an adder tree. The input, weights and intermediate results of the operations are stored in on-chip BRAM memory.

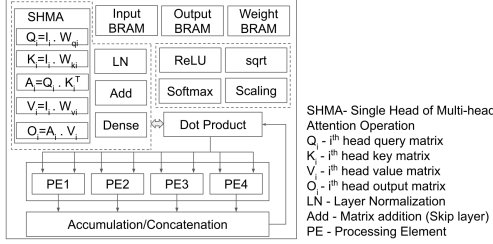


Fig. 5: The proposed Tiny-VBF accelerator architecture

The operations involved in computing Query, Key and Value of  $i^{th}$  head is shown in Fig. 6. The input  $I_i$  will be multiplied by Weight  $W_i$  to get the resultant matrix with size  $(np, k)$  where,  $np$  is the number of patches and  $k$  is the projection dimension divided by the number of heads. The operation will be performed on the processing elements  $PE1, PE2, PE3$  and  $PE4$ . Here, the  $inp1, inp2, \dots, inp8$  are loaded with the values of the row and column of input and weight, and perform dot products using the PEs. The output values will be accumulated and stored to the output memory.

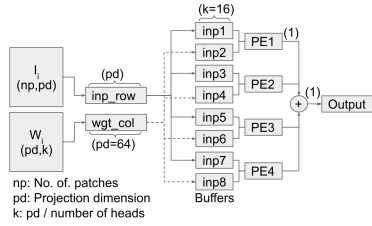


Fig. 6: Operation involved in computing Query, Key and Value of MHAL

Similarly, the attention score for each head will be computed as shown in Fig. 7. The Query and Key matrices have dimension  $(np, k)$  and the resulting matrix has a dimension of  $(np, np)$ . Moreover, the dense operation and the output of the single head operation, that is the result of multiplying attention score and Value matrices, will be computed by using the architecture shown in Fig. 8(a). The partial result needs to be accumulated with the next corresponding row and column operation, resulting in the final output, which is then stored in the output memory. Finally, Fig. 8(b) represents the

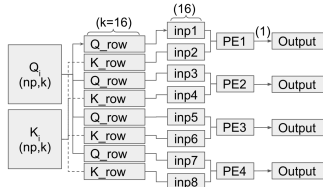


Fig. 7: Architecture for computing the attention score of single head operation architecture of the PE. It involves 16 element multiplications followed by an adder tree.

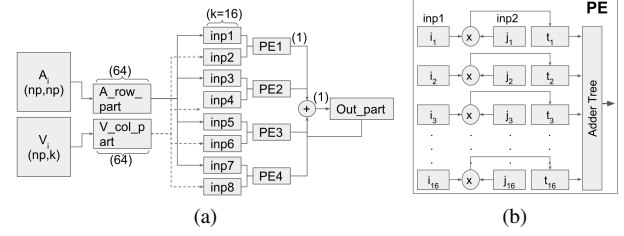


Fig. 8: (a) Architecture for computing the output of single head operation or dense layer operation and (b) Architecture of single Processing Element

#### IV. EXPERIMENTAL RESULTS

In this section, a comprehensive examination of the proposed Tiny-VBF is presented, using diverse in-silico and in-vitro datasets that include point targets and cyst targets.

Fig. 9(a) demonstrates the results of each ultrasound beamforming method over single angle PICMUS contrast simulation data. The performance of the state-of-the-art Tiny-CNN model on this data is restricted to that of the DAS. There remains noise and other artifacts in the region of cysts in case of DAS and Tiny-CNN, but it is reduced greatly in case of Tiny-VBF and our benchmark MVDR. From the Fig. 9(b), it is evident that the Tiny-VBF and target MVDR exhibit sharp changes in pixel intensity along the lateral position on the image boundary of cysts, as opposed to the standard DAS and Tiny-CNN.

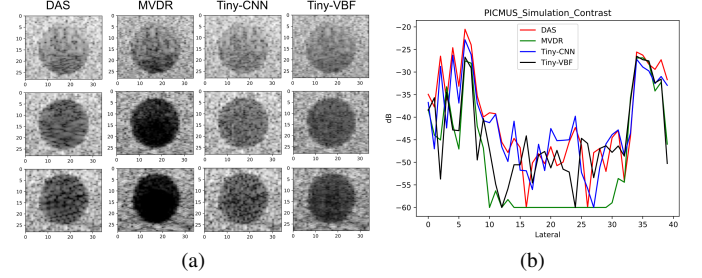


Fig. 9: (a) The in-silico contrast images at three different depths, ranging from near field to far field (at depths of 13 mm, 25 mm, and 37 mm) and (b) Lateral variations in the beamformed cyst from the in-silico contrast data at depth of 37 mm

Similarly, Fig. 10 shows the performance of each beamforming method over in-vitro contrast data. The results were obtained from near and far fields (at depths of 15 mm and 35 mm) from the ultrasound probes. It also demonstrates a sharp edge over the cyst region when compared to the DAS and Tiny-CNN.

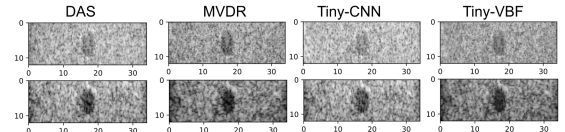


Fig. 10: The in-vitro contrast images at depths of 15 mm and 35 mm

Table I provides a quantitative comparison of the contrast for both simulation (in-silico) and phantom (in-vitro) data. The Contrast Ratio (CR) is higher with Tiny-VBF compared to both DAS and Tiny-CNN, exhibiting improvements of 8% and 10.7% in the in-silico data, and 4.2% and 8% in the in-vitro data, respectively. There is a slight reduction in CNR and



GCNR compared to DAS in the simulation data, but it showed better in the phantom data.

TABLE I: Contrast metrics (mean) of Simulation and Phantom data

Beamformer	Simulation (dB)			Phantom (dB)		
	CR	CNR	GCNR	CR	CNR	GCNR
DAS	13.78	<b>2.37</b>	<b>0.83</b>	11.70	1.04	<b>0.83</b>
MVDR	<b>21.66</b>	1.95	0.78	<b>15.09</b>	<b>2.63</b>	0.72
Tiny-CNN	13.45	2.04	0.83	<b>11.30</b>	1.05	0.79
Tiny-VBF	<b>14.89</b>	1.75	0.74	<b>12.20</b>	<b>1.39</b>	0.67

Fig. 11 illustrates the results of different beamformers over resolution-distortion simulation data. The point targets are arranged horizontally against an anechoic background in two distinct depth zones. In comparison to Tiny-CNN and DAS, both the Tiny-VBF model and MVDR exhibit superior lateral and axial resolution.

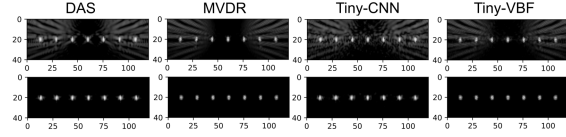


Fig. 11: B-mode images of in-silico resolution-distortion dataset at depths of 15 mm and 35 mm

Fig. 12 provides a visualization of the lateral resolution for point targets located at two different depths. It is evident that the MVDR and Tiny-VBF beamformers significantly reduce the mainlobe width and sidelobes in contrast to DAS and Tiny-CNN.

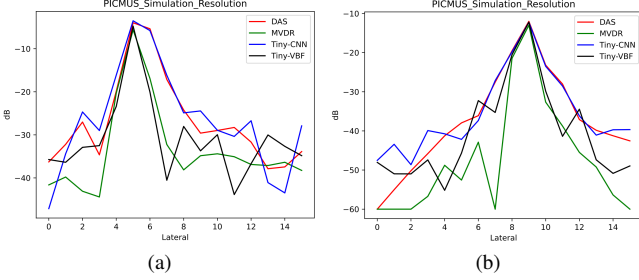


Fig. 12: Lateral point spread functions at (a) 15.12 mm and (b) 35.15 mm where x-axis represents the lateral point position (pixel position) and y-axis represents the normalized amplitude values

Similarly, the results over point targets in the experimental data is shown in Fig. 13. The performance of the Tiny-VBF model is consistently shown better than the DAS and Tiny-CNN in the experimental data.

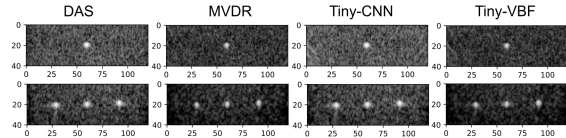


Fig. 13: B-mode images of in-vitro resolution-distortion dataset at depths of 14 mm and 33 mm

Additionally, graphs in Fig. 14 illustrate the lateral variation results from in-vitro data, demonstrating the heightened performance achieved with the Tiny-VBF model.

Moreover, Table II provides a quantitative comparison of axial and lateral resolution for both simulation and phantom data. It is evident that the proposed Tiny-VBF consistently surpasses the state-of-the-art DAS and Tiny-CNN in both

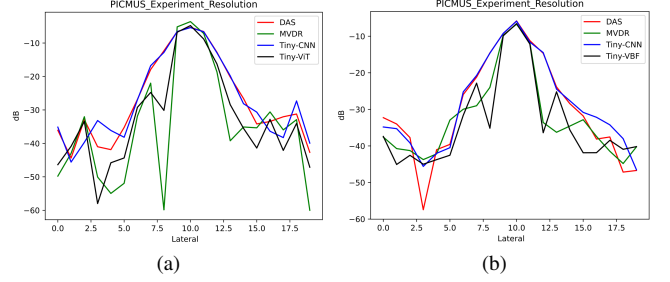


Fig. 14: Lateral point spread functions at (a) 14.01 mm and (b) 32.79 mm where x-axis represents the lateral point position (pixel position) and y-axis represents the normalized amplitude values

cases.

TABLE II: Quantitatively analysis of axial and lateral resolution of Simulation and Phantom data

Beamformer	Simulation (mm)		Phantom (mm)	
	Axial	Lateral	Axial	Lateral
DAS	0.364	0.6	0.459	0.6
MVDR	<b>0.297</b>	<b>0.45</b>	0.459	<b>0.48</b>
Tiny-CNN	0.368	0.6	0.466	0.72
Tiny-VBF	<b>0.303</b>	<b>0.45</b>	<b>0.444</b>	<b>0.48</b>

The inference time of Tiny-VBF model for each frame with frame size of  $368 \times 128$  is estimated around *0.230 sec* on *Intel Xeon 2vCPU@2.2GHz*. Whereas the inference time of Tiny-CNN, MVDR and CNN [8] are approximately *0.520 sec*, *240 sec* and *4 sec*, respectively.

#### A. Evaluation on FPGA

We quantized the Tiny-VBF model with different quantization levels and deployed it on FPGA. In addition, We used two hybrid quantization schemes as represented in Table III.

TABLE III: Hybrid quantization bit-width representation

	Hybrid-1	Hybrid-2
Weights	8 bits	8 bits
Softmax	24 bits	24 bits
Mul/Add ops	20 bits	16 bits
Intermediate output	20 bits	16 bits

The reconstructed images of simulation and phantom data, utilizing floating-point data and different quantization levels including hybrid quantizations, is shown in Fig. 15. We observed a significant degradation in image quality with the use of 16-bit quantization.

Table IV displays the resolution of both simulation and phantom data with different quantization levels. The quantitative values were similar to original values in case of quantization levels 24 bits, 20 bits and hybrid quantizations. Similarly, contrast metrics of both simulation and phantom data are shown in Table V. There is slight variation in the contrast values as compared to the floating point implementation.

TABLE IV: Resolution metrics of Simulation and Phantom data on FPGA

Quantization Bit-width	Simulation (mm)		Phantom (mm)	
	Axial	Lateral	Axial	Lateral
Float	<b>0.303</b>	<b>0.45</b>	0.444	<b>0.48</b>
24 bits	<b>0.303</b>	<b>0.45</b>	0.444	<b>0.48</b>
20 bits	0.310	<b>0.45</b>	<b>0.421</b>	0.54
Hybrid-1	0.309	<b>0.45</b>	0.429	0.54
Hybrid-2	0.309	<b>0.45</b>	0.429	0.54

Furthermore, Table VI presents the resource utilization of the Tiny-VBF model at various quantization bit-widths. The

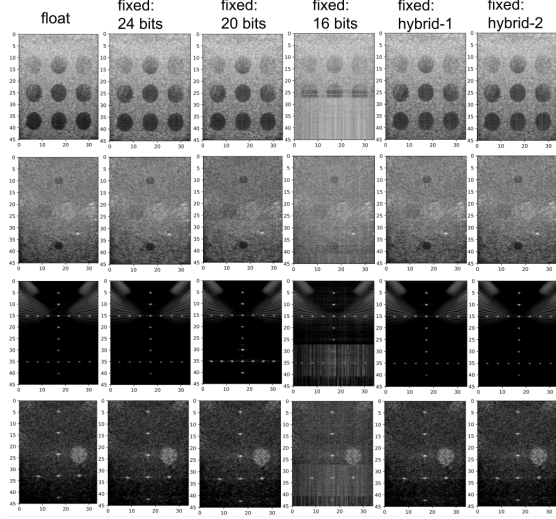


Fig. 15: B-mode images generated from FPGA

TABLE V: Contrast metrics of Simulation and Phantom data on FPGA

Quantization Bit-widths	Simulation (dB)			Phantom (dB)		
	CR	CNR	GCNR	CR	CNR	GCNR
Float	<b>14.89</b>	1.75	0.74	12.20	1.39	0.67
24 bits	14.07	<b>1.84</b>	<b>0.75</b>	13.0	1.22	<b>0.69</b>
20 bits	14.30	1.45	0.73	<b>13.05</b>	1.22	0.67
Hybrid-1	13.34	1.74	0.73	12.72	1.37	0.68
Hybrid-2	13.26	1.75	0.72	12.62	<b>1.40</b>	0.67

quantization of parameters results in a significant reduction in resource requirements. In the case of Hybrid-2 quantization, the resource consumption is reduced by more than 50% compared to its floating-point implementation.

TABLE VI: Resource utilization of Tiny-VBF model on FPGA

Resource	Float	24 bits	20 bits	16 bits	Hybrid-1	Hybrid-2
LUT	124935	88457	84594	59840	72415	<b>61951</b>
FF	91470	50454	43333	34920	38287	<b>29105</b>
BRAM	161.5	158	156	<b>82</b>	150	110
DSP	533	279	148	274	<b>146</b>	274
LUTRAM	17589	11556	9442	6795	5352	<b>5324</b>
Power (W)	4.489	4.369	4.174	<b>3.989</b>	4.229	4.174

## V. CONCLUSION

In this study, we introduced Tiny-VBF, a vision transformer-based image reconstruction for ultrasound imaging. The Tiny-VBF enhanced ultrasound beamforming quality when compared to the standard DAS method, in both in-silico and in-vitro data. The Tiny-VBF exhibit a minimal computational burden, requiring only 0.34 GOPs/Frame for a frame size of 368 x 128, as compared to the state-of-the-art deep learning models. We introduced an accelerator architecture and deployed our Tiny-VBF on the Zynq Ultrascale+ MPSoC ZCU104 FPGA. Our evaluation included network quantization and an assessment of the Tiny-VBF performance in terms of both quality and resource utilization. Notably, our hybrid quantization approach achieved a remarkable reduction of over 50% in resource consumption without compromising image quality. To the best of our knowledge, this represents the first implementation of a vision transformer-based beamformer and

the pioneering deployment of a deep learning-based beamformer on an FPGA.

## ACKNOWLEDGMENT

This work is supported in part by the NYUAD Center for Artificial Intelligence and Robotics, funded by Tamkeen under the NYUAD Research Institute Award CG010.

## REFERENCES

- [1] Mickael Tanter and Mathias Fink, "Ultrafast Imaging in Biomedical Ultrasound," in *IEEE Trans Ultrason Ferroelectr Freq Control*, 2014.
- [2] Thomas L. Szabo, "Diagnostic Ultrasound Imaging: Inside Out, 2014.
- [3] G. Montaldo, M. Tanter, J. Bercoff, N. Benech and M. Fink, "Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography," in *IEEE Trans Ultrason Ferroelectr Freq Control*, 2009.
- [4] Johan-fredrik Synnevag, Andreas Austeng and Sverre Holm, "Benefits of minimum-variance beamforming in medical ultrasound imaging," in *IEEE Trans Ultrason Ferroelectr Freq Control*, 2009.
- [5] Junying Chen, Jinhui Chen, Renxin Zhuang and Huaqing Min, "Multi-Operator Minimum Variance Adaptive Beamforming Algorithms Accelerated With GPU," in *IEEE Transactions on Medical Imaging*, 2020.
- [6] Ben Luijten et al., "Adaptive Ultrasound Beamforming Using Deep Learning," in *IEEE Transactions on Medical Imaging*, 2020.
- [7] Mathews, Roshan P and Mahesh Raveendranatha Panicker, "Towards Fast Region Adaptive Ultrasound Beamformer for Plane Wave Imaging Using Convolutional Neural Networks," in *International Conference of the IEEE Engineering in Medicine and Biology Society*, 2021.
- [8] Sobhan Goudarzi and Hassan Rivaz, "Deep reconstruction of high-quality ultrasound images from raw plane-wave data: A simulation and in vivo study," in *Ultrasonics*, 2022.
- [9] Lu JY, Lee PY, Huang CC, "Improving Image Quality for Single-Angle Plane Wave Ultrasound Imaging With Convolutional Neural Network Beamformer," in *IEEE Trans Ultrason Ferroelectr Freq Control*, 2022.
- [10] Sobhan Goudarzi, Amir Asif and Hassan Rivaz, "Ultrasound Beamforming using MobileNetV2," in *IEEE International Ultrasonics Symposium (IUS)*, 2020.
- [11] Jun Luo, Qijun Huang, Sheng Chang, Xiaoying Song and Hao Wang, "Compact Beamformer Design with High Frame Rate for Ultrasound Imaging," in *Sensors and Transducers*, 2014.
- [12] Alexey Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv:2010.11929*, 2020.
- [13] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *arXiv:2103.14030*, 2021.
- [14] Derek Allman, Austin Reiter and Muyinatu A. Lediju Bell, "A machine learning method to identify and remove reflection artifacts in photoacoustic channel data," in *IEEE Int. Ultrason. Symp.*, 2017.
- [15] Adam Luchies, Brett Byram, "Deep neural networks for ultrasound beamforming," in *IEEE Int. Ultrason. Symp.*, 2017.
- [16] Dimitris Perdios, Adrien Besson, Marcel Arditi, Jean-Philippe Thiran, "A Deep Learning Approach to Ultrasound Image Recovery," in *IEEE Int. Ultrason. Symp.*, 2017.
- [17] Zixia Zhou et al., "High Spatial-Temporal Resolution Reconstruction of Plane-Wave Ultrasound Images With a Multichannel Multiscale Convolutional Neural Network," in *IEEE Trans Ultrason Ferroelectr Freq Control*, 2018.
- [18] H. Liebgott, A. Rodriguez-Molares, F. Cervenansky, J.A. Jensen and O. Bernard, "Plane-wave imaging challenge in medical ultrasound," in *IEEE Int. Ultrason. Symp.*, 2016.
- [19] Malamal, Gayathri, and Mahesh Raveendranatha Panicker, "Towards Diffuse Beamforming For Specular Reflectors: A Pixel-Level Reflection Tuned Apodization Scheme For Ultrasound Imaging," in *IEEE International Symposium on Biomedical Imaging*, 2021.
- [20] Madhavanunni, A. N., and Mahesh Raveendranatha Panicker, "A nonlinear beamforming for enhanced spatiotemporal sensitivity in high frame rate ultrasound flow imaging," in *Comput. Biol. Med.*, 2022.
- [21] Dongwoon Hyun et al., "Deep learning for ultrasound image formation: CUBDL evaluation framework and open datasets," in *IEEE Trans Ultrason Ferroelectr Freq Control*, 2021.