US 20240265694A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2024/0265694 A1**

Lal et al. (43) **Pub. Date: Aug. 8, 2024**

(54) **ULTRASONIC COMPUTATION HARDWARE FOR CONVOLUTIONAL NEURAL NETWORK COMPUTING AND OTHER COMPUTATION APPLICATIONS**

(71) Applicant: **Cornell University**, Ithaca, NY (US)

(72) Inventors: **Amit Lal**, Ithaca, NY (US); **Yutong Liu**, Ithaca, NY (US); **Juneho Hwang**, Ithaca, NY (US); **Justin Kuo**, Ithaca, NY (US)

(21) Appl. No.: **18/282,141**

(22) PCT Filed: **Apr. 20, 2022**

(86) PCT No.: **PCT/US2022/025586**

§ 371 (c)(1),
(2) Date: **Sep. 14, 2023**

Related U.S. Application Data

(60) Provisional application No. 63/177,220, filed on Apr. 20, 2021.

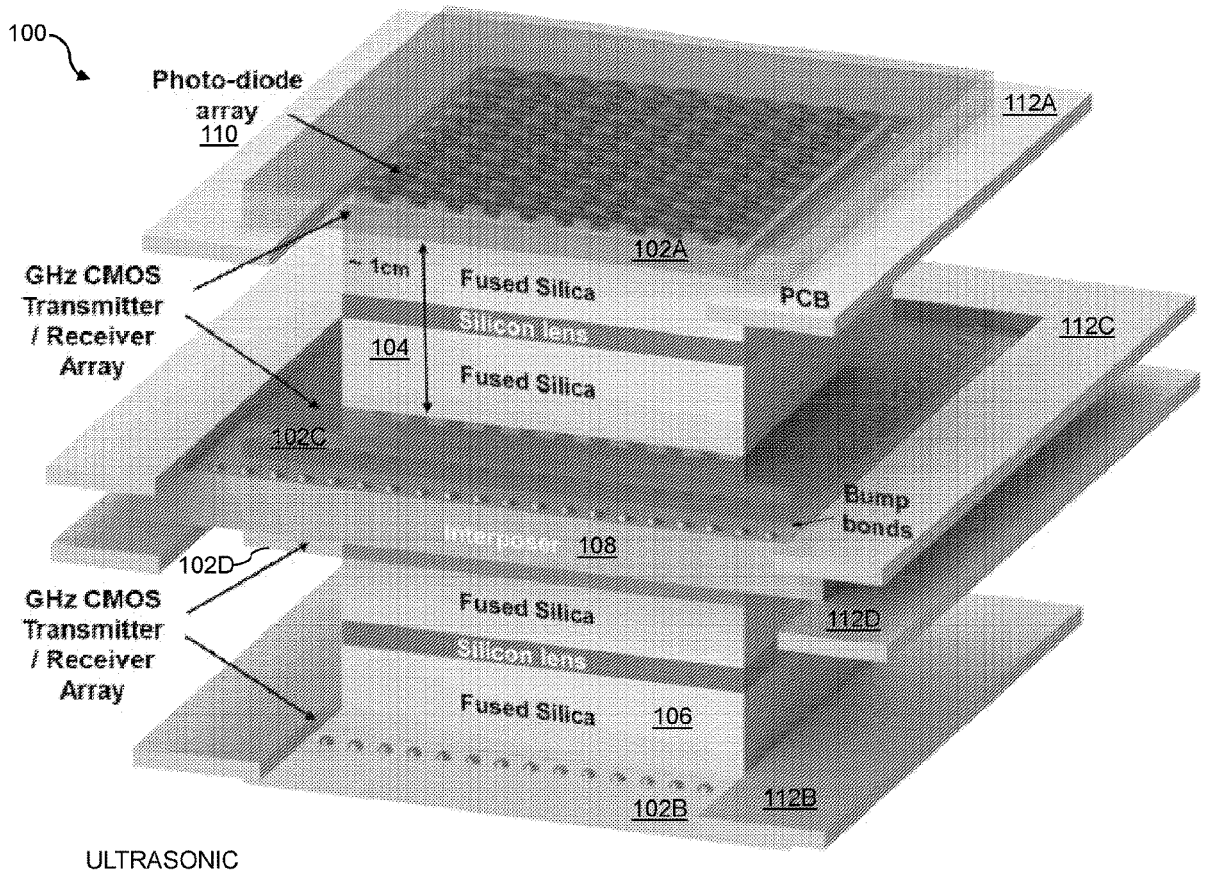
Publication Classification

(51) **Int. Cl.**
G06V 10/94 (2006.01)
G06V 10/82 (2006.01)

(52) **U.S. Cl.**
CPC ***G06V 10/955*** (2022.01); ***G06V 10/82*** (2022.01)

(57) **ABSTRACT**

An ultrasonic computation apparatus includes first and second ultrasonic transducer arrays arranged on opposing ends thereof, and further comprises first and second ultrasonic propagation regions arranged between the first and second ultrasonic transducer arrays and proximate respective ones of the first and second ultrasonic transducer arrays, and an intermediate computational structure arranged between the first and second ultrasonic propagation regions. Respective first and second input signals applied to respective ones of the first and second ultrasonic transducer arrays cause propagation of corresponding ultrasonic waves through the respective first and second ultrasonic propagation regions towards the intermediate computational structure. The intermediate computational structure is configured to receive the propagating ultrasonic waves from the respective first and second ultrasonic propagation regions and to generate from the received propagating ultrasonic waves an additional signal that is a function of the first and second signals.



ULTRASONIC COMPUTATION ENGINE

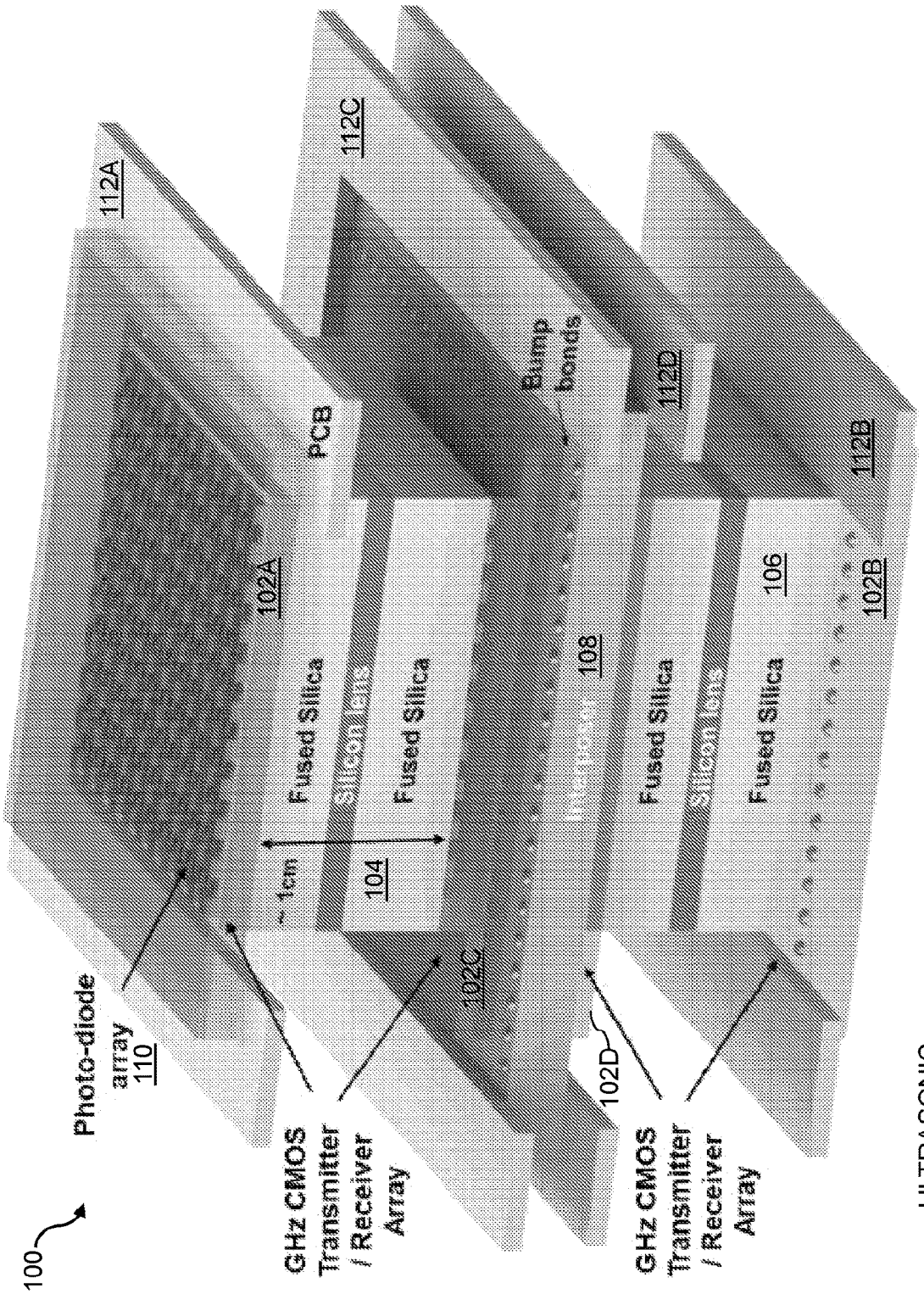

100
Photo-diode array 110
112A
112C
112D
112B
102A
102C
102D
102B
PCB
Bump bonds
108
~ 1cm
Fused Silica
104
Silicon Layer
106
GHz CMOS Transmitter / Receiver Array
GHz CMOS Transmitter / Receiver Array
ULTRASONIC COMPUTATION ENGINE

FIG. 1

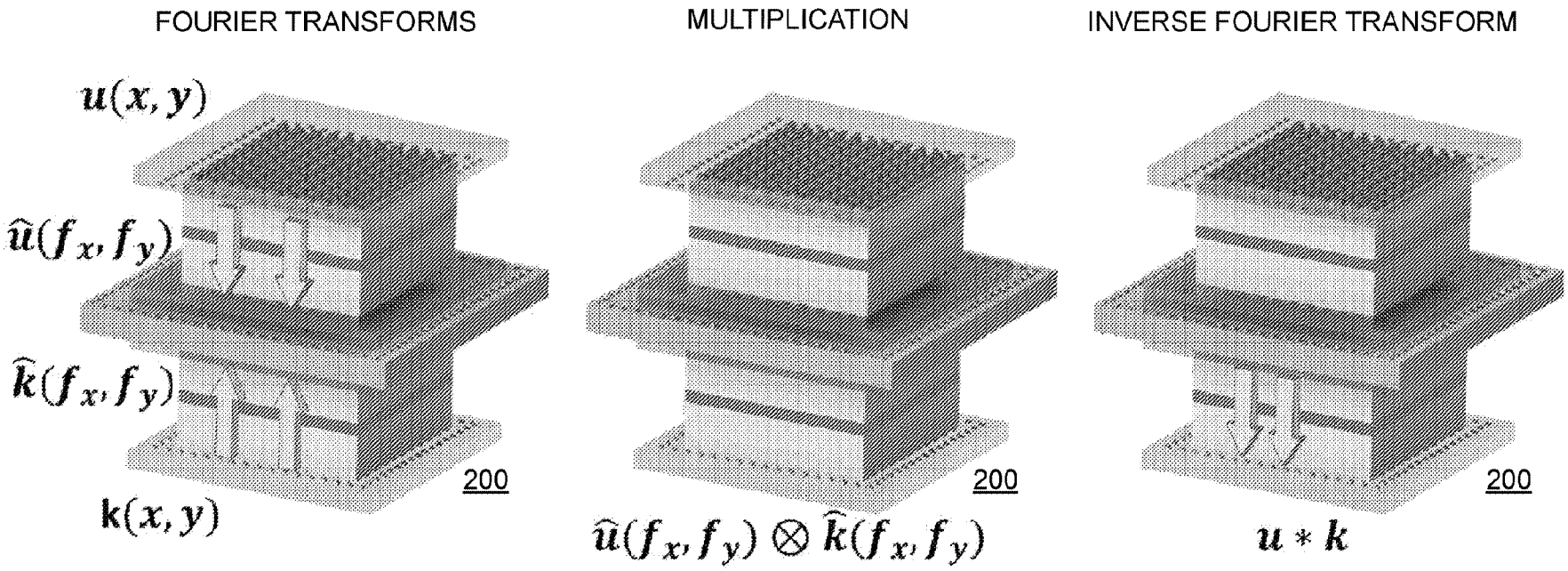

FOURIER TRANSFORMS
MULTIPLICATION
INVERSE FOURIER TRANSFORM
$u(x,y)$
$\hat{u}(f_x,f_y)$
$\hat{k}(f_x,f_y)$
$k(x,y)$
200
$\hat{u}(f_x,f_y) \otimes \hat{k}(f_x,f_y)$
200
$u * k$
200
(1) Image collected by photodiodes or data bus is projected into FFT block. Kernel image(s) are projected from bottom block
(2) FFTs are multiplied in analog domain in the receive arrays
(3) Product image is projected into bottom FFT block to obtain the convolution

FIG. 2

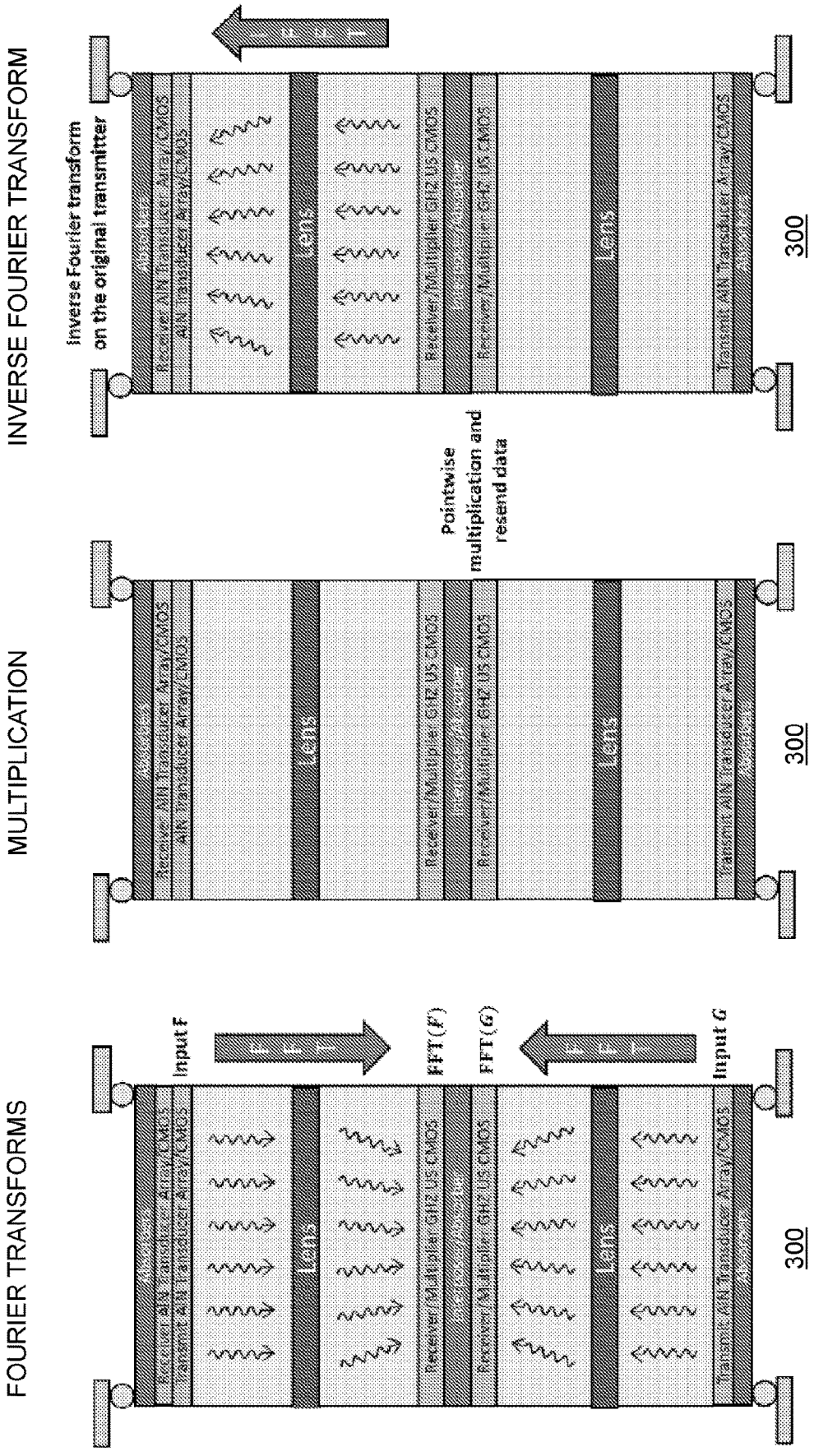

FOURIER TRANSFORMS
MULTIPLICATION
INVERSE FOURIER TRANSFORM
Input F
Input G
FFT(F)
FFT(G)
Pointwise multiplication and resend data
inverse Fourier transform on the original transmitter
Lens
300

FIG. 3

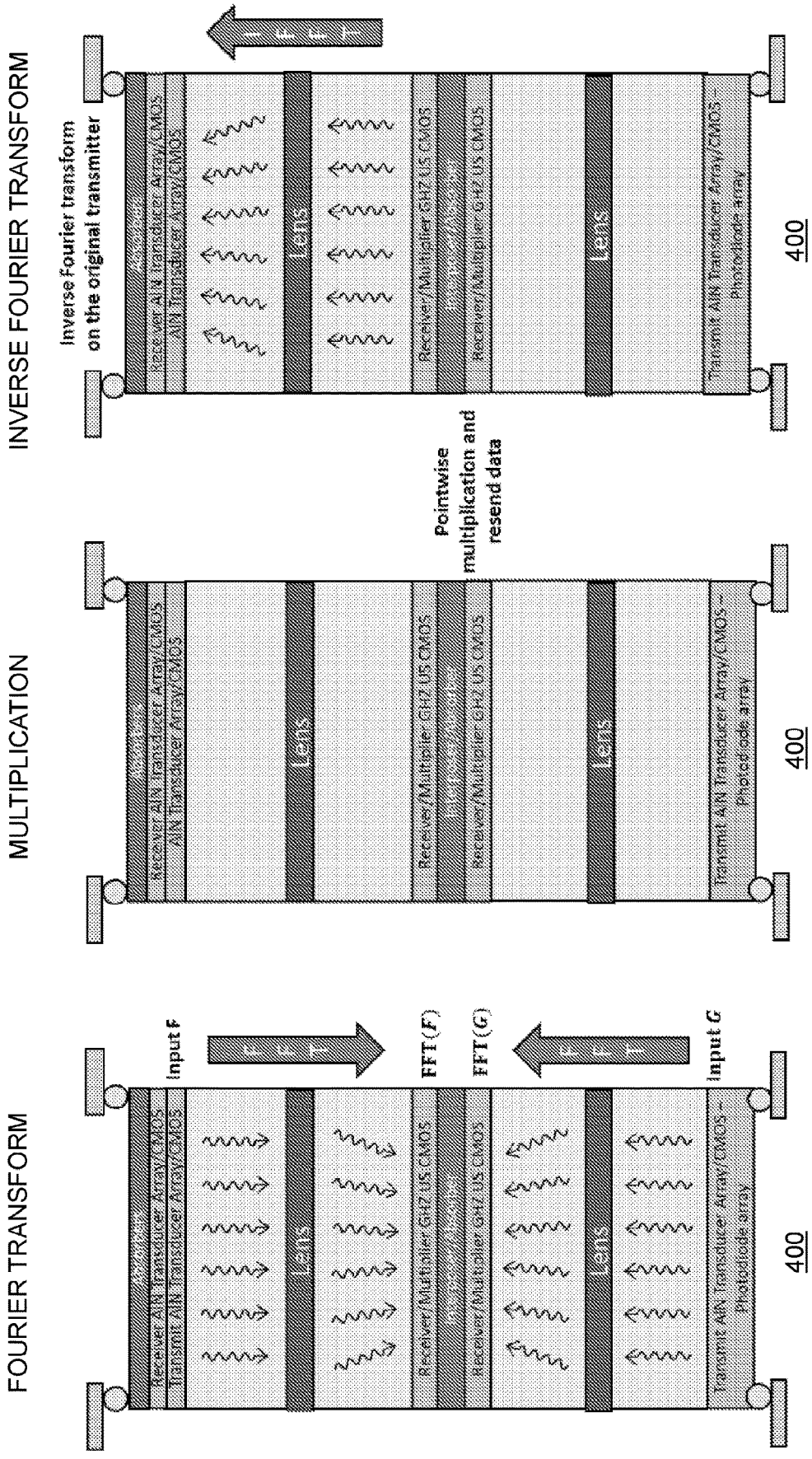

FOURIER TRANSFORM
MULTIPLICATION
INVERSE FOURIER TRANSFORM
Input F
Input G
FFT(F)
FFT(G)
Inverse Fourier transform on the original transmitter
Pointwise multiplication and resend data
Lens
Lens
Lens
400
400
400

FIG. 4

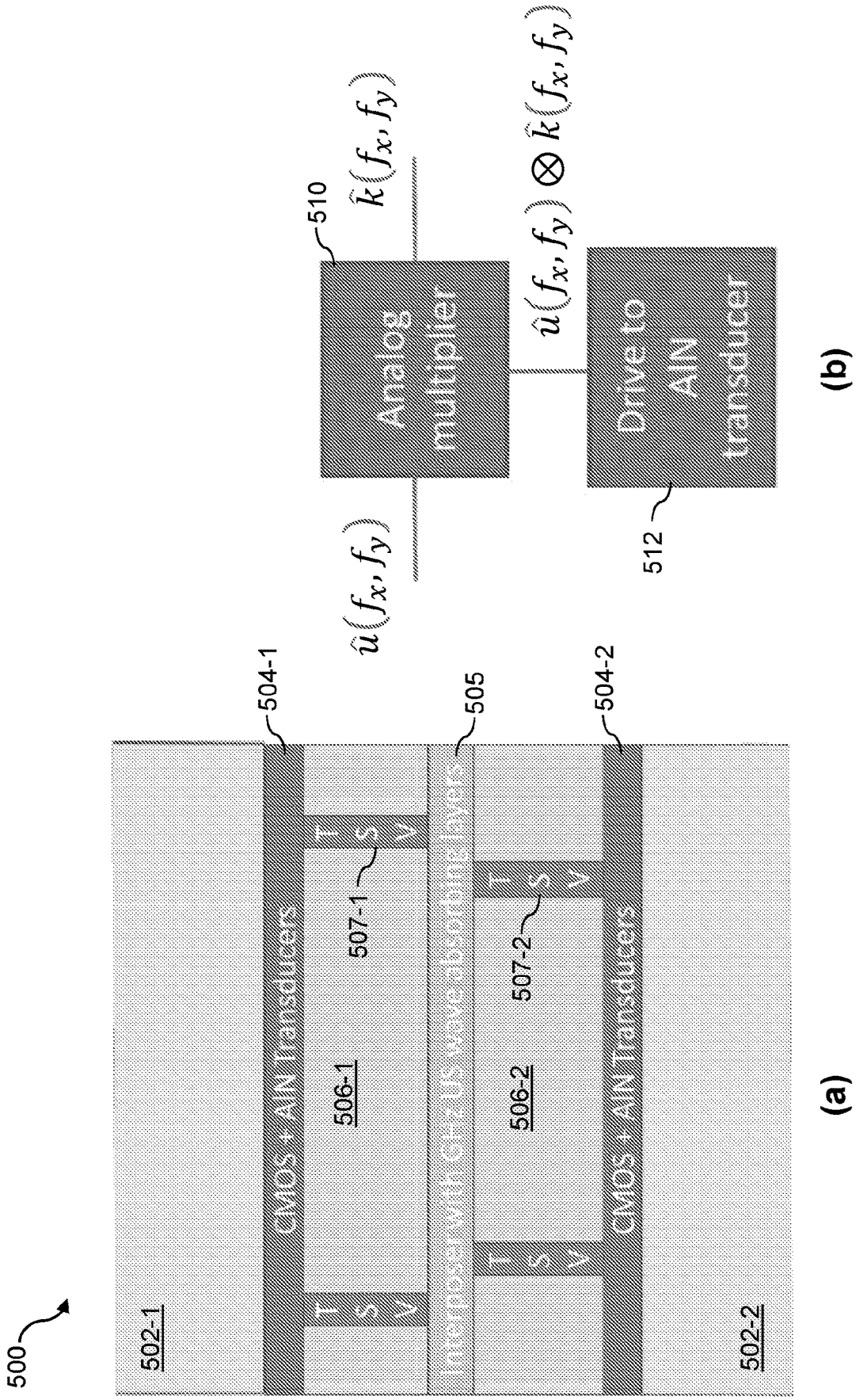

500
502-1
CMOS + AlN transducers
T
S
V
506-1
507-1
504-1
505
Interleaved with CMOS waveguide layers
T
S
V
507-2
506-2
504-2
CMOS + AlN Transducers
502-2
(a)
$\hat{u}(f_x, f_y)$
510
$\hat{k}(f_x, f_y)$
Analog multiplier
$\hat{u}(f_x, f_y) \otimes \hat{k}(f_x, f_y)$
Drive to AlN transducer
512
(b)

FIG. 5

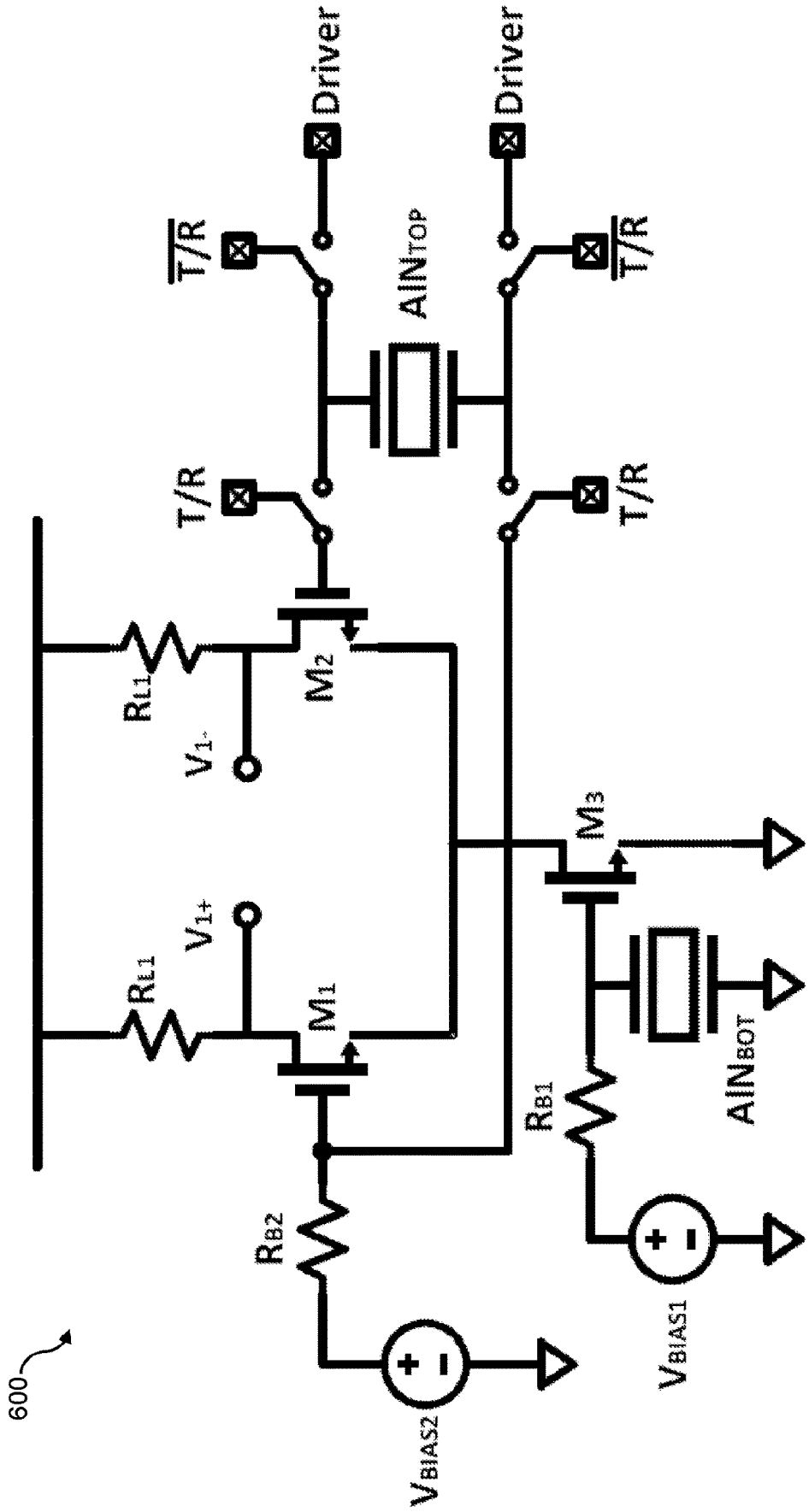

600
Driver
Driver
T/R
T/R
AlN_TOP
T/R
T/R
R_L1
M_2
V_1-
V_1+
R_L1
M_1
M_3
AlN_BOT
R_B1
R_B2
V_BIAS1
V_BIAS2

FIG. 6

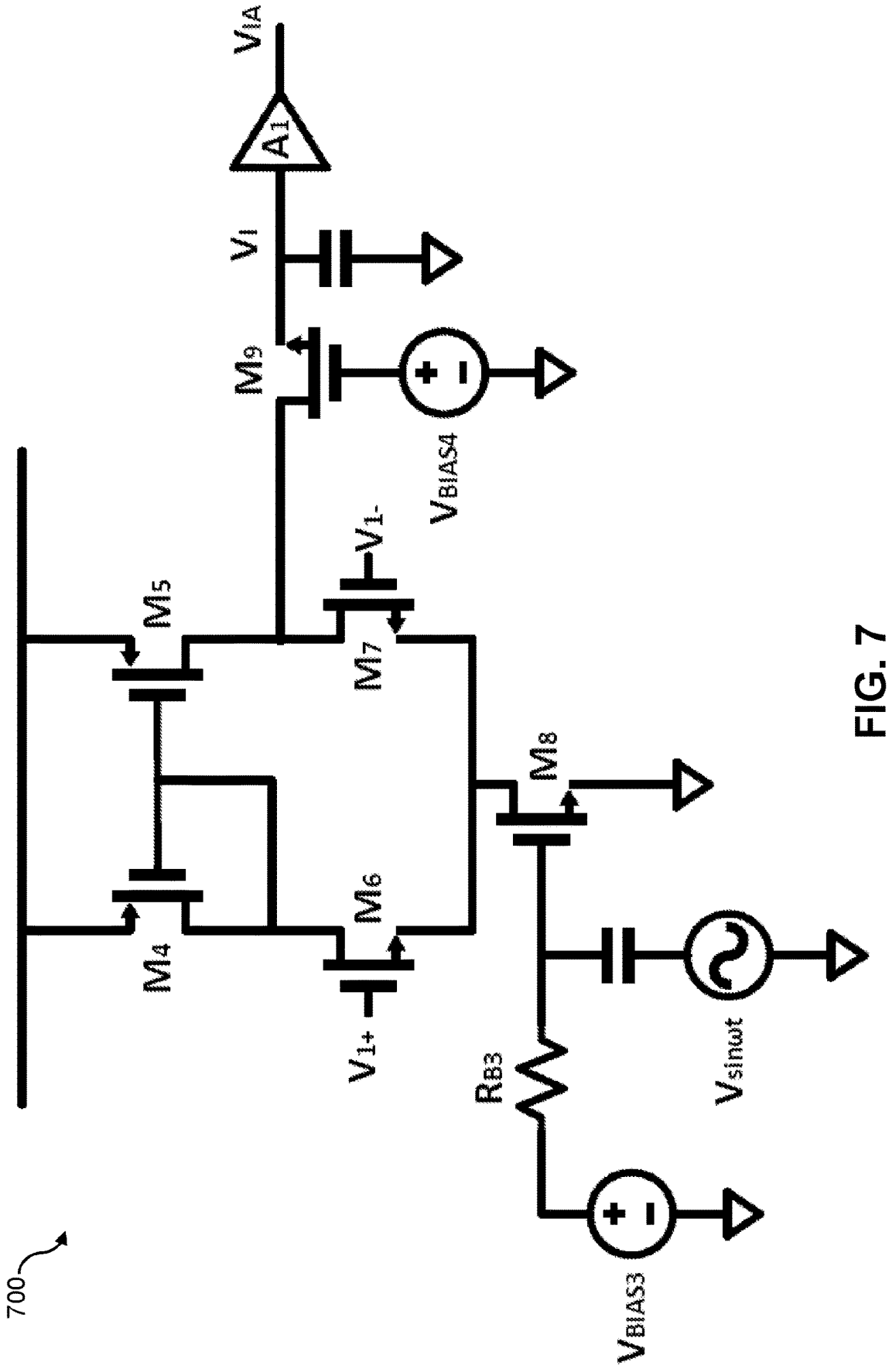

700
V_{IA}
A_1
V_I
M_9
V_{BIAS4}
V_{1-}
M_5
M_7
M_8
M_4
M_6
V_{1+}
R_{B3}
$V_{sin\omega t}$
V_{BIAS3}

FIG. 7

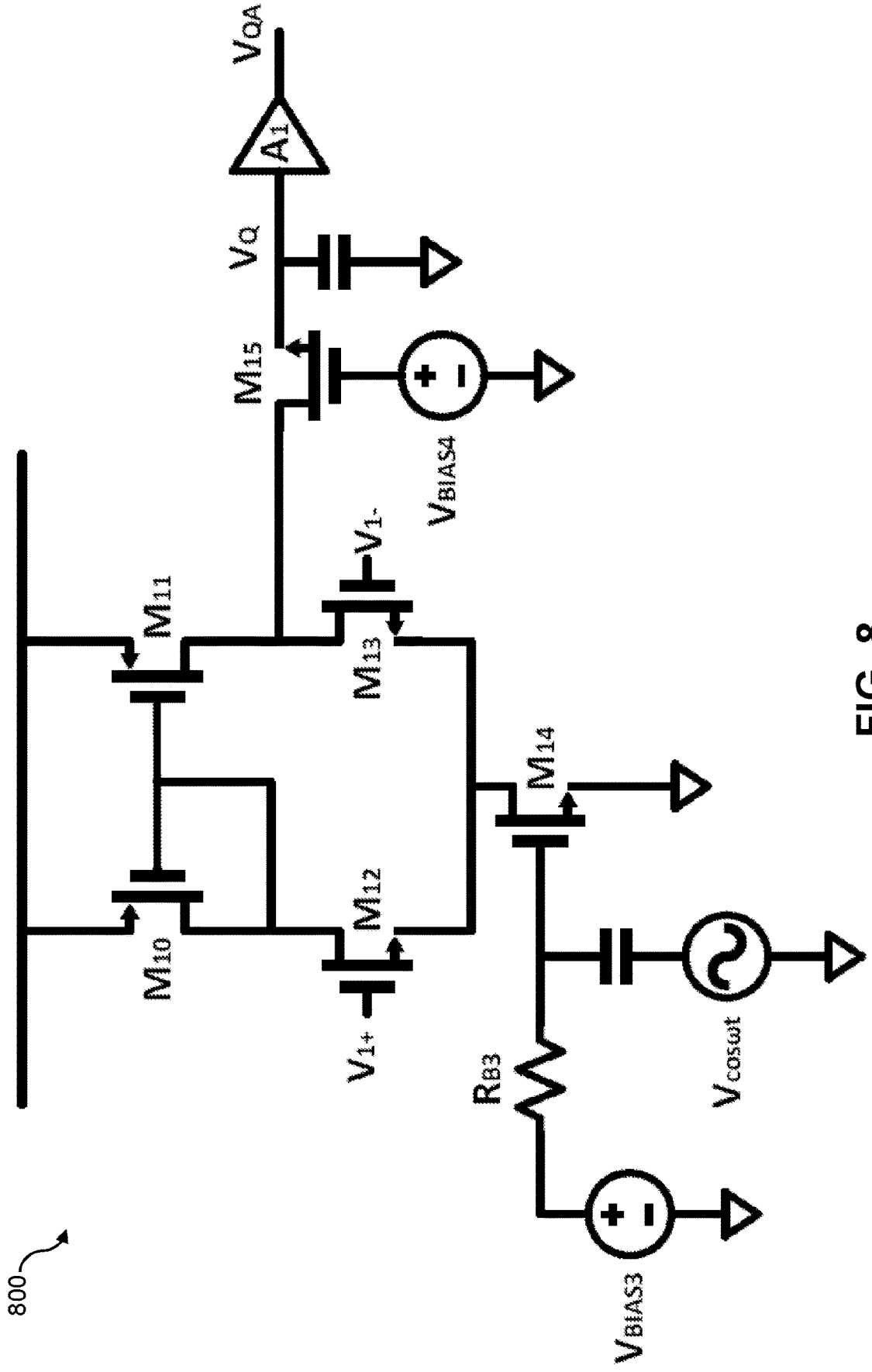

800
V_{OA}
A1
V_Q
M_{15}
V_{BIAS4}
V_{1-}
M_{11}
M_{13}
M_{14}
M_{10}
M_{12}
V_{1+}
R_{B3}
$V_{cos\omega t}$
V_{BIAS3}

FIG. 8

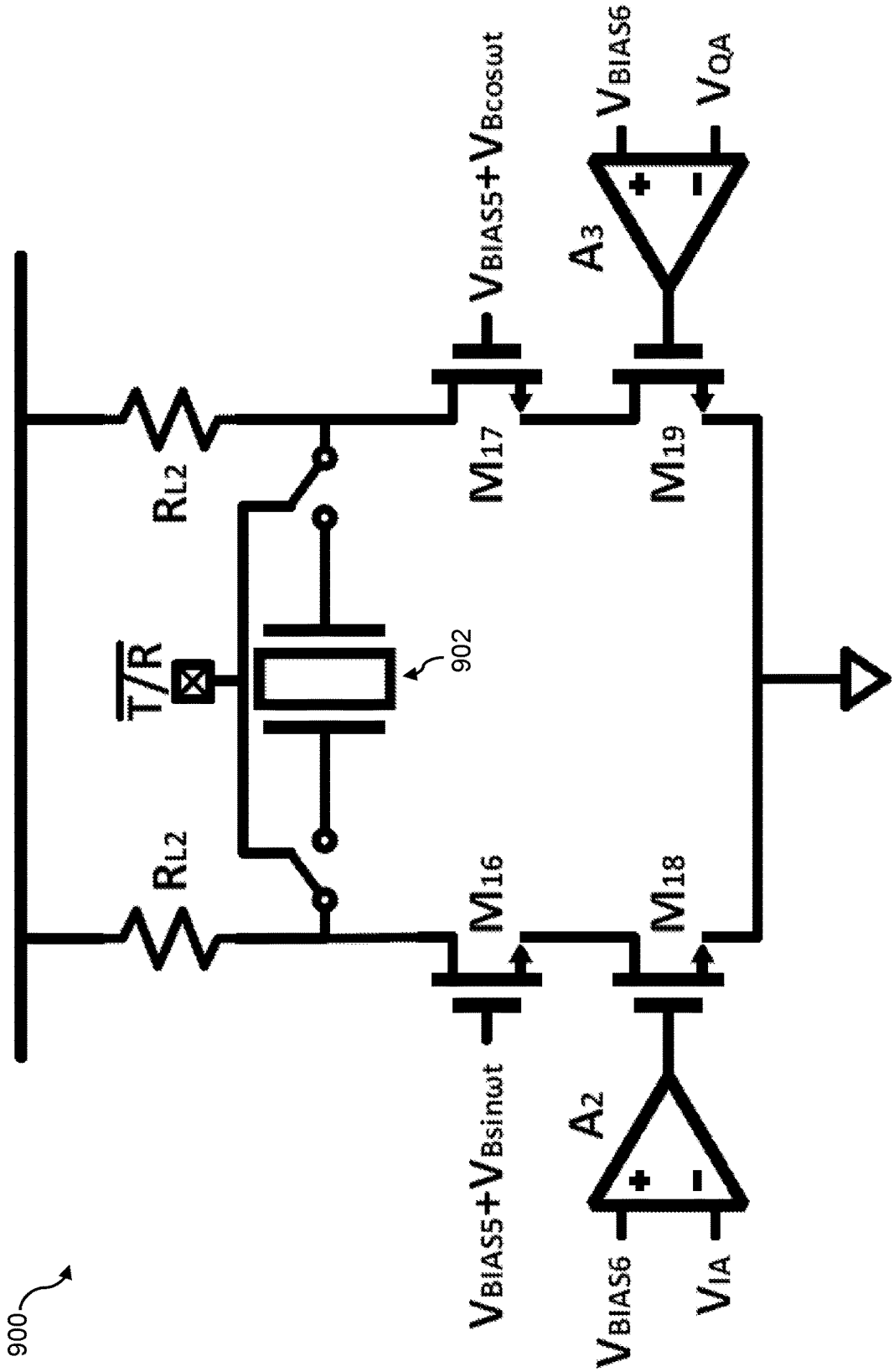

900
$V_{BIAS5}+V_{Bsin\omega t}$
V_{BIAS6}
V_{IA}
A_2
M_{16}
M_{18}
R_{L2}
R_{L2}
T/R
902
$V_{BIAS5}+V_{Bcos\omega t}$
V_{BIAS6}
V_{QA}
A_3
M_{17}
M_{19}

FIG. 9

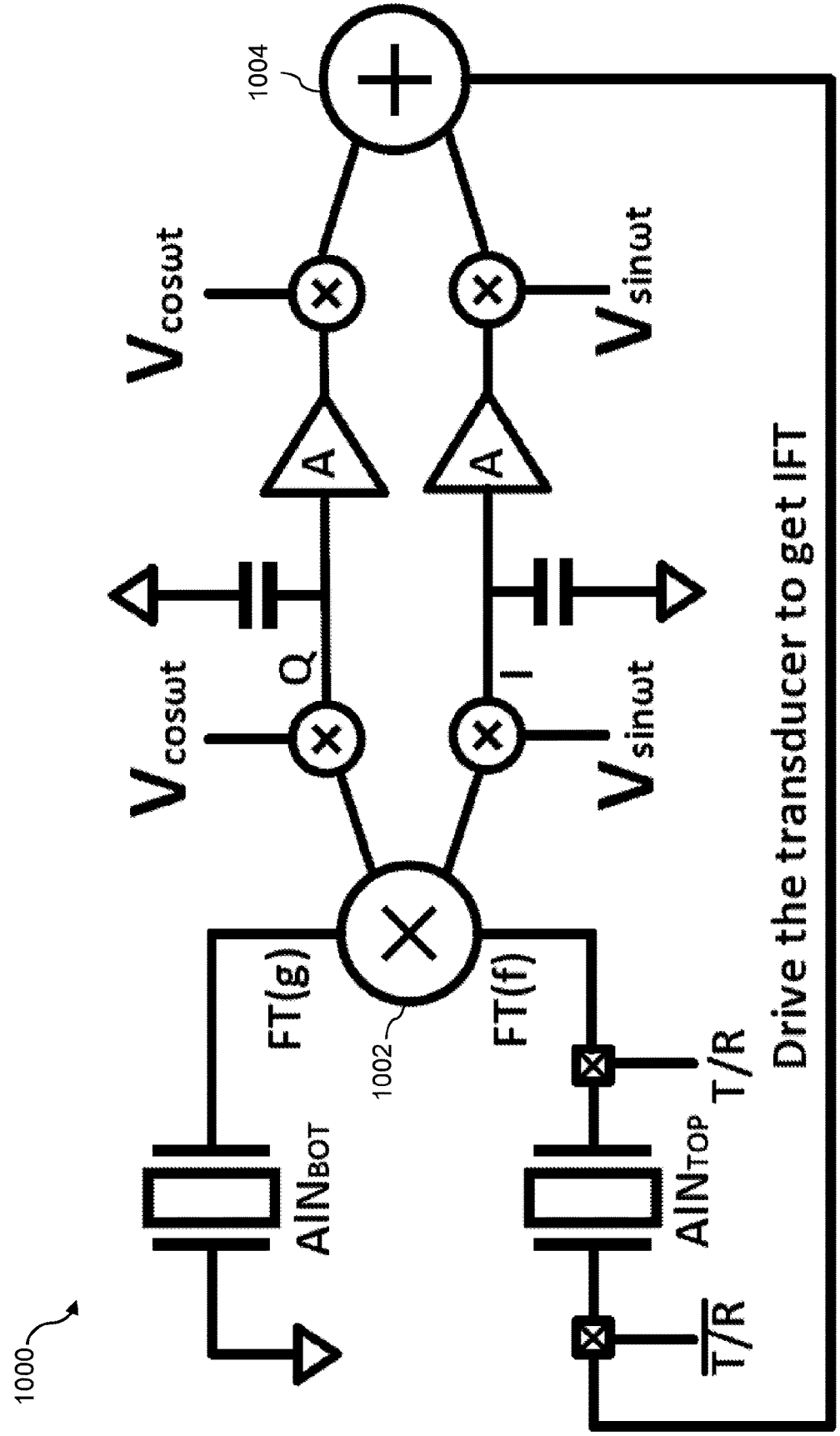

1000
1004
$V_{cos\omega t}$
$V_{sin\omega t}$
A
A
Q
I
$V_{cos\omega t}$
$V_{sin\omega t}$
FT(g)
FT(f)
1002
T/R
T/R
AlN_{BOT}
AlN_{TOP}
Drive the transducer to get IFT

FIG. 10

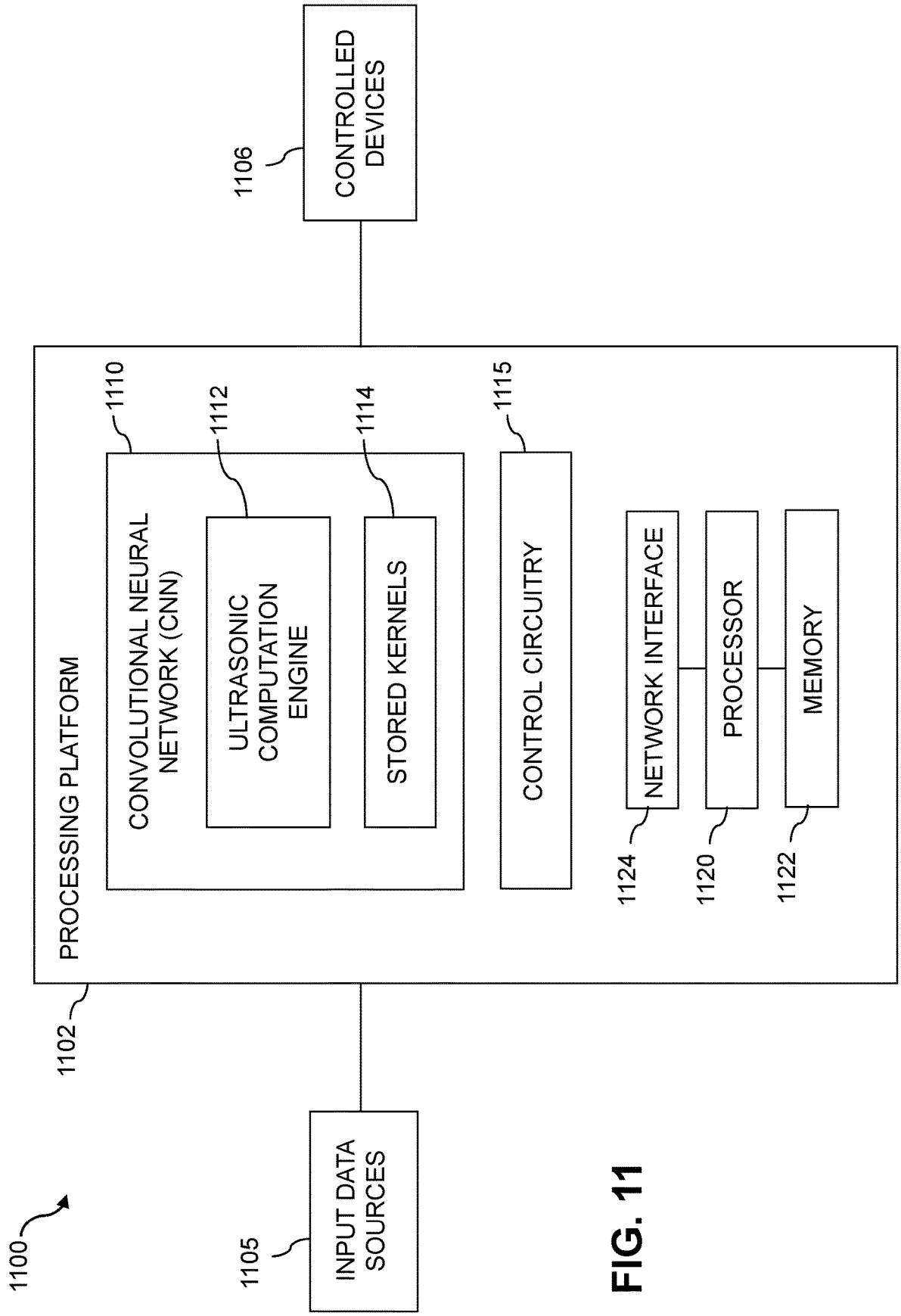

1100
1105
INPUT DATA SOURCES
1102
PROCESSING PLATFORM
1110
CONVOLUTIONAL NEURAL NETWORK (CNN)
1112
ULTRASONIC COMPUTATION ENGINE
1114
STORED KERNELS
1115
CONTROL CIRCUITRY
1124
NETWORK INTERFACE
1120
PROCESSOR
1122
MEMORY
1106
CONTROLLED DEVICES

FIG. 11

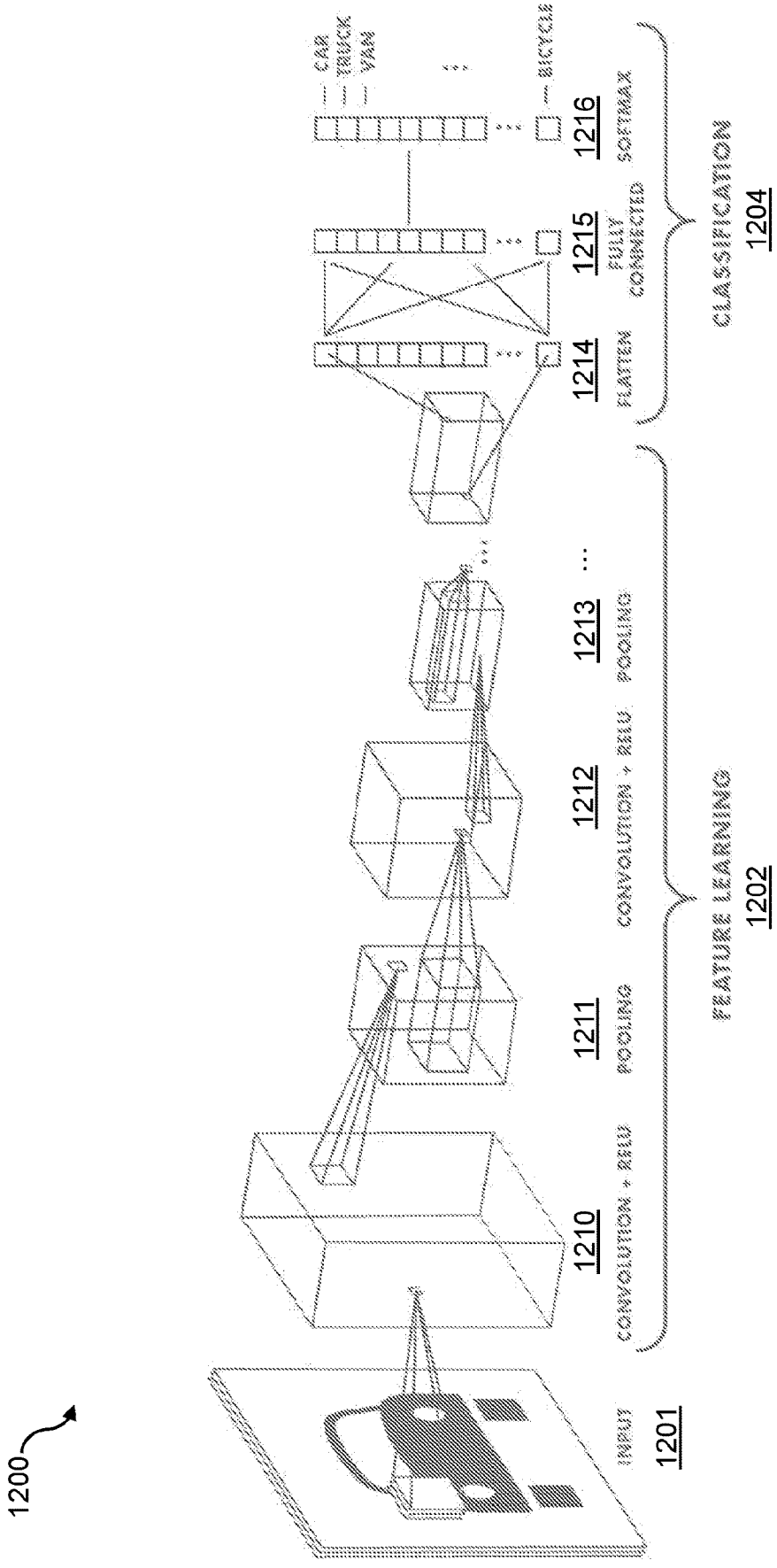

1200
1201
INPUT
1210
CONVOLUTION + RELU
1211
POOLING
1212
CONVOLUTION + RELU
1213
POOLING
FEATURE LEARNING
1202
1214
FLATTEN
1215
FULLY CONNECTED
1216
SOFTMAX
CAR
TRUCK
VAN
BICYCLE
CLASSIFICATION
1204

FIG. 12

ULTRASONIC COMPUTATION HARDWARE FOR CONVOLUTIONAL NEURAL NETWORK COMPUTING AND OTHER COMPUTATION APPLICATIONS

RELATED APPLICATION

[0001] The present application claims priority to U.S. Provisional Patent Application Ser. No. 63/177,220, filed Apr. 20, 2021, which is incorporated by reference herein in its entirety.

STATEMENT OF GOVERNMENT SUPPORT

[0002] This invention was made with U.S. government support under Grant No. HR00117298430001 of the Defense Advanced Research Projects Agency (DARPA)—Defense Sciences Office (DSO). The U.S. government has certain rights in the invention.

FIELD

[0003] The field relates generally to ultrasonic devices, and more particularly to ultrasonic computation hardware, which can be utilized to perform computations in information processing systems.

BACKGROUND

[0004] As the amounts of data created in a wide variety of different applications continue to grow exponentially, there is need to process such data more efficiently, such as with enhanced processing speed but with lower power consumption. For example, cameras with ever-increasing pixel counts and continuous sensor data collection have created vast amounts of visual data, which needs to be processed quickly to be useful in applications such as driverless cars, drones and autonomous robots. However, conventional circuit-based approaches are running up against the limits of Moore's law as well as associated thermal limitations of integrated circuits. Accordingly, improved computational arrangements are needed to more efficiently process data in these and numerous other contexts.

SUMMARY

[0005] Illustrative embodiments disclosed herein provide ultrasonic computation hardware that can, for example, perform highly efficient convolution operations in neural network computing applications, such as those involving driverless cars, drones and autonomous robots. Such ultrasonic computation hardware can similarly provide efficient computation in a wide variety of other computation applications.

[0006] In one embodiment, an ultrasonic computation apparatus includes first and second ultrasonic transducer arrays arranged on opposing ends thereof, and further comprises first and second ultrasonic propagation regions arranged between the first and second ultrasonic transducer arrays and proximate respective ones of the first and second ultrasonic transducer arrays, and an intermediate computational structure arranged between the first and second ultrasonic propagation regions. Respective first and second input signals applied to respective ones of the first and second ultrasonic transducer arrays cause propagation of corresponding ultrasonic waves through the respective first and second ultrasonic propagation regions towards the intermediate computational structure. The intermediate computational structure is configured to receive the propagating ultrasonic waves from the respective first and second ultrasonic propagation regions and to generate from the received propagating ultrasonic waves an additional signal that is a function of the first and second signals.

[0007] In some embodiments, the ultrasonic computation apparatus is configured to perform high-efficiency convolutions of an input image with kernel images in a convolutional neural network (CNN), deep neural network (DNN) or other type of neural network or machine learning system. Under conventional circuit-based approaches, the mathematical complexity of such convolution operations is very high and can limit the speed at which the CNN or DNN can perform image recognition.

[0008] Some embodiments disclosed herein advantageously provide an ultrasonic computation apparatus configured to generate respective Fourier transforms of the input image and a given one of the kernels, to multiply the resulting Fourier transforms in the analog domain, and to perform an inverse Fourier transform to obtain the convolution of the input image and the given one of the kernel images. Such arrangements substantially reduce the complexity of the convolution operations, allowing the operations to be performed at high speed and with low power consumption, thereby significantly enhancing the overall performance of the CNN or DNN in its image processing tasks.

[0009] As noted above, a wide variety of other types of computations can be efficiently performed in other applications using ultrasonic computation hardware as disclosed herein. Accordingly, it should be understood that the foregoing arrangements are only examples, and numerous alternative arrangements are possible.

[0010] These and other illustrative embodiments include but are not limited to systems, methods, apparatus, processing devices, integrated circuits, and computer program products comprising processor-readable storage media having software program code embodied therein.

BRIEF DESCRIPTION OF THE FIGURES

[0011] FIG. **1** shows an ultrasonic computation engine in an illustrative embodiment.

[0012] FIG. **2** illustrates the manner in which the ultrasonic computation engine of FIG. **1** performs a convolution of an input image with one or more kernel images in an illustrative embodiment.

[0013] FIGS. **3** and **4** show additional examples of ultrasonic computation engines performing convolution operations in illustrative embodiments.

[0014] FIG. **5** in part (a) thereof shows a more detailed view of a portion of an ultrasonic computation engine and in part (b) thereof a corresponding block diagram, in illustrative embodiments.

[0015] FIG. **6** is a schematic diagram of a multiplier circuit implemented in an ultrasonic computation engine in an illustrative embodiment.

[0016] FIGS. **7** and **8** are schematic diagrams of respective in-phase (I) and quadrature-phase (Q) implementations of mixer and DC converter circuits for converting output of the multiplier circuit of FIG. **6** to respective I and Q components.

[0017] FIG. **9** is a schematic diagram of a product transmission circuit implemented in an ultrasonic computation engine in an illustrative embodiment.

[0018] FIG. **10** is a schematic diagram showing an example of intermediate multiplication circuitry that incorporates a multiplier circuit, I and Q mixer and DC converter circuits, and a product transmission circuit, all implemented in an ultrasonic computation engine in an illustrative embodiment.

[0019] FIG. **11** is a block diagram of an information processing system that incorporates a neural network having as a component thereof an ultrasonic computation engine in an illustrative embodiment.

[0020] FIG. **12** shows a more detailed view of an example implementation of the neural network of FIG. **11** that has as a component thereof an ultrasonic computation engine in an illustrative embodiment.

DETAILED DESCRIPTION

[0021] Illustrative embodiments include ultrasonic computation engines and other types of ultrasonic computation hardware. Such ultrasonic computation hardware in some embodiments can be implemented, for example, in information processing systems comprising one or more processing platforms each having at least one computer, server or other processing device. For example, a given such processing platform implementing a neural network or other type of machine learning system can incorporate one or more ultrasonic computation engines that are utilized by the machine learning system to perform various types of computations, such as convolutions of input images with kernel images in a CNN, DNN or other type of neural network or machine learning system. A number of examples of such systems will be described in detail herein. It should be understood, however, that the disclosed embodiments are more generally applicable to a wide variety of other computation contexts, illustratively involving other arrangements of ultrasonic computation hardware and associated processing devices, memories and additional or alternative components. Accordingly, the term "information processing system" as used herein is intended to be broadly construed so as to encompass these and other arrangements.

[0022] Some illustrative embodiments provide methods and systems for a computing architecture configured to perform neural network operations.

[0023] There are significant problems with and limitations in current neural networks, which restrict their performance and functionality. For example, a common feature of CNNs is to perform the convolution of a set of kernel images with an input image. The convolutions are then weighted to estimate if the image contains the feature image. The mathematical complexity of the convolution operation is very high and can limit the speed at which image recognition can occur. It is also known that the Fourier transform of a convolution of two functions is the product of the Fourier transforms of the two functions.

[0024] Some embodiments disclosed herein recognize that there are potential problems, limitations, and/or disadvantages with conventional neural networks and associated computational arrangements as discussed herein. Various embodiments may be advantageous in that they may solve or reduce one or more of the potential problems, limitations, and/or disadvantages discussed herein.

[0025] For example, some embodiments provide ultrasonic computation hardware that can provide significantly enhanced computational efficiency in applications such as driverless cars, drones, and autonomous robots. In these and other applications, cameras and other types of sensors transmit data to an edge device of an information processing system configured in accordance with a core-edge architecture.

[0026] Some embodiments disclosed herein can provide orders of magnitude efficiency gains for performing certain types of computations in these and numerous other applications. For example, illustrative embodiments of the present disclosure can perform real-time convolutions of an input image with a kernel image in a highly efficient manner, leading to substantial improvement in image recognition speeds. Additional improvements in terms of power consumption as well as weight and cost are also provided in performing such image-related computations. Moreover, illustrative embodiments can be readily integrated with photodiode arrays or other types of image sensor arrays of optical imagers.

[0027] Illustrative embodiments disclosed herein therefore overcome significant drawbacks of conventional arrangements. Although these illustrative embodiments are described in conjunction with particular applications, the disclosed techniques can be adapted in a straightforward manner to other contexts and use cases.

[0028] In some embodiments, the ultrasonic computation hardware generates and detects GigaHertz (GHz) ultrasonic waves in silicon substrates. For example, thin-film piezoelectric film based transducers are illustratively formed on top of CMOS wafers, and the embedded CMOS is used to drive and sense strain waves. An array of two-dimensional (2D) piezoelectric pixels can launch waves into the bulk silicon to produce a Fourier transform of the input phase and magnitude of voltages applied at the input frame. The CMOS integration utilized in some embodiments allows very low energy and fast operation. The transducers illustratively include piezoelectric ultrasonic transducers formed using an Aluminum Nitride (AlN) process on a CMOS substrate, although other types of transducers can be used.

[0029] FIG. **1** shows an ultrasonic computation engine **100** in an illustrative embodiment. The ultrasonic computation engine **100** is an example of what is more generally referred to herein as "ultrasonic computation hardware" or an "ultrasonic computation apparatus." The particular configuration and functionality of the ultrasonic computation engine **100** is presented by way of illustrative example only, and should not be construed as limiting in any way.

[0030] The ultrasonic computation engine **100** comprises GHz CMOS transmitter/receiver arrays **102**A, **102**B, **102**C and **102**D. Different ones of the transmitter/receiver arrays **102** may have different configurations. For example, a given such array may be transmit only, receive only, or configured to both transmit and receive. Accordingly, the term "transmitter/receiver" as used herein is intended to be construed as transmitter and/or receiver. First and second ones of these ultrasonic transducer arrays **102**A and **102**B are arranged on opposing ends of the ultrasonic computation engine **100**. The ultrasonic computation apparatus **100** further comprises first and second ultrasonic propagation regions **104** and **106** arranged between the first and second ultrasonic transducer arrays **102**A and **102**B and proximate respective ones of the first and second ultrasonic transducer arrays **102**A and **102**B,

and an intermediate computational structure arranged between the first and second ultrasonic propagation regions **104** and **106**.

[0031] The term "proximate" as used herein is intended to be broadly construed, so as to encompass, for example, arrangements in which a transducer array is adjacent to an ultrasonic propagation region or otherwise arranged relative to an ultrasonic propagation region in a manner that permits the transducer array to transmit and/or receive ultrasonic waves from the ultrasonic propagation region.

[0032] The intermediate computational structure in this embodiment more particularly comprises ultrasonic transducer arrays **102**C and **102**D and an interposer layer **108** arranged between the ultrasonic transducer arrays **102**C and **102**D. At least the first ultrasonic transducer array **102**A of the first and second ultrasonic transducer arrays **102**A and **102**B has a photodiode array **110** formed thereon as illustrated in the figure. The ultrasonic transducer arrays **102**A, **102**B, **102**C and **102**D are supported at least in part by respective printed circuit boards (PCBs) **112**A, **112**B, **112**C and **112**D within the ultrasonic computation engine **100**. As indicated above, this particular arrangement of components of ultrasonic computation engine **100** is exemplary only, and can be varied in other embodiments.

[0033] In the ultrasonic computation engine **100**, respective first and second input signals applied to respective ones of the first and second ultrasonic transducer arrays **102**A and **102**B cause propagation of corresponding ultrasonic waves through the respective first and second ultrasonic propagation regions **104** and **106** towards the intermediate computational structure. The intermediate computational structure is configured to receive the propagating ultrasonic waves from the respective first and second ultrasonic propagation regions **104** and **106** via the respective ultrasonic transducer arrays **102**C and **102**D and to generate from the received propagating ultrasonic waves an additional signal that is a function of the first and second signals.

[0034] For example, the intermediate computational structure in the ultrasonic computation engine **100** is illustratively configured to generate the additional signal at least in part as a product of transformed versions of the respective first and second signals as represented by the corresponding received propagating ultrasonic waves.

[0035] More particularly, the first ultrasonic transducer array **102**A and the first ultrasonic propagation region **104** are configured such that the propagation of ultrasonic waves corresponding to the first signal through the first ultrasonic propagation region **104** causes a Fourier transform of the first signal to be received at a first side of the intermediate computational structure proximate the first ultrasonic propagation region **104**.

[0036] Similarly, the second ultrasonic transducer array **102**B and the second ultrasonic propagation region **106** are configured such that the propagation of ultrasonic waves corresponding to the second signal through the second ultrasonic propagation region **106** causes a Fourier transform of the second signal to be received at a second side of the intermediate computational structure proximate the second ultrasonic propagation region **106**.

[0037] The propagation regions **104** and **106** each comprise an ultrasonic lens, illustratively formed in silicon, that results in production of an equivalent Fourier transform of the ultrasonic waves at the corresponding receiver array **102**C or **102**D. The thickness of the two sides of the propagation regions are illustratively chosen to be the focal length of the lens. The transit time of the wavefront from the transmitter array to the receiver array can be estimated as $t_{transit}=2F/c_{sound}$ where F is the focal length of the lens and c_{Sound} is the speed of sound in the propagation regions. The focal length is typically proportional to the aperture size A of the ultrasonic transducer array by the f_{stop} number. If N pixels can fit in the aperture, and each pixel occupies space proportional to the ultrasonic wavelength ($\lambda=c_{sound}/f$), where f is the ultrasonic frequency, an N×N pixel Fourier transform can be obtained in a time $t_{compute}\sim t_{transit}\sim N\lambda/c_{sound}=N*T$. Here T is the period of the ultrasonic frequency. Hence, an N×N pixel Fourier transform can be obtained in illustrative embodiments in a time proportional to the number of pixels N. In contrast, digital domain Fourier transforms can require a number of computational operations given by $O(N^2 \text{ Log } N)$.

[0038] The analog computation of a Fourier transform as disclosed herein can therefore be performed at high speeds, particularly when N is large, and the ultrasonic frequencies are comparable to the clock speeds being used in computing systems. As has been evident in microcontroller development over the last decade, clock frequencies have saturated to a few GHz, owing to the excessive power consumption, also known as the power-wall. An ultrasonic Fourier transform of the type disclosed herein, if the ultrasonic frequency is in the Ghz range, can achieve small pixels owing to small wavelengths, and can provide speed gains over digital computation. The operation at higher frequency also enables reduction of the overall length of the ultrasonic travel distance, enabling small mm-cm scale systems that be integrated onto common PCB assemblies.

[0039] In some embodiments, the ultrasonic frequencies utilized in the ultrasonic computation engine **100** are greater than about 1 GHz, although other ultrasonic frequencies can be used in other embodiments, and the term "ultrasonic" as used herein is therefore intended to be broadly construed. For example, in other embodiments, ultrasonic frequencies in a range of about 0.5 GHz to 5 GHz can be used, although this is again just an example and should not be construed as limiting in any way.

[0040] As indicated above, the lenses within the respective ultrasonic propagation regions **104** and **106** are illustratively formed in silicon. For example, the lenses can be made in silicon at least in part by filling silicon cavities with silicon dioxide (SiO_2), although other materials and techniques can be used. As another example, the lenses can be made in fused quartz, by etching it in pillars, and filling with silicon.

[0041] The ultrasonic propagation regions **104** and **106** are illustratively also formed at least in part using silicon, and may comprise, for example, fused silica, as indicated in the illustrative embodiment of FIG. **1**. Other low-loss materials such as, for example, sapphire and fused quartz, can additionally or alternatively be used. It should also be noted, in some embodiments, a particular substrate may be anisotropic, such that ultrasonic waves propagate at different speeds in different directions, and transmission compensation may be used in embodiments that utilize such substrates, as will be appreciated by those skilled in the art.

[0042] It is assumed in the FIG. **1** embodiment that at least the first ultrasonic transducer array **102**A comprises a transmit/receive ultrasonic transducer array that is illustratively configured both to transmit ultrasonic waves to its corresponding ultrasonic propagation region **104**, and to receive

ultrasonic waves from its corresponding ultrasonic propagation region **104**. In some embodiments, the second transducer **102**B may be similarly configured. Alternatively, the second ultrasonic transducer array **102**B may be configured to as transmit only, so as to only transmit ultrasonic waves to its corresponding ultrasonic propagation region **106**.

[0043] The Fourier transforms of the respective first and second signals are illustratively received by the ultrasonic transducer arrays **102**C and **102**D at the respective first and second sides of the intermediate computational structure. The intermediate computational structure is illustratively configured to generate the additional signal as a product of the Fourier transforms of the respective first and second signals.

[0044] The generation of the additional signal in the intermediate computational structure illustratively causes propagation of additional ultrasonic waves through one of the first and second ultrasonic propagation regions **104** and **106** towards one of the first and second ultrasonic transducer arrays **102**A and **102**B such that a function of the additional signal is received at that one of the first and second ultrasonic transducer arrays **102**A and **102**B.

[0045] For example, the additional signal generated in the intermediate computational structure illustratively comprises a product of Fourier transforms of respective ones of the first and second signals and the function of the additional signal received at one of the first and second ultrasonic transducer arrays **102**A and **102**B comprises an inverse Fourier transform of the product of the Fourier transforms, where the inverse Fourier transform of the product of the Fourier transforms is a convolution of the first and second signals.

[0046] The ultrasonic computation engine **100** therefore computes the convolution of the first and second signals, in a highly efficient manner. Other types of computations can be performed using the ultrasonic computation engine in other embodiments.

[0047] In some embodiments, the first signal comprises an input image applied to the first ultrasonic transducer array **102**A and the second signal comprises a kernel image applied to the second ultrasonic transducer array **102**B. The first signal is illustrated received by the first ultrasonic transducer array **102**A via the photodiode array **110**. The photodiode array **110** is an example of what is more generally referred to herein as an "image sensor array." Other types of image sensor arrays can be used in other embodiments.

[0048] The kernel image is illustratively one of a plurality of kernel images of a CNN that comprises or otherwise utilizes the ultrasonic computation engine **100**. The ultrasonic computation engine **100** in such an arrangement is configured to generate a plurality of convolutions of the input image with respective ones of the kernel images as respective convolution operations of the CNN. The convolutions in some embodiments are utilized as new images from which features are extracted, and which can then be used to train the CNN. Although CNNs are utilized in some embodiments herein, it is to be appreciated that the disclosed technique are more generally applicable to performing high-efficiency computations in DNNs or other types of neural networks, as well as other types of machine learning systems.

[0049] As indicated above, the ultrasonic computation engine **100** performs convolutions using the property that the inverse Fourier transform of the product of Fourier transforms of the input and kernel images is the convolution of the input and kernel images. The ultrasonic computation engine **100** may be viewed as comprising stacked upper and lower block structures that produce Fourier transforms of the respective input image and kernel image, with multiplication of those Fourier transforms being performed in the intermediate computational structure in the analog domain using analog multipliers. The product is then transmitted back through the lower block structure to form the inverse Fourier transform of the product, which provides the convolution of the input image and the kernel image.

[0050] The time needed to compute the convolution in some embodiments is given approximately by 2L/c, where L is the thickness of each of the upper and lower block structures, and c is the sound speed. This results in a time of about 4 microseconds (μs) for a large arrays, enabling convolutions to be performed at speeds on the order of megaframes per second. The volume of the ultrasonic computation engine **100** in some embodiments is on the order of about 2-3 cubic centimeters (cc), which is readily compatible with operation in cars, drones and robots, and in numerous other applications.

[0051] In some embodiments, the ultrasonic transducer arrays **102** illustratively comprise respective CMOS integrated piezoelectric ultrasonic transducer arrays, implemented in some embodiments as respective 2D AlN arrays. Typical transducer capacitances are in 10-100 femtoFarad (fF) range, and can be 1V CMOS compatible, and duty cycled for low power consumption operation. Phase can be easily measured using IQ mixers and row-column based ADC arrays.

[0052] In the ultrasonic computation engine **100**, the first and second ultrasonic propagation regions **104** and **106** each include a first sub-region adjacent its corresponding one of the first and second ultrasonic transducer arrays **102**A and **102**B, a second sub-region adjacent the intermediate computational structure, and a lens arranged between the first and second sub-regions, as illustrated.

[0053] The first and second sub-regions in the FIG. **1** embodiment illustratively comprise fused silica, and the lens comprises a silicon lens, with the height of the corresponding first and second ultrasonic propagation regions **104** and **106** each being approximately 1 centimeter (cm), although other materials and dimensions may be used.

[0054] The various layers of the ultrasonic transducer arrays **102** are illustratively interconnected at least in part using flip-chip bonding arrangements with associated bond bumps, although alternative techniques can be used.

[0055] As will be described in more detail elsewhere herein, the intermediate computational structure illustratively comprises intermediate multiplication circuitry including a plurality of multiplier circuits, a plurality of in-phase and quadrature-phase mixer and converter circuits coupled to corresponding outputs of the multiplier circuits, and a plurality of product transmission circuits coupled to corresponding outputs of the mixer and converter circuits, arranged to facilitate the performance of the computations mentioned above. Additional or alternative circuitry can be incorporated into the intermediate computational structure in order to perform other types of computations.

[0056] FIG. **2** illustrates the manner in which an ultrasonic computation engine **200**, similar to the ultrasonic computation engine **100** of FIG. **1**, performs a convolution of an

input image with one or more kernel images. The figure shows three different instances of the ultrasonic computation engine **200**, arranged from left to right, with the left-most instance illustrating computation of the Fourier transforms of the respective input image and the kernel image by projection through the respective upper and lower ultrasonic propagation regions to the intermediate computational structure, the middle instance illustrating the multiplication of the two Fourier transforms in the intermediate computational structure, and the right-most instance illustrating the computation of the inverse Fourier transform of the product of the Fourier transforms of the input image and the kernel image by projection from the intermediate computational structure through the lower ultrasonic propagation region. PCBs similar to those utilized in the ultrasonic computation engine **100** are not shown in this figure for simplicity and clarity of illustration.

[0057] In the FIG. **2** embodiment, the first and second transducer layers at the top and bottom of the ultrasonic computation engine **200** comprise respective GHz ultrasonic CMOS layers configured to transmit the input image and the kernel image, respectively. The input image and the kernel image in this embodiment are more particularly denoted as u(x, y) and k(x, y), respectively. Projection through the respective upper and lower ultrasonic propagation regions results in the transforming of the input image u(x, y) and the kernel image k(x, y) into their respective Fourier transforms, which are given by $\hat{u}(f_x, f_y)$ and $\hat{k}(f_x, f_y)$.

[0058] The intermediate computational structure computes the product of the two Fourier transforms $\hat{u}(f_x, f_y)$ and $\hat{k}(f_x, f_y)$, with the resulting product being given by $\hat{u}(f_x, f_y)\otimes\hat{k}(f_x, f_y)$. This product is projected through the lower ultrasonic propagation region to obtain at the second ultrasonic transducer array the convolution u*k.

[0059] The various steps (1), (2) and (3) associated with the computation of the convolution u*k in FIG. **2** may be further described as follows, with the Fourier transforms being more particularly referred to as fast Fourier transforms (FFTs) generated by respective top and bottom ultrasonic FFT blocks of the ultrasonic computation engine **200**.

[0060] In step (1), the input image collected by a photodiode array or data bus is projected into the top ultrasonic FFT block and the kernel image collected in a similar manner is projected into the bottom ultrasonic FFT block, to generate the respective FFTs at the intermediate computational structure of the ultrasonic computation engine.

[0061] In step (2), the two FFTs are multiplied in the analog domain in the receive arrays of the intermediate computational structure.

[0062] In step (3), the resulting product image is projected into the bottom ultrasonic FFT block to obtain the convolution of the input image and the kernel image.

[0063] FIGS. **3** and **4** show additional examples of ultrasonic computation engines performing convolution operations in illustrative embodiments.

[0064] Referring now to FIG. **3**, an ultrasonic computation engine **300**, similar to the ultrasonic computation engines **100** and **200** previously described, performs a convolution of an input image with one or more kernel images. The figure shows three different instances of the ultrasonic computation engine **300**, arranged from left to right, with the left-most instance illustrating computation of the Fourier transforms of the respective input image and the kernel image by projection through the respective upper and lower ultrasonic propagation regions to the intermediate computational structure, the middle instance illustrating the multiplication of the two Fourier transforms in the intermediate computational structure, and the right-most instance illustrating the computation of the inverse Fourier transform of the product of the Fourier transforms of the input image and the kernel image by projection from the intermediate computational structure through the lower ultrasonic propagation region. Again, PCBs similar to those utilized in the ultrasonic computation engine **100** are not shown in this figure for simplicity and clarity of illustration.

[0065] In the ultrasonic computation engine **300**, a first ultrasonic transducer array at the top of the stacked structure comprises a transmit AlN ultrasonic transducer array implemented in CMOS combined with a receive AlN ultrasonic transducer array, also implemented in CMOS. The receive AlN ultrasonic transducer array can be formed, for example, by bonding a receive chip to the transmit chip implementing the transmit AlN ultrasonic transducer array. Alternatively, the receive AlN ultrasonic transducer array can be integrated with the transmit AlN ultrasonic transducer array into a single transmit/receive chip. A second ultrasonic transducer array at the bottom of the stacked structure comprises a transmit AlN ultrasonic transducer array implemented in CMOS. Both the first and second ultrasonic transducer arrays in this embodiment are associated with respective absorber layers as shown. The ultrasonic computation engine **300** further comprises the above-noted upper and lower ultrasonic propagation regions, each illustratively comprising a bulk ultrasonic transmission medium such as fused silica sub-regions and a silicon lens arranged as shown, and the above-noted intermediate computational structure which comprises a pair of GHz ultrasonic receiver/multiplier arrays also implemented in CMOS and an interposer/absorber layer between the receiver/multiplier arrays.

[0066] The ultrasonic computation engine **300** first computes FFTs of respective inputs F and G where F is illustratively an input image and G is a kernel image, as illustrated in the right-most instance in FIG. **3**. The FFTs of F and G are multiplied in the analog domain in the intermediate computational structure, as shown in the middle instance, illustratively by pointwise multiplication utilizing multiplier circuits of a type described elsewhere herein.

[0067] Such analog multiplication advantageously enables very fast calculations, limited by the speed of the transistor level current and voltage changes. Direct analog computation avoids the need for analog-to-digital conversion (ADC) of the received signals, followed by digital multiplication, although such digital multiplication can be used in other embodiments. The use of analog components also enables miniaturization of each pixel to maximize the number of pixels for a given chip area.

[0068] The resulting product of the FFTs of F and G is then transmitted using an ultrasonic transducer array of the intermediate computational structure, via the upper ultrasonic propagation region, to the first ultrasonic transducer array at the top of the stacked structure, thereby performing an inverse FFT (IFFT) such that the convolution F*G is produced at the first ultrasonic transducer array. The product in other embodiments could instead be transmitted using an ultrasonic transducer array of the intermediate computational structure, via the lower ultrasonic propagation region, to the second ultrasonic transducer array at the bottom of the stacked structure.

[0069] FIG. **4** illustrates the operation of another ultrasonic computation engine **400**, which is similar to that described above for the ultrasonic computation engine **300**. However, in the ultrasonic computation engine **400**, the second ultrasonic transducer array at the bottom of the stacked structure comprises a transmit AlN ultrasonic transducer array that is integrated in CMOS with a photodiode array. Such an arrangement allows an input image G to be directly captured by the photodiode array and then transmitted by the transmit AlN ultrasonic transducer array into the lower ultrasonic propagation region towards the intermediate computational structure. The photodiode collected photons generate a current which can be used to modulate the amplitude and phase of each AlN transducer.

[0070] A similar integration with a photodiode array could additionally or alternatively be implemented for the first ultrasonic transducer array at the top of the stacked structure, so as to allows an input image F to be directly captured by the photodiode array and then transmitted by the transmit AlN ultrasonic transducer array of the first ultrasonic transducer array into the upper ultrasonic propagation region towards the intermediate computational structure. The operation of the ultrasonic computation engine **400** is otherwise similar to that of the ultrasonic computation engine **300** as previously described.

[0071] FIG. **5** in part (a) thereof shows a more detailed view of a portion of an ultrasonic computation engine **500** and in part (b) thereof a corresponding block diagram, in illustrative embodiments.

[0072] Referring first to part (a), the portion of the ultrasonic computation engine **500** as illustratively comprises an intermediate computational structure arranged between respective upper and lower ultrasonic propagation regions **502-1** and **502-2**. The intermediate computational structure comprises first and second receiver layers **504-1** and **504-2** arranged proximate respective ones of the upper and lower ultrasonic propagation regions **502-1** and **502-2**, and an interposer layer **505** arranged between the first and second receiver layers **504-1** and **504-2**. A plurality of through-silicon vias (TSVs), such as TSVs **507-1** and **507-2**, are configured to connect the first and second receiver layers **504-1** and **504-2** through respective regions **506-1** and **506-2** to the interposer layer **505**. In this embodiment, the receiver layers **504-1** and **504-2** comprise respective AlN ultrasonic transducer arrays implemented in CMOS, and the interposer layer **505** comprises an interposer GHz ultrasonic wave chip with absorbing layers. The interposer layer **505** may also be implemented in CMOS to facilitate interconnection with the receiver layers **504-1** and **504-2**. The absorber layers of the interposer layer **505** illustratively comprise thin film absorbers, such as one or more layers of nickel and copper. The absorber layers prevent the ultrasonic waves propagating through the upper and lower ultrasonic propagation regions **502-1** and **502-2** from interfering with each other. In this embodiment, CMOS circuitry of the type described elsewhere herein is utilized to implement the multiplication of the two FFTs in the intermediate computational structure between the upper and lower ultrasonic propagation regions **502-1** and **502-2**.

[0073] Referring now to part (b), the block diagram illustrates the computations performed in the intermediate computational structure in the portion of the ultrasonic computation engine **500** shown in part (a). As shown, an analog multiplier **510** of the interposer layer **505** receives the Fourier transforms, which are illustratively denoted $\hat{u}(f_x, f_y)$ and $\hat{k}(f_x, f_y)$, and multiplies the Fourier transforms to generate the product $\hat{u}(f_x, f_y) \otimes \hat{k}(f_x, f_y)$. The product is provided by the analog multiplier **510** to drive circuitry **512** of the interposer layer **505** for driving an AlN ultrasonic transducer array to propagate the product back through one of the upper or lower ultrasonic propagation regions **502-1** and **502-2** to compute the inverse Fourier transform of the product, which provides the convolution u*k.

[0074] The particular arrangement illustrated in FIG. **5** for the intermediate computational structure of the ultrasonic computation engine **500** can be similarly implemented in one or more of the other ultrasonic computation engines **100**, **200**, **300** and **400** as previously described. Additional or alternative structures and components can be utilized in implementing the intermediate computational structure in other embodiments.

[0075] The particular arrangements of ultrasonic computation engines illustrated in FIGS. **1** through **5** are presented by way of illustrative example only, and alternative stacked structures with other arrangements of components can be used in other embodiments.

[0076] For example, in some embodiments, ultrasonic waves from a first image can be transmitted from a front-side ultrasonic transducer array and then reflected back from a back-side component to be received and stored by the front-side ultrasonic transducer array. A first set of received signals, illustratively corresponding to the Fourier transform of the first image, can then be stored in the analog domain using a first set of storage capacitors to store received signal values for each pixel. A second image is then similarly transmitted from the front-side ultrasonic transducer array and then reflected back from the back-side component, with a second set of received signals, illustratively corresponding to the Fourier transform of the second image, being stored in the analog domain using a second set of storage capacitors to store received signal values for each pixel. The two Fourier transforms are then multiplied in the analog domain as previously described and the resulting product is then transmitted through the medium and received by the front-side ultrasonic transducer array. In such an embodiment, the opposite side of the device corresponding to the back-side component may comprise one or more layers with varying ultrasonic impedance elements for lensing of the received ultrasonic signals.

[0077] These and numerous other variants of ultrasonic computation engines can be implemented in accordance with the present disclosure. For example, some embodiments disclosed herein can incorporate components, materials, parameters or other aspects of one or more of the ultrasonic devices disclosed in U.S. patent application Ser. No. 17/057,868, entitled "Ultrasonic Fourier Transform Analog Computing Apparatus, Method and Applications," which is commonly assigned herewith and incorporated by reference herein in its entirety.

[0078] As indicated previously, the intermediate computational structure of a given ultrasonic computation engine illustratively comprises intermediate multiplication circuitry including a plurality of multiplier circuits, a plurality of in-phase and quadrature-phase mixer and converter circuits coupled to corresponding outputs of the multiplier circuits, and a plurality of product transmission circuits coupled to corresponding outputs of the mixer and converter circuits, arranged to facilitate the performance of the computations

mentioned above. Examples of such circuitry will now be described with reference to FIGS. **6** through **10**, although as noted above, additional or alternative circuitry can be incorporated into the intermediate computational structure in order to perform other types of computations.

[0079] FIG. **6** shows an example multiplier circuit **600** implemented to perform multiplications at a pixel level in an ultrasonic computation engine in an illustrative embodiment. The multiplier circuit **600** includes transistors M_1, M_2 and M_3, resistors R_{B1} and R_{B2}, two instances of resistor R_{L1}, and first and second bias voltage sources V_{BIAS1} and V_{BIAS2}, arranged as shown. Ultrasonic transducer components corresponding to respective top and bottom AlN receiver layers are denoted AlN_{Top} and AlN_{Bot}.

[0080] A received input signal from pixel transducer AlN_{Top} of the top AlN receiver layer is denoted $V_{AlN_{Top}}$ and is gated through two transmit/receive switches driven by the control signal T/R. If the control signal T/R is "on" (e.g., at a logic high level), its corresponding switches are closed, and the multiplier circuit **600** operates in a mode in which the received input signal $V_{AlN_{Top}}$ is applied to the multiplier portion of the multiplier circuit **600** to generate a product of that signal and a received input signal $V_{AlN_{Bot}}$ from the pixel transducer AlN_{Bot} of the bottom AlN receiver layer. If T/R is "off" (e.g., at a logic low level), its corresponding switches are open, and the switches driven by the complementary control signal $\overline{T/R}$ are closed, such that the multiplier circuit **600** operates in a mode in which an output signal from the driver circuitry of FIG. **9** is applied to transducer AlN_{Top} to drive the product out through the top ultrasonic propagation region to obtain the inverse Fourier transform.

[0081] In the multiplier circuit **600**, the voltages of the applied signal $V_{AlN_{Top}}$ modulate the current through transistors M_1 and M_2 differentially to generate a differential output across the terminals V_{1+}, and V_{1-}. A current tail transistor M_3 is driven by bias voltage source V_{BIAS1} and the signal $V_{Aln_{Bot}}$ from the transducer AlN_{Bot} of the bottom AlN receiver layer. The output impedance of the bias voltage source V_{BIAS1} can be high such that the capacitive output of the AlN receiver layer directly effects the voltage at the M_3 gate. Alternatively, another differential amplifier can be used to drive the gate of M_3, with the differential amplifier being driven by the AlN receiver layer signal and a DC bias.

[0082] The current control in the current tail transistor M_3 and the differential M_1/M_2 control generate the product of the input voltages as the output is proportional to the transconductance g_m which is controlled by the tail current. Other embodiments can utilize alternative multiplier circuits such as, for example, quadrant multipliers, balanced multipliers or other types of multipliers, as will be appreciated by those skilled in the art.

[0083] FIGS. **7** and **8** show respective example in-phase (I) and quadrature-phase (Q) implementations of mixer and DC converter circuits **700** and **800** for converting an output of the multiplier circuit **600** of FIG. **6** to respective I and Q components. The mixer and DC converter circuit **700** of FIG. **7** includes transistors M_4 through M_9, resistor R_{B3}, first and second capacitors, amplifier A_1, bias voltage sources V_{BIAS3} and V_{BIAS4}, and a signal source $V_{sin\ \omega t}$, arranged as shown. The mixer and DC converter circuit **800** of FIG. **8** is configured in a similar manner, but with transistors Mio through M_{15}, and a signal source $V_{cos\ \omega t}$.

[0084] In FIG. **7**, a sin(ωt) component from signal source $V_{sin\ \omega t}$ is multiplied with the input voltage from the multiplier circuit **600** and then rectified using transistor M_9 and averaged using the output capacitor, and then amplified using amplifier A_1 to produce the output in-phase or I component V_{1A}.

[0085] Similarly, in FIG. **8**, a cos(ωt) component from signal source $V_{cos\ \omega t}$ is multiplied with the input voltage from the multiplier circuit **600** and then rectified using transistor M_{15} and averaged using the output capacitor, and then amplified using amplifier A_1 to produce the output quadrature-phase or Q component V_{QA}. The circuits **700** and **800** therefore generate I and Q components of the product from the multiplier circuit **600**.

[0086] Once the I and Q components of the product are generated, and the initial ultrasonic energy from propagation of the input image and kernel image through the respective upper and lower ultrasonic propagation regions is absorbed in the absorber layers, the product is transmitted through one of the upper and lower ultrasonic propagation regions to obtain the inverse Fourier transform as previously described.

[0087] FIG. **9** shows an example product transmission circuit **900** implemented in an ultrasonic computation engine in an illustrative embodiment. The product transmission circuit **900** includes transistors M_{16} through M_{19}, two instances of resistor R_{L2}, and amplifiers A_2 and A_3, arranged as shown, and utilizes bias signals V_{BIAS5} and V_{BIAS6}. The amplifiers A_2 and A_3 receive at inverting inputs thereof the signals V_{1A} and VoA from the respective circuits **700** and **800** of FIGS. **7** and **8**. The amplifiers A_2 and A_3 receive at non-inverting inputs thereof the bias signal V_{BIAS6}.

[0088] The product transmission circuit **900** drives a transducer **902** for transmitting the product signal. This transducer illustratively corresponds to the above-noted transducer AlN_{Top}. The transducer **902** is connected across the output of two common-source amplifiers comprising respective transistors M_{16} and M_{17} such that the difference of the I and Q input signal voltages appears across the transducer **902**. For example, if the input on the left side is given by V_{PI} sin(ωt), and the input on the right side is given by —V_{PQ} cos(ωt), then the voltage across the transducer **902** is given by V_{PI} sin(ωt)+V_{PQ} cos(ωt). This phase shifts the output to generate via the transducer **902** a corresponding acoustic wave with the phase and amplitude information. Transistors M_{18} and M_{19} control the amplitude of the two components and their polarity can be reversed using a variety of circuits that effectively subtract and level shift, as will be appreciated by those skilled in the art.

[0089] FIG. **10** shows example intermediate multiplication circuitry **1000** that incorporates a multiplier circuit such as multiplier circuit **600**, I and Q mixer and DC converter circuits such as circuits **700** and **800**, and a product transmission circuit such as circuit **900**, all implemented in an ultrasonic computation engine in an illustrative embodiment. The intermediate multiplication circuitry **1000** generally operates in a manner similar to that previously described in conjunction with the circuits **600**, **700**, **800** and **900** of FIGS. **6** through **9**. This circuitry is part of what is generally referred to herein as an intermediate computational structure of an ultrasonic computation engine.

[0090] In the circuitry **1000**, received input signals denoted FT(f) and FT(g) from the respective top and bottom AlN receiver layer pixel transducers AlN_{Top} and AlN_{Bot} are multiplied in multiplier **1002**, and the I and Q components of the product are extracted, and then stored on two capacitors. The stored components are amplified in respective

amplifiers, each denoted A, and then mixed again with sin and cos terms and summed in signal combiner **1004** to provide the product that is transmitted via the transducer AlN_{Top} in ultrasonic waves through the top ultrasonic propagation region to obtain the inverse Fourier transform.

[0091] As indicated previously, in other embodiments, the circuitry can be configured such that the product is transmitted via the transducer AlN_{Bot} in ultrasonic waves through the bottom propagation region to obtain the inverse Fourier transform. Numerous other circuitry arrangements are possible in other embodiments.

[0092] Ultrasonic computation engines of the type described above can be implemented in a processing platform that integrates such components with a CNN, DNN or other type of machine learning system to provide significant improvements in computational efficiency and associated classification performance.

[0093] FIG. **11** shows an example of an information processing system **1100** that incorporates a processing platform **1102**. The processing platform **1102** illustratively interacts with input data sources **1105** and controlled devices **1106**. The processing platform **1102** comprises a CNN **1110** that includes an ultrasonic computation engine **1112** and a plurality of stored kernels **1114**. Also included in the processing platform **1102** is control circuitry **1115**, which illustratively utilizes classification outputs of the CNN **1110** to control one or more of the controlled devices **1106**. The processing platform **1102** in some embodiments may be part of a larger device, such as a car, drone or robot, each of which may be considered an "information processing system" as that term is broadly used herein. The processing platform **1102** further comprises a processor **1120**, a memory **1122** and a network interface **1124**.

[0094] The processing platform **1102** in this embodiment implements a computing architecture that utilizes ultrasonic computation engine **1112** to improve the computational efficiency of the CNN **1110**. Although shown as a component of the CNN **1110** in this embodiment, the ultrasonic computation engine **1112** can in other embodiments be implemented at least in part externally to the CNN **1110**. Also, there may be multiple ultrasonic computation engines, rather than a single ultrasonic computation engine as illustrated.

[0095] The CNN **1110** illustratively processes input images which may be received from one or more of the input data sources **1105**. Additionally or alternatively, input images may be generated from a pixel array or other image sensor array integrated with the ultrasonic computation engine in the manner described elsewhere herein. The processing of these and other input images in CNN **1110** illustratively includes various types of image classification operations which classify a particular received image as being an image of a particular type of object or arrangement of multiple objects. In performing such operations, the CNN **1110** applies an input image and a particular kernel image from the stored kernels **1114** to the ultrasonic computation engine **1112** to generate a convolution of the input image and the kernel image in the manner previously described herein. The CNN **1110** similarly utilizes the ultrasonic computation engine **1112** to generate a plurality of convolutions of the input image with respective different ones of a plurality of kernel images from the stored kernels **1114** as respective ones of a plurality of convolution operations of the CNN **1110**.

[0096] The processor **1120** may comprise, for example, a microprocessor, an application-specific integrated circuit (ASIC), a field-programmable gate array (FPGA), a central processing unit (CPU), a tensor processing unit (TPU), a graphics processing unit (GPU), an arithmetic logic unit (ALU), a digital signal processor (DSP), or other similar processing device component, as well as other types and arrangements of processing circuitry, in any combination. At least a portion of the functionality of a computing architecture provided by one or more processing devices as disclosed herein can be implemented using such circuitry.

[0097] In some embodiments, the processor **1120** comprises one or more graphics processor integrated circuits of the processing platform **1102**. Such graphics processor integrated circuits are illustratively implemented in the form of one or more GPUs. Accordingly, in some embodiments, an information processing system implementing a CNN or other type of machine learning system is configured to include a GPU-based processing platform. Such a GPU-based processing platform can be cloud-based and configured to implement a CNN or other type of machine learning system as disclosed herein. Other embodiments can be implemented using similar arrangements of one or more TPUs.

[0098] Numerous other arrangements are possible. For example, in some embodiments, one or more computing architectures and associated neural networks can be implemented on a single device, such as a car, drone or robot, utilizing one or more processors and one or more ultrasonic computation engines of that device. Such embodiments are examples of "on-device" implementations of computing architectures.

[0099] The memory **1122** stores software program code for execution by the processor **1120** in implementing portions of the functionality of the processing platform **1102**. For example, at least portions of the functionality of a computing architecture or other machine learning system component can be implemented using program code stored in memory.

[0100] A given such memory that stores such program code for execution by a corresponding processor is an example of what is more generally referred to herein as a processor-readable storage medium having program code embodied therein, and may comprise, for example, electronic memory such as SRAM, DRAM or other types of random access memory, flash memory, read-only memory (ROM), magnetic memory, optical memory, or other types of storage devices in any combination.

[0101] Articles of manufacture comprising such processor-readable storage media are considered embodiments of the invention. The term "article of manufacture" as used herein should be understood to exclude transitory, propagating signals.

[0102] Other types of computer program products comprising processor-readable storage media can be implemented in other embodiments.

[0103] In addition, illustrative embodiments may be implemented in the form of integrated circuits comprising processing circuitry configured to implement processing operations associated with a computing architecture or its associated machine learning system. For example, at least a portion of a computing architecture as disclosed herein is illustratively implemented in at least one neural network integrated circuit of a processing device of the processing

platform. Such a neural network integrated circuit can be at least partially integrated with one or more ultrasonic computation engines as disclosed herein.

[0104] The network interface **1124** is configured to allow the processing platform **1102** to communicate over one or more networks with other system elements, and may comprise one or more conventional transceivers.

[0105] FIG. **12** shows a more detailed view of an example implementation **1200** of the CNN **1110** of FIG. **11**. It is assumed that the CNN **1110** comprises or is otherwise associated with at least one ultrasonic computation engine **1112** that is configured to perform high-efficiency convolution operations for the CNN **1110** in conjunction with feature learning and associated image classification.

[0106] In the example implementation **1200**, the CNN **1110** receives input **1201**, illustratively a series of input images, and performs feature learning **1202** followed by classification **1204**. In feature learning **1202**, the CNN **1110** includes a first convolution plus rectified linear unit (ReLU) layer **1210** followed by a first pooling layer **1211**, and a second convolution plus ReLU layer **1212** followed by a second pooling layer **1213**, and may include one or more additional layers of similar type, as indicated in the figure. In classification **1204**, the CNN **1110** includes a flatten layer **1214**, a fully-connected layer **1215**, and a softmax layer **1216**. The CNN **1110** in this embodiment is more particularly configured to learn features of input images by performing convolutions with kernel images, and then to utilize those learned features to perform image classification. For example, as illustrated in the figure, the CNN **1110** can learn features of various types of vehicles, so as to be able to classify additional images as each corresponding to a particular type of vehicle, such as a car, truck, van, bicycle, etc.

[0107] It is to be appreciated that the particular arrangement of components and other system elements described above is presented by way of illustrative example only, and numerous alternative embodiments are possible. For example, other embodiments of information processing systems can be configured to implement computing architectures and associated machine learning functionality of the type disclosed herein.

[0108] In other embodiments, different types of neural networks can be used. For example, although some embodiments are implemented using CNNs, other embodiments can comprise additional or alternative neural networks, such as DNNs configured for deep learning, or other types of neural networks implementing other types of machine learning or artificial intelligence algorithms.

[0109] Also, the particular process operations described in illustrative embodiments herein, although illustrated as being performed sequentially, can in other embodiments be performed at least in part in parallel with one another, or in different orders. Also, additional or alternative processing operations can be used in other embodiments.

[0110] As indicated previously, embodiments disclosed herein provide significant advantages over conventional approaches.

[0111] These and other advantages referred to herein are illustratively present in some embodiments, but need not be present in all embodiments. Other embodiments can provide different arrangements of advantages, depending upon their particular implementations.

[0112] The particular configurations as shown in the figures are non-limiting and should be considered illustrative examples only. Numerous other types of ultrasonic computation engines and associated computing architectures can be used in other embodiments. Also, other types of machine learning and/or artificial intelligence arrangements, illustratively implementing other types of neural networks, can be used in other embodiments. Accordingly, illustrative embodiments herein are not limited to use with particular ultrasonic computation engine implementations, or to use with CNNs, DNNs or other particular types of neural networks or machine learning systems.

[0113] An information processing system comprising one or more ultrasonic computation engines as disclosed herein can be configured to support a wide variety of distinct applications, in numerous diverse contexts.

[0114] It is therefore to be appreciated that the particular use cases described herein are examples only, intended to demonstrate utility of illustrative embodiments, and should not be viewed as limiting in any way.

[0115] Automated actions taken based on outputs generated by a machine learning system of the type disclosed herein can include particular actions involving interaction between a processing platform and other related equipment utilized in one or more of the use cases described herein. For example, control signals or other outputs generated by a machine learning system can control one or more components of a related system. In some embodiments, the machine learning system and the components in the related system are implemented on the same processing platform, which may comprise, for example, a car, a drone, a robot or another type of system or device, and in other embodiments the machine learning system and the components of the related system are implemented on separate processing platforms, illustratively interconnected by one or more networks.

[0116] As indicated above, the particular arrangements shown and described in conjunction with the figures are presented by way of illustrative example only, and numerous alternative embodiments are possible. The various embodiments disclosed herein should therefore not be construed as limiting in any way. Numerous alternative arrangements of computing architectures comprising one or more ultrasonic computation engines can be utilized in other embodiments. Those skilled in the art will also recognize that alternative processing operations and associated system configurations can be used in other embodiments.

[0117] It is therefore possible that other embodiments may include additional or alternative system elements, relative to the entities of the illustrative embodiments. Accordingly, the particular system configurations and associated computation implementations can be varied in other embodiments.

[0118] A given processing device or other component of an information processing system as described herein is illustratively configured utilizing a corresponding processing device comprising a processor coupled to a memory. The processor executes software program code stored in the memory in order to control the performance of processing operations and other functionality. The processing device also comprises a network interface that supports communication over one or more networks.

[0119] The processor may comprise, for example, a microprocessor, an ASIC, an FPGA, a CPU, a TPU, a GPU, an ALU, a DSP, or other similar processing device component, as well as other types and arrangements of processing circuitry, in any combination. For example, at least a portion

of the functionality of a computing architecture and/or an associated machine learning system provided by one or more processing devices as disclosed herein can be implemented using such circuitry.

[0120] The memory stores software program code for execution by the processor in implementing portions of the functionality of the processing device. A given such memory that stores such program code for execution by a corresponding processor is an example of what is more generally referred to herein as a processor-readable storage medium having program code embodied therein, and may comprise, for example, electronic memory such as SRAM, DRAM or other types of random access memory, ROM, flash memory, magnetic memory, optical memory, or other types of storage devices in any combination.

[0121] As mentioned previously, articles of manufacture comprising such processor-readable storage media are considered embodiments of the invention. The term "article of manufacture" as used herein should be understood to exclude transitory, propagating signals. Other types of computer program products comprising processor-readable storage media can be implemented in other embodiments.

[0122] In addition, embodiments of the invention may be implemented in the form of integrated circuits comprising processing circuitry configured to implement processing operations associated with implementation of a computing architecture and/or associated machine learning system as disclosed.

[0123] An information processing system as disclosed herein may be implemented using one or more processing platforms, or portions thereof.

[0124] For example, one illustrative embodiment of a processing platform that may be used to implement at least a portion of an information processing system comprises cloud infrastructure including virtual machines implemented using a hypervisor that runs on physical infrastructure. Such virtual machines may comprise respective processing devices that communicate with one another over one or more networks.

[0125] The cloud infrastructure in such an embodiment may further comprise one or more sets of applications running on respective ones of the virtual machines under the control of the hypervisor. It is also possible to use multiple hypervisors each providing a set of virtual machines using at least one underlying physical machine. Different sets of virtual machines provided by one or more hypervisors may be utilized in configuring multiple instances of various components of the information processing system.

[0126] Another illustrative embodiment of a processing platform that may be used to implement at least a portion of an information processing system as disclosed herein comprises a plurality of processing devices which communicate with one another over at least one network. Each processing device of the processing platform is assumed to comprise a processor coupled to a memory. A given such network can illustratively include, for example, a global computer network such as the Internet, a WAN, a LAN, a satellite network, a telephone or cable network, a cellular network such as a 4G or 5G network, a wireless network implemented using a wireless protocol such as Bluetooth, WiFi or WiMAX, or various portions or combinations of these and other types of communication networks.

[0127] Again, these particular processing platforms are presented by way of example only, and an information processing system may include additional or alternative processing platforms, as well as numerous distinct processing platforms in any combination, with each such platform comprising one or more computers, servers, storage devices or other processing devices.

[0128] A given processing platform implementing a computing architecture comprising at least one ultrasonic computation engine as disclosed herein can be implemented or otherwise supported by cloud infrastructure or other types of virtualization infrastructure.

[0129] It should therefore be understood that in other embodiments different arrangements of additional or alternative elements may be used. At least a subset of these elements may be collectively implemented on a common processing platform, or each such element may be implemented on a separate processing platform.

[0130] Also, numerous other arrangements of computers, servers, storage devices or other components are possible in an information processing system. Such components can communicate with other elements of the information processing system over any type of network or other communication media.

[0131] It should again be emphasized that the embodiments of the invention as described herein are intended to be illustrative only. Other embodiments of the invention can be implemented utilizing a wide variety of different types and arrangements of ultrasonic computation engines, computing architectures, information processing systems, processing devices, machine learning systems and additional or alternative components, than those utilized in the particular illustrative embodiments described herein, and in numerous alternative processing contexts. In addition, the particular assumptions made herein in the context of describing certain embodiments need not apply in other embodiments. These and numerous other alternative embodiments will be readily apparent to those skilled in the art.

What is claimed is:

1. An ultrasonic computation apparatus comprising:

first and second ultrasonic transducer arrays arranged on opposing ends of the ultrasonic computation apparatus;

first and second ultrasonic propagation regions arranged between the first and second ultrasonic transducer arrays and proximate respective ones of the first and second ultrasonic transducer arrays; and

an intermediate computational structure arranged between the first and second ultrasonic propagation regions;

wherein respective first and second input signals applied to respective ones of the first and second ultrasonic transducer arrays cause propagation of corresponding ultrasonic waves through the respective first and second ultrasonic propagation regions towards the intermediate computational structure; and

wherein the intermediate computational structure is configured to receive the propagating ultrasonic waves from the respective first and second ultrasonic propagation regions and to generate from the received propagating ultrasonic waves an additional signal that is a function of the first and second signals.

2. The apparatus of claim **1** wherein the intermediate computational structure is configured to generate the additional signal at least in part as a product of transformed versions of the respective first and second signals as represented by the corresponding received propagating ultrasonic waves.

3. The apparatus of claim 1 wherein the first ultrasonic transducer array and the first ultrasonic propagation region are configured such that the propagation of ultrasonic waves corresponding to the first signal through the first ultrasonic propagation region causes a Fourier transform of the first signal to be received at a first side of the intermediate computational structure proximate the first ultrasonic propagation region.

4. The apparatus of claim 3 wherein the second ultrasonic transducer array and the second ultrasonic propagation region are configured such that the propagation of ultrasonic waves corresponding to the second signal through the second ultrasonic propagation region causes a Fourier transform of the second signal to be received at a second side of the intermediate computational structure proximate the second ultrasonic propagation region.

5. The apparatus of claim 4 wherein the intermediate computational structure is configured to generate the additional signal as a product of the Fourier transforms of the respective first and second signals.

6. The apparatus of claim 1 wherein the generation of the additional signal in the intermediate computational structure causes propagation of additional ultrasonic waves through one of the first and second ultrasonic propagation regions towards one of the first and second ultrasonic transducer arrays such that a function of the additional signal is received at said one of the first and second ultrasonic transducer arrays.

7. The apparatus of claim 6 wherein the additional signal generated in the intermediate computational structure comprises a product of Fourier transforms of respective ones of the first and second signals and the function of the additional signal received at said one of the first and second ultrasonic transducer arrays comprises an inverse Fourier transform of the product of the Fourier transforms, the inverse Fourier transform of the product of the Fourier transforms being a convolution of the first and second signals.

8. The apparatus of claim 1 wherein the first signal comprises an input image applied to the first ultrasonic transducer array and the second signal comprises a kernel image applied to the second ultrasonic transducer array.

9. The apparatus of claim 8 wherein the kernel image is one of a plurality of kernel images of a convolutional neural network and the apparatus is configured to generate a plurality of convolutions of the input image with respective ones of the kernel images as respective convolution operations of the convolutional neural network.

10. The apparatus of claim 1 wherein at least one of the first and second ultrasonic transducer arrays comprises a transmit/receive ultrasonic transducer array that is configured to transmit ultrasonic waves to its corresponding ultrasonic propagation region and to receive ultrasonic waves from its corresponding ultrasonic propagation region.

11. The apparatus of claim 1 wherein at least one of the first and second ultrasonic transducer arrays is adapted to receive a corresponding one of the first and second signals via an image sensor array that is coupled to the ultrasonic transducer array and is configured to generate the corresponding one of the first and second signals as an image signal that is applied from the image sensor array to the ultrasonic transducer array.

12. The apparatus of claim 1 wherein at least one of the ultrasonic propagation regions comprises:

a first sub-region adjacent its corresponding one of the first and second ultrasonic transducer arrays;

a second sub-region adjacent the intermediate computational structure; and

at least one lens arranged between the first and second sub-regions.

13. The apparatus of claim 1 wherein the intermediate computational structure comprises intermediate multiplication circuitry comprising:

a plurality of multiplier circuits;

a plurality of in-phase and quadrature-phase mixer and converter circuits coupled to corresponding outputs of the multiplier circuits; and

a plurality of product transmission circuits coupled to corresponding outputs of the mixer and converter circuits.

14. The apparatus of claim 1 wherein the intermediate computational structure comprises:

first and second receiver layers arranged proximate respective ones of the first and second ultrasonic propagation regions; and

an interposer layer arranged between the first and second receiver layers;

wherein a plurality of vias are configured to connect the first and second receiver layers through the interposer layer.

15. An apparatus comprising:

at least one processing device comprising a processor coupled to a memory; and

one or more ultrasonic computation engines each coupled to the at least one processing device;

the at least one processing device being configured to implement a machine learning system;

the one or more ultrasonic computation engines each being configured to perform computations for the machine learning system;

wherein each of the one or more ultrasonic computation engines comprises:

first and second ultrasonic transducer arrays arranged on opposing ends of the ultrasonic computation engine and adapted to receive respective first and second signals;

first and second ultrasonic propagation regions arranged between the first and second ultrasonic transducer arrays and proximate respective ones of the first and second ultrasonic transducer arrays; and

an intermediate computational structure arranged between the first and second ultrasonic propagation regions;

wherein the intermediate computational structure is configured to perform at least a portion of the computations for the machine learning system, including generating an additional signal that is a function of the first and second signals.

16. The apparatus of claim 15 wherein the machine learning system comprises a convolutional neural network and one or more of the computations each comprise a convolution operation of the convolutional neural network.

17. The apparatus of claim 16 wherein the first signal comprises an input image applied to the first ultrasonic transducer array and the second signal comprises a kernel image applied to the second ultrasonic transducer array.

18. The apparatus of claim 17 wherein the kernel image is one of a plurality of kernel images of the convolutional neural network and a given one of the one or more ultrasonic

computation engines is configured to generate a plurality of convolutions of the input image with respective different ones of the kernel images as respective ones of a plurality of convolution operations of the convolutional neural network.

19. A method comprising:

applying first and second input signals to respective first and second ultrasonic transducer arrays arranged on opposing ends of an ultrasonic computation engine, the ultrasonic computation engine further comprising first and second ultrasonic propagation regions arranged between the first and second ultrasonic transducer arrays and proximate respective ones of the first and second ultrasonic transducer arrays, and an intermediate computational structure arranged between the first and second ultrasonic propagation regions;

propagating ultrasonic waves corresponding to respective ones of the first and second input signals from respective ones of the first and second ultrasonic transducer arrays through the respective first and second ultrasonic propagation regions towards the intermediate computational structure;

receiving in the intermediate computational structure the propagating ultrasonic waves from the respective first and second ultrasonic propagation regions; and

generating in the intermediate computational structure, based at least in part on the received propagating ultrasonic waves, an additional signal that is a function of the first and second signals.

20. The method of claim **19** wherein generating the additional signal in the intermediate computational structure comprises generating the additional signal at least in part as a product of transformed versions of the respective first and second signals as represented by the corresponding received propagating ultrasonic waves.

* * * * *