



Ultrasound Signal Processing with GPUs – Introduction to Parallel Programming

Ultrasound Hardware & Software



License / Attribution



 Materials for the short-course "Ultrasound Signal Processing with GPUs – Introduction to Parallel Programming" are licensed by us4us Ltd. the IPPT PAN under the <u>Creative Commons Attribution-NonCommercial 4.0 International License.</u>

- Some slides and examples are borrowed from the course "The GPU Teaching Kit" that is licensed by NVIDIA and the University of Illinois under the <u>Creative Commons Attribution-</u> NonCommercial 4.0 International License.
 - All the borrowed slides are marked with



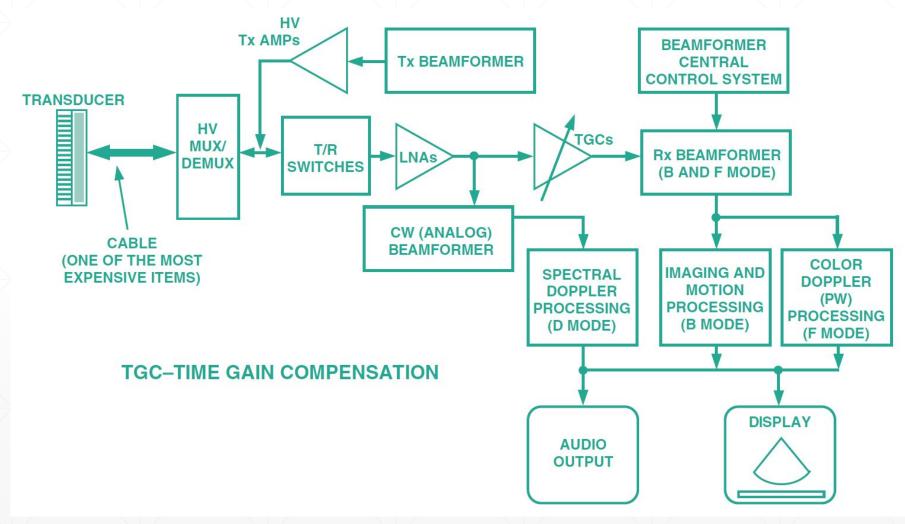


2

Ultrasound Systems



Ultrasound Scanner



Source: Eberhard Brunner, How Ultrasound System Considerations Influence Front-End Component Choice, Analog Dialogue 36-03 (2002)

4

What we need Ultrasound Research Systems for?

Applications

- Introducing new ultrasound modalities in Biomed, NDT (on-line/off-line)
 - Raw channel data acquisition (RF or I/Q)
- Developing Demo systems (on-line)
- Testing/evaluating new probes

Processing

- Off-line acquire first, process later off-line
- On-line acquire and process in real-time
 - high-speed data transfer & computing resources are needed!

Features ...

- Programmable TX/RX
- High-level SDK (support for Matlab, Python)
- High-performance processing for real-time

5

Ultrasound Research Systems





	SARUS	ULA-OP 256	UARP	SonixTouch	Verasonics	us4R-lite™	us4R™
					(Vantage 256)		
Channels	Up to 1024 Tx/Rx	Up to 256 Tx/Rx	Up to 256 Tx/Rx	128 Tx/Rx	256 Tx/Rx	256Tx / 64Rx	1024Tx / 256Rx
Tx Voltage	Up to 200 V_{pp}	Up to 200 V_{pp}	Up to 200 V_{pp}	Up to 50 V_{pp}	3 to 190 V_{pp}	Up to 180V _{pp}	Up to 180V _{pp}
Tx Frequency	1 to 30 MHz	1 to 20 MHz	0.5 to 15 MHz	1 to 20 MHz	0.5 to 20 MHz	1 to 30MHz	1 to 30MHz
					(standard config.)		
Тх Туре	Linear	Linear	5-Level	3-Level	3-Level	3-Level	3-Level
ADC	70 MHz @ 12 bits	78 MHz @ 12 bits	programmable	80 MHz @ 10 bits/	programmable	65 MHz @ 14-bits	65 MHz @ 14-bits
	programmable	programmable	sampling rate up to	40 MHz @ 12 bits	sampling rate up to		or 80 MHz @ 12-bits
	downsampling	downsampling	80 MHz @ 12 bits		62.5 MHz @ 14 bits		00 W112 @ 12-51t3
	with filtration						
RAM Buffer	128 GB	80 GB	16 GB	16 GB	16 GB	8 GB	32 GB
Connection to PC	sixty-four 1Gb/s	USB 3.0	PCIe 3.0	USB 2.0	PCIe 3.0 x8	Thunderbolt-3	PCIe 3.0 x16
	Ethernet links					or PCle 3.0 x8	or Dual PCle 3.0 x16
	coupled through					Pole 3.0 Xo	Dual Pole 3.0 x 10
	four 10Gb/s						
	optical links						
Real-time PROCESSING:	FPGA	FPGA / DSP	FPGA	FPGA	FPGA / CPU	CPU / GPU	CPU / GPU

Source: E. Boni, A. C. H. Yu, S. Freear, J. A. Jensen and P. Tortoli, "Ultrasound Open Platforms for Next-Generation Imaging Technique Development," in *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 65, no. 7, pp. 1078-1092, July 2018, doi: 10.1109/TUFFC.2018.2844560

us4R-lite™



us4R™



- □ A low-cost, portable ultrasound platform for research & educational applications
- □ 256 TX / 64 RX optimized for Plane-Wave
- High-speed data interfaces:
 - ☐ Thunderbolt-3 (2.5GB/s)
 - □ PCle gen3 x8 (6GB/s)
- □ GPU-ready using external PC/notebook with GPU or optional embedded NVIDIA Jetson TX2
- □ ARRUS™ SDK for Windows / Linux
 - open-source!
 - optimized for real-time data acquisition and software processing.

- ☐ High-performance ultrasound research platform
- ☐ Up to 1024 TX / 256 RX (support for 2D/3D)
- ☐ High-speed data interfaces:
 - □ single PCle gen3 x16 (12GB/s)
 - □ dual PCle gen3 x16 (24GB/s)
- Versatile possible integration with various hardware:
 - ☐ GPU and multi-GPU (Nvidia, AMD, Intel)
 - □ PCIe FPGA Accelerator cards (Altera, Xilinx)
- ARRUS™ SDK for Windows / Linux
 - open-source!
 - optimized for real-time data acquisition and software processing.

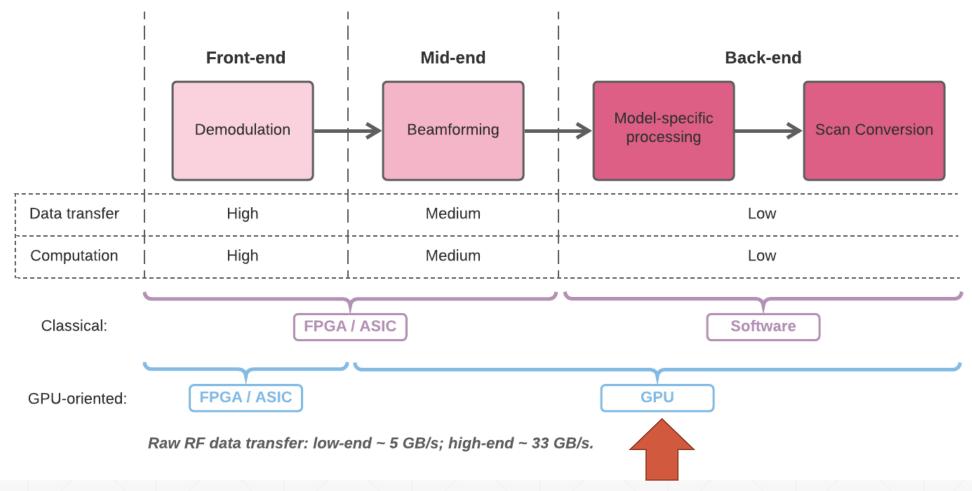
7

Software Defined Ultrasound

Systems Architecture



Ultrasound processing chain



SOFTWARE BEAMFORMING

9

Software-Defined Ultrasound (SDU)

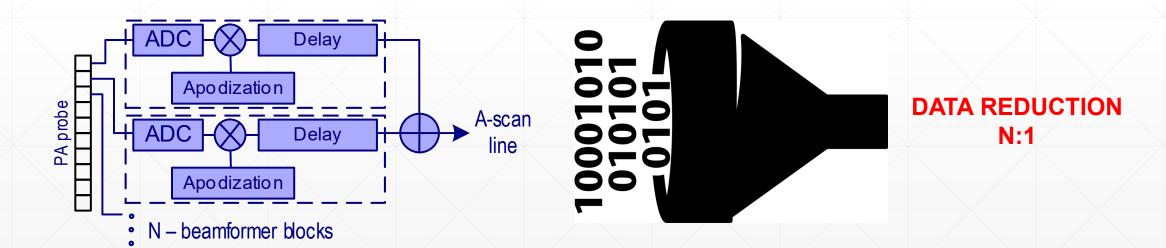
- Software-Defined Ultrasound is a new paradigm of applying full software processing to raw ultrasound echo signals (RF or I/Q).
- Instead of a fixed (usually hardware-based) processing pipeline, we can have a fully programmable and extensible software framework.
- Thanks to the high-performance GPU technology, now, we can implement advanced parallel processing algorithms in real-time.
- us4us® ultrasound platforms have been designed to provide real-time streaming of raw RF data for Software-Defined Ultrasound processing.
- Ultrafast acquisition technology with parallel processing enables implementation of novel multi-modal technologies.

Imaging of Natural Waves · Electromechanical Wave Imaging **Ultrafast Doppler Imaging** Pulse Wave Fuses color flow imaging and PW Doppler Cardiovascular applications · High-sensitivity Doppler imaging **Shear Wave Elastography** · Tumor vascularization imaging · Real-Time and Quantitative Imaging · Wide clinical applications **Tissue Motion Blood Motion Ultrafast Ultrasonic Imaging** Reaching the physical speed limits of Ultrasound Acquisition at a range of kHz frames per second **Micro Bubbles** Neurovascular coupling Functional US (fUS) imaging **Ultrafast Contrast Imaging** of the brain Molecular imaging · Novel imaging tool in neuroscience High contrast · fUS imaging on awake animals Transfontanellar and peroperative fUS

Adopted from: M. Tanter, M. Fink, Ultrafast Imaging in Biomedical Ultrasound, IEEE TUFFC 61(1):102-119, DOI: 10.1109/TUFFC.2014.6689779, 2014

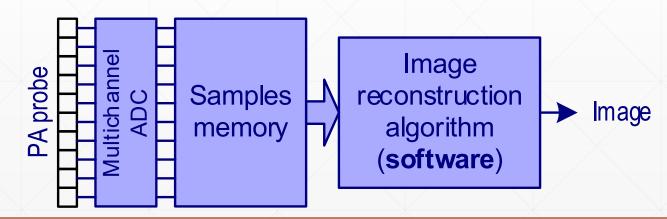
Classical approach – Beamforming

- A classical Beamforming approach
 - Fixed hardware-based image reconstruction
- Line-by-line acquisition and processing
- N:1 data reduction (N signals are delayed and summed together)



Software Defined Ultrasound – features

- A fully programmable TX/RX
- A fully software-programmable signal processing chain
- Direct access and processing of raw channel data / ultrasound echoes
- Imaging based on the Synthetic Aperture Focusing Technique:
 - Depth-independent imaging
 - Ultrafast methods (1000s FPS v.s. 10s FPS for classical Beamforming)
 - Advanced ultrasound modalities/algorithms (e.g. Doppler vector flow)



us4us® Software Architecture

Application

us4us ARRUS SDK*

(C/C++, Matlab, Python)

Parallel Processing API (CUDA, OpenCL, OneAPI, etc.)



us4R™



us4R-lite™

VPU, CPU **GPU** SoC,

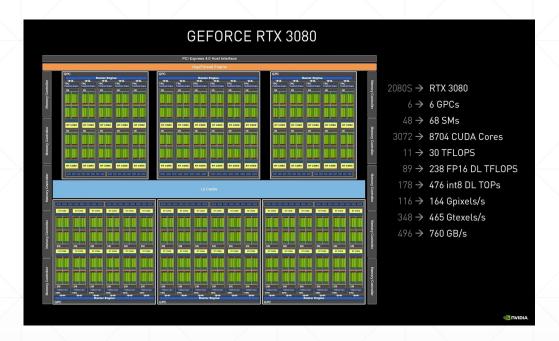
FPGA

*Windows & Linux support

GPU processing



NVIDIA GPU ecosystem



THE JETSON FAMILY

From AI at the Edge to Autonomous Machines



JETSON NANO 5 - 10W 0.5 TFLOPS (FP16) 45mm x 70mm \$129



JETSON TX2 Series (TX2, TX2 4GB, TX2i*) 7.5 - 15W* 1.3 TFLOPS (FP16) 50mm x 87mm Starting at \$249



JETSON AGX XAVIER Series (AGX Xavier 8GB, AGX Xavier)

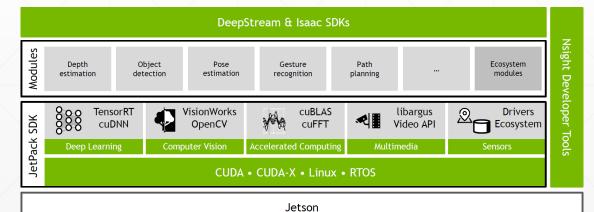
5.5 - 11 TFLOPS (FP16) 20 - 32 TOPS (INT8) 100mm x 87mm Starting at \$599

Al at the edge — Fully autonomous machines

Multiple devices - Same software

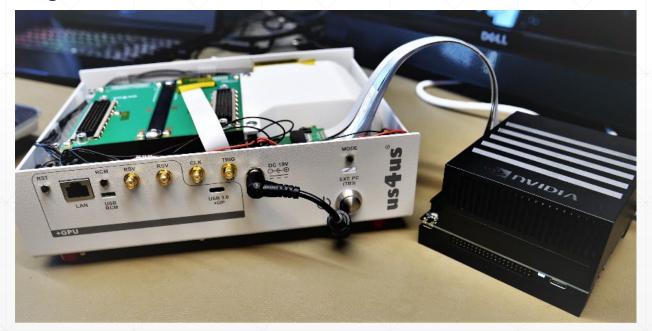
Listed prices are for 1000u+ | Full specs at developer.nvidia.com/jetson

*i = for industrial environments



us4R-lite™ with NVIDIA Xavier

- We have executed a feasibility study of integrating us4R-lite[™] platform with NVIDIA Jetson AGX Xavier[™].
 - The Jetson AGX Xavier module delivers up to 32 TOPS of AI performance and benefits from NVIDIA's rich set of AI tools and workflows, enabling developers to train and deploy neural networks quickly.
- We envision integrating the Xavier in the next revision of the us4R-lite™ systems.



NVIDIA Clara AGX

- An embedded AI computer and software developer framework for medical devices
- Designed to boost development of Software-Defined Ultrasound solutions
- Built for real-time AI and advanced signal processing on highthroughput streaming data

Find out more Clara AGX for healthcare here.



us4R-lite™ and Clara AGX integration



- We have now successfully integrated NVIDIA Clara AGX with our smaller system in a laboratory setting.
- Our immediate plans include integration of low-level software to enable integration with NVIDIA machine learning libraries.
- In future, we plan to fully integrate Clara AGX with both the us4R[™] and the us4Rlite [™] and make these available to our clients for purchase alongside corresponding software.

IUS GPU short-course 2022 18

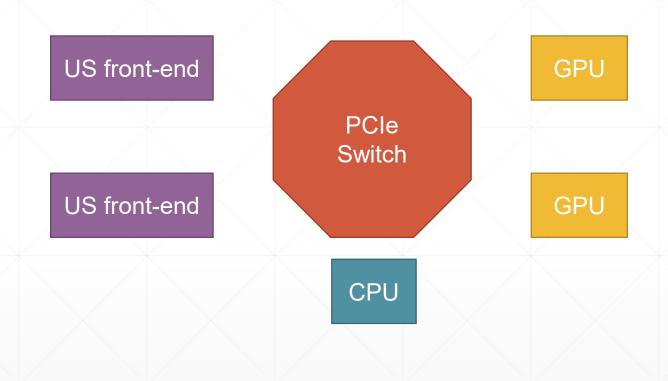


Software Defined Ultrasound

Applications



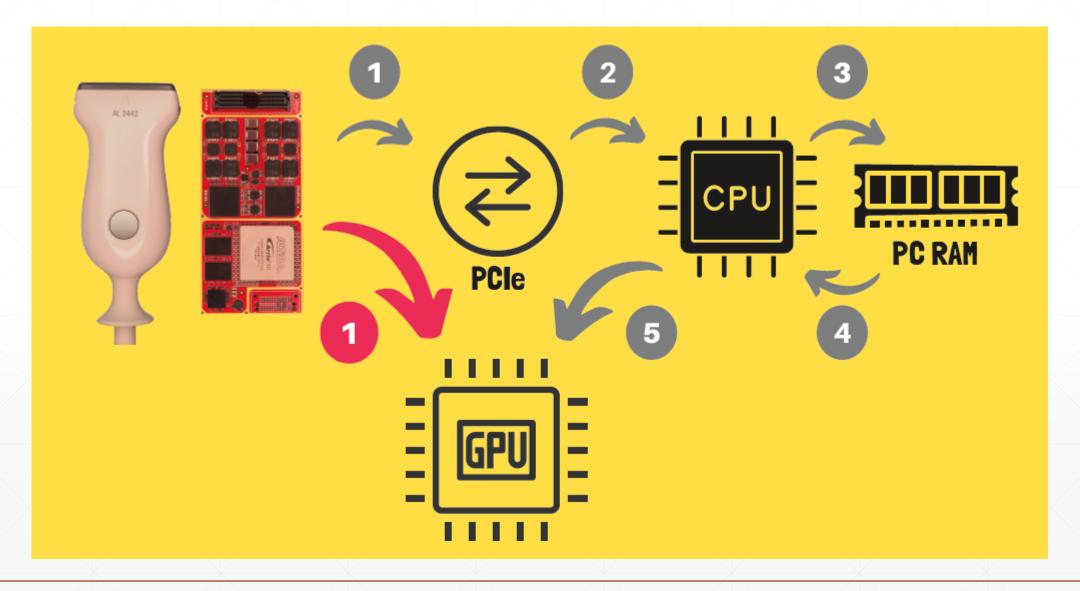
Data Flow



ADC FPGA Buf RAM CPU RAM GPU RAM

> PCle > > PCle >

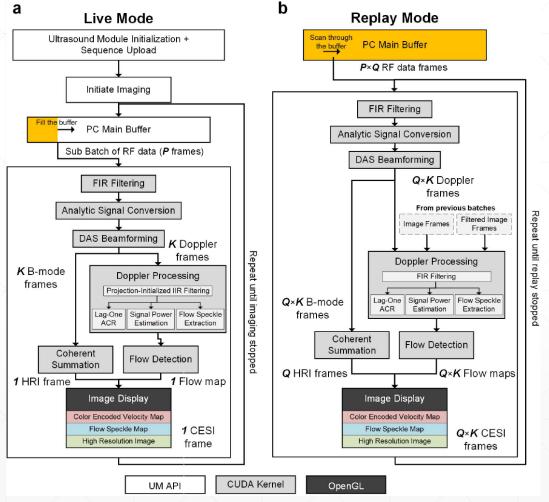
GPUDirect RDMA



IUS GPU short-course 2022 21

Advanced ultrafast acquisition and GPU processing





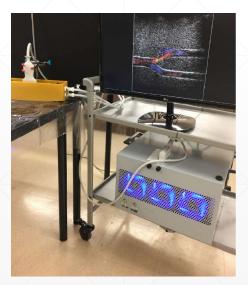
Source: B.Y.S. Yiu, M. Walczak, M. Lewandowski, A.C.H. Yu, Live Ultrasound Color-Encoded Speckle Imaging Platform for Real-Time Complex Flow Visualization In Vivo. IEEE TUFFC, 2019 Apr;66(4):656-668. doi: 10.1109/TUFFC.2019.2892731

On a single Nvidia GTX-1080 card!

Scenario (PRF)	Imaging Depth	Center Freq.	Sampling Freq.	Samples /Channel	Data/s
Superficial (10 kHz)	2 cm	15 MHz	50 MHz	1299	4.6 GB/s*
Carotid (10 kHz)	4 cm	6 MHz	25 MHz	1299	4.6 GB/s*
Kidney (5 kHz)	8 cm	10 MHz	25 MHz	2598	4.6 GB/s*
Heart (4 kHz)	18 cm	3 MHz	25 MHz	5845	4.3 GB/s [#]

^{*}Assuming the use of a 192-element linear array and 2 bytes per data sample

[#] Assuming the use of a 96-element phased array and 2 bytes per data sample

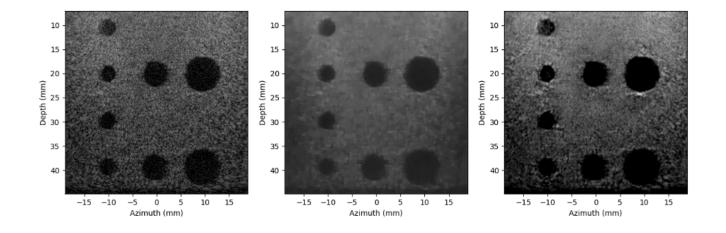


USPlatform (us4us, Poland)

22

us4R-lite™ with NVIDIA Clara AGX for Deep-Learning

- After successfully integrating Clara AGX with the us4R-lite system, we performed a test-case of real-time imaging using a neural network for despeckling and image enhancement.
 - Ultrasound raw data acquisition transfer rate: 3.2 GB/s
 - Signal processing throughput for 64 PWI: 67 fps // NN Despeckling: 57 fps // MimickNet post-processing: 57 fps



LOOK:

- IUS-2021 presentation: "Integration of Ultrasound Research System with AI Workstation NVIDIA Clara AGX"
- NVIDIA Clara AGX us4R-lite Ultrasound Container >> https://ngc.nvidia.com/catalog/containers/nvidia:clara-agx:agx-us4us-ultrasound

Implementation based on Dongwoon Hyun et al., Beamforming and Speckle Reduction Using Neural Networks, IEEE TUFFC, doi:10.1109/TUFFC. 2019.2903795.

And NOW ...

- You almost made it
- There are another exciting Exercises:
 - CUDA streams & Processing

- Probably, you run out of coffee by now ...
- We have a few case-studies (<u>believe me, they are worth seeing</u>):
 - Doppler processing
 - Color-encoded speckle imaging platform for real-time complex flow visualization in-vivo