High-throughput Cellular Imaging with High-Speed Asymmetric-Detection Time-Stretch Optical Microscopy under FPGA Platform

Ho-Cheung Ng*†, Maolin Wang*, Bob M.F. Chung‡, B. Sharat Chandra Varma§,
Manish Kumar Jaiswal*, Sam M.H. Ho*, Kevin K. Tsia*, Ho Cheung Shum‡ and Hayden Kwok-Hay So*

*Department of Electrical & Electronic Engineering, The University of Hong Kong, Hong Kong
†Department of Computing, Imperial College London, UK
‡Department of Mechanical Engineering, The University of Hong Kong, Hong Kong
§Department of Avionics, Indian Institute of Space Science and Technology, India

*{hcng, mlwang, manishkj, mhho, tsia, hso}@eee.hku.hk,†h.ng16@imperial.ac.uk,
‡{c3m8f, ashum}@hku.hk, §sharatcvarma@iist.ac.in

***Abstract*—Asymmetric-Detection Time-Stretch Optical Microscopy (ATOM) is a recently emerged technology that provides ultra-fast cell imaging with a frame rate up to MHz – orders-of-magnitude higher than any classical imaging systems.**

However, existing measuring instruments are unable to fully exploit the capability of ATOM. For example, the volume of imaging data-set of ATOM quickly increases beyond the capacity of available onboard buffer of a modern high-speed oscilloscope.

This paper presents an open source, FPGA-based solution which serves as a dual role of collecting low-level signals from ATOM frontend as well as processing and transferring data to backing store. Optical signals are sampled by a high-speed analog-to-digital converter and the resulting values are collected by an FPGA. The quantized values received are then further processed and divided into four segments for subsequent data transfer with $10\,\mathrm{Gbit}$ Ethernet. Four computing units are attached to these channels with direct connection in order to reliably receive the data for post-processing.

Experiments show that, with decent quality images for single-cell analysis, the proposed system can store $10\times$ more dataset than existing high-end oscilloscope. With $8\times$ decrease in equipment cost, the proposed FPGA-based system will definitely be beneficial for many bioimaging applications with ATOM technology such as rare cancer cell imaging and identification.

I. Introduction

High-throughput micro-particle or cellular imaging and screening have been commonly used in life science research and medical diagnosis. These applications include rare cancer cells detection, aberrant cells screening or emulsion droplets synthesis, which require highly accurate characterization from a massive population of samples (>100,000 particles) [1].

However, existing choice of biomedical imaging tools is limited by the trade-off between throughput and accuracy. Current state-of-the-art photo flow cytometers reveal the imaging details of each cell at the expense of screening throughput, while classical flow cytometers can provide multiple characteristics per cell at high throughput by forgoing spatial information. Such restriction is usually the bottleneck for single-cell analysis with a large sample of cells [2].

Recently, ATOM has emerged as a complementary bioimaging tool that allows ultrafast single-cell imaging with the frame rate or line scan rate as high as MHz, which is comparable to conventional non-imaging flow cytometry [2], [3]. More importantly, ATOM is capable to deliver high-throughput label-free imaging with excellent contrast to reveal multiple parameters of each cell such as size, irregularity or even sub-cellular structure at high-speed data rate.

To fully exploit the functionality of ATOM in real-life applications, a powerful backing store is also required to retrieve such high-speed and high-amount of imaging data. Current state-of-the-art device and technology, such as any CPU-based computing node or high-end oscilloscopes, are unable to deliver excellent storage performance due to their insufficient memory capacity.

In this regard, we present an open source, FPGA-based solution that provides the ability to process and transfer the optical data from ATOM frontend to back-end storage via $10\,\mathrm{Gbit}$ Ethernet connections. It highlights the unique capability of FPGAs in processing low-level signal from the photo detector at line rate as well as alleviating the storage bottleneck by delivering the raw data to multiple computing nodes with the utilization of extensive IO resources on FPGA.

In the particular case of the proposed framework, objects of interest such as microemulsion droplets or cancer cells are flowed through a microfluidic channel where images are captured by ATOM frontend. The optical signal retrieved is then converted electronically by a photo-detector and the output is subsequently sampled by a 8-bit $4\,\mathrm{GSamples/s}$ analog-digital converter. The quantized data is collected by the FPGA where the low-level signal is further processed and sent out to four $10\,\mathrm{Gbit}$ Ethernet ports. Multiple computing nodes are attached to these Ethernet ports so as to gather the sampled data for image reconstruction thereafter.

As such, we consider the main contribution of this work rests on the following aspects:

- We present an FPGA-based solution that has distinguished advantages in data acquisition at line rate where the storage capacity for a continuous data stream is $10\times$

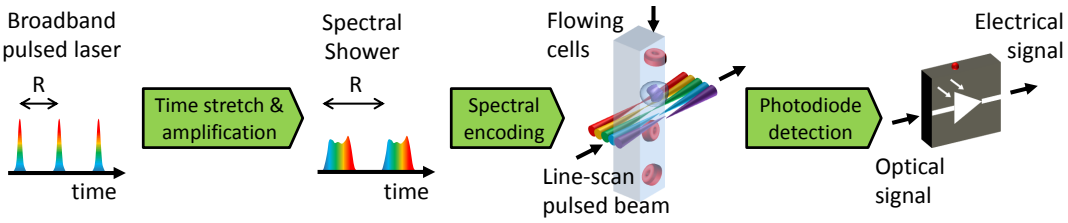


Fig. 1: Simplified block diagram of the ATOM imaging system. The samples are flowed through a microfluidic channel which are imaged by a pulse of spectrally encoded shower.

(can be further enhanced to $20\times$) more than that of real-time oscilloscopes.

- The proposed FPGA framework is capable of delivering high quality images which are more than sufficient for subsequent bioimaging analysis.
- We present a case that an open-source, FPGA-based solution can lower the equipment cost by $8\times$ for data retrieval with ATOM technology.

In the next section, we describe the underlying principle of the ATOM imaging system. Then we elaborate the proposed FPGA system in Section III and discuss the experimental results in Section IV. We make conclusions in Section V.

II. Background

The overall functional goal of the proposed system is to provide intensive storage capacity for the continuous image stream from ATOM frontend. There exist similar acquisition systems such as RASMUS [4] or PEN-DAE [5] which also target at medical diagnosis. However, the proposed system is unique since it specifically aims at high-throughput imaging with ATOM technology. Depending on the experimental setup, these images can be cells, microemulsion droplets or algae. This section describes and elaborates the background about the ATOM imaging system. The processing steps to produce the images from the optical frontend are also discussed. For detailed information about ATOM, please refer to [2].

A. ATOM Imaging System

Figure 1 displays a simplified block diagram of the ATOM imaging system. The basic principle of ATOM lies in ultra-fast retrieval of image information encoded on the spectrum of a broadband laser pulse by converting it into serial temporal data format in real time.

To begin with, a laser source repeatedly generates pulses at a frequency of R. Each laser pulse is then passed through an off-the-shelf dispersive fiber to realise wavelength-to-time mapping with extraordinary large group velocity dispersion (GVD). This process is called time-stretch. To compensate the intrinsic loss of optical power during time-stretching process, inline optical amplifiers are applied to increase the signal-to-noise ratio.

The time-stretched laser beam is then spatially diffracted into a one-dimensional spectral shower (i.e. rainbow) with a spectral encoder. This spectral shower, which is orthogonal to the microfluidic channel, is then applied to scan the flowing samples such as cells or droplets, forming a line scan. Basically, the information of the samples are encoded by mapping the spatial coordinates onto different wavelengths of the spectrum.

As the laser pulse is emitted regularly with a constant frequency and the samples are flowed at a linear speed in the microfluidic channel, the image of an entire cell can be formed when each line scan captures the spatial information of the samples at different positions. The repeated time-stretched waveforms are then fed to the photo detector which are consecutively converted into electrical signals.

B. Image Formation

The output from the photo detector is sampled by a high-speed ADC and the resulting data stream is collected by the FPGA. Then the FPGA processes and directs the received data to the backing store via $10\,\mathrm{Gbit}$ Ethernet channels for subsequent image processing.

The image formation process is normally initiated after the completion of data acquisition. It begins with realignment of the input data stream into segments which correspond to each line of the original image according to the input laser repetition rate R. Once aligned, the resulting image can be formed and objects of interest such as cells can be located by running different detection algorithms.

III. System Organization and Design Methodology

This section details the design methodology and implementation of the proposed FPGA-based data acquisition system. As shown in Figure 2, our current hardware system is based on ROACH-2 platform developed by CASPER [6]. It runs with BORPH operating system that enables run-time controls over the FPGA [7]. A 8-bit $4\,\mathrm{GSamples/s}$ ADC serves as the interface between the ATOM frontend and the onboard FPGA. The data collected are sent over to the storage nodes via $10\,\mathrm{Gbit}$ Ethernet channels.

A. Data Acquisition System

Analog-to-Digital Converter Board — The ADC chosen is ADC1-5000 DMUX 1:1 [8] which converts the optical input from ATOM frontend into electronic signal at $4\,\mathrm{GSamples/s}$. This sampling rate basically depends on the clock frequency supplied by the signal generator HP8648C. Note the ADC can

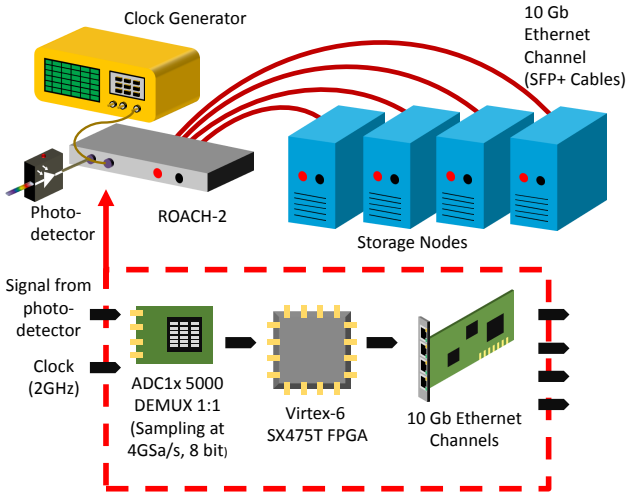


Fig. 2: Overview of the proposed FPGA-based data acquisition system.

achieve a sampling rate of 4 GHz with a clock frequency of 2 GHz by the mean of inter-leaving.

ROACH-2 Revision-2 & 10Gbit Ethernet — The quantized values, which are represented as 8 bit × 16 outputs, are collected by the onboard Virtex-6 FPGA. The FPGA, on the other hand, runs at 250 MHz based on the clock frequency from the signal generator. Essentially, the FPGA clock is always equal to $S/16$, where S = sampling frequency to ensure that the FPGA is able to receive all the sampled data.

The data captured is then divided into multiple segments on the FPGA and each portion is transferred to the four computing nodes for storage via the 10 Gbit Ethernet with UDP.

Moreover, based on our experiments, the maximum throughput of the 10 Gbit Ethernet is around 9.5 Gbit due to the existence of the packet header. Therefore the overall throughput that can be delivered by the four Ethernet ports is around 38 Gbit, which is more than sufficient to stream all the frontend data generated by the ADC (8 bit × 4 GHz = 32 Gbit).

Computing/ Storage Nodes — Due to the limited processing speed of CPU as well as the writing speed of SSD, receiving packets from the four 10 Gbit channels on a single machine without packet drops is basically impossible. Therefore, the backing store is composed of four separate computers in which each of them is equipped with a 10 Gbit network interface card to collect the transferred data.

Furthermore, the design of the backing store has undertaken multiple considerations to achieve reliable transmission. To begin with, each node is connected to ROACH-2 with a direct, dedicated SFP+ cable so as avoid any congestion or packet drops incurred by Ethernet switches. Also, the socket receive buffer for UDP is resized to about 1.5 GB on each machine in order to enhance reliability. Finally, the capturing process on each node is performed with `n2Disk` software which places all the packets received on the 32 GB memory onboard for fast, temporary storage.

When the memory space is used up on all the machines, data acquisition from the ATOM frontend will be paused and data on the main memory will be written to solid-state drive. The capturing process will be re-initiated again once all the data is permanently placed on the long-term storage.

B. Data Processing on FPGA

As mentioned above, the FPGA is mainly responsible for collecting, processing and directing the data from ATOM frontend to the backing store. Figure 3 displays a diagram which indicates all the major blocks on Virtex-6 FPGA for the proposed system.

ADC Block & FIFO Buffer — The ADC block provided by CASPER outputs 8 bit × 16 of data at every FPGA clock cycle, providing a constant bandwidth of 4 GSamples/s. These 16 sampled values are concatenated as two 64 bit of data, and are sent to four separate FIFO buffers according to the selection signal over the two de-multiplexers.

The selection signal `sel` is generated from a 1 bit counter and it periodically outputs a value of 0 and 1. When `sel`== 0, the de-multiplexers direct the top and the bottom 64 bit to FIFO-A and FIFO-C respectively. While in the case `sel`== 1, FIFO-B and FIFO-C are the destinations for the concatenated 64 bit values. This process is to enable uniform distributions of the ADC outputs across the four FIFO buffers.

Packeter Blocks — The key component of the FPGA design is the Packeter which accepts quantized values from the FIFO buffer and transfers the data to 10 Gbit Ethernet block. In particular, this module has to send out data according to a rate that is specified by input parameters `length` and `period`. Essentially, input `length` represents the size of the Ethernet payload while input `period` indicates the number of cycles that is required to pump the payload into the Ethernet block. These parameters can also be represented by the equation,

$$period = length * 2, \tag{1}$$

where

$$period < max\_payload\_size/8\,bit, \tag{2}$$

where $max\_payload\_size$ is the maximum payload size, i.e., 65 536 bit, that can be supported by the Ethernet block.

Therefore, if $max\_payload\_size$ == 65536, then $period$ == 2048 and $length$ == 1024 in order to satisfy both (1) and (2). Default value of `length` is set to be 1024 to maximize the throughput of the Ethernet channels.

10Gbit Ethernet Blocks — The Ethernet block is provided by CASPER that accepts 64 bit inputs from Packeter and constructs UDP packets based on the accepted contents. The destination IP address and port number can be specified with a hardware register on FPGA using BORPH operating system.

As mentioned, the maximum throughput of the Ethernet channel is around 9.5 Gbit, therefore the digitized data from the ATOM frontend is distributed across the four Ethernet blocks evenly to sustain a reliable transmission. This is

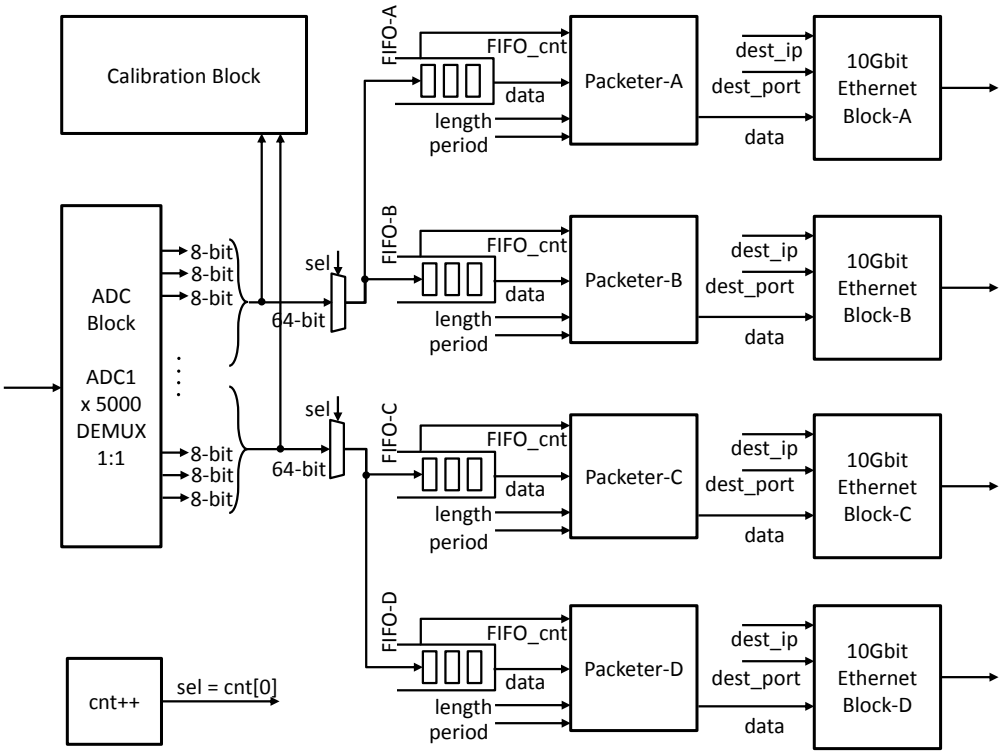


Fig. 3: Block diagram of various modules on FPGA for the proposed system.

achieved by partitioning the $8\,\mathrm{bit} \times 16$ sampled data to the four FIFO buffers, and also maintaining a constant rate of data output from the Packeter which are specified by parameters `length` and `period`.

C. Design Methodology

The FPGA design described in Section III-B is developed with the Simulink-based design flow equipped with the CASPER library. This platform is an invaluable tool for our target application where processing of low-level signal input from ATOM frontend can be implemented with the predefined blocks available in Simulink + ISE. The co-simulation environment also helps developing some of the timing critical blocks in Verilog such as Packeter, and integrating it with the rest of the Simulink design.

Moreover, ROACH-2 from CASPER can achieve excellent system observability and controllability with BORPH operating system, where a hardware-software co-design interface is created to enable software based control of the hardware. The state of the hardware can be read out through software registers, and the hardware registers can also be written with the software interface. This helps intensively in beginning and ending the data capturing process, and also provides a way to specify the IP addresses for the data destination.

In the future, when scientists need perform extensive algorithm exploration on the final received images, the tight integration between Simulink and Matlab will be useful in the sense that the cell image data can be effectively cross compared between software and hardware in the same environment. Such environment will also be beneficial for work extended on top of the current FPGA implementation, such as studies on in-situ object detection and classification with machine-learning algorithm.

IV. Experimental Results

To study the design implications of the proposed system in high-speed data acquisition with ATOM technology, two sets of imaging experiments were performed to demonstrate the feasibility and performance.

A. System Parameters

In each of the experiments, the repetition rate of the laser beam was $R = 11.6\,\mathrm{MHz}$. In the ATOM imaging system, the sampling frequency of the ADC has a direct effect on the horizontal resolution of the resulting images. As part of our optimization effort, a stretched fiber with $1.6\,\mathrm{ns/nm}$ GVD was picked to support the relative low sampling rate of the ADC board so as to deliver high quality images.

The electronic signal from the photo detector was sampled by ADC1x5000-8 DMUX 1:1 with a sampling rate of $4\,\mathrm{GHz}$. The digitized 8-bit values were collected with the Xilinx Virtex-6 XC6VSX475T-1FFG1759C FPGA on ROACH-2. The four storage nodes were connected to ROACH-2 with four separate SFP+ 24 AWG cables manufactured by Molex Incorporated, and each node is equipped with an Intel i5-3470 $2.90\,\mathrm{GHz}$ CPU, $32\,\mathrm{GB}$ RAM and $256\,\mathrm{GB}$ SSD.

During the imaging process, the data received was placed on the main memory for fast, temporary storage. When the

memory space was eventually consumed, the capturing process would be paused until all the data is transferred to the SSD. Finally, when the imaging was completed, the data collected in each of the machine was transferred to and multiplexed in one computer to reconstruct the images offline.

B. Experiment I: Microemulsion Droplets Imaging

Experimental Set-up — To study the quality of the images processed by the proposed FPGA system, microemulsion droplets were created with droplet generation device, flowed in the microfluidic channel and imaged with ATOM technology.

Microemulsion droplets are micrometer scale emulsion droplets that composed of two immiscible fluids and surfactant molecules. In this experiment, the droplets generated were around $60\,\mu\mathrm{m}$ in size and each composed of an outer oil phase and an inner water phase. Beads were inserted in some droplets with $10\,\mu\mathrm{m}$ polystyrene beads from Bangs Laboratories.

To perform ultra-high speed microfluidic flow and imaging, the generated droplets were subsequently reinjected into a separate microfluidic device. Droplets in device were spatially separated along the flow direction, and were accelerated in a linear speed around $1\,\mathrm{m/s}$.

The decision to image microemulsion droplets in this experiment is essentially a method to evaluate the feasibility and practicality of the proposed system. To begin with, microemulsion droplets share similar properties with human cells in terms of their size and dimensions. Furthermore, as the internal features of microemulsion droplets can be clearly displayed with ATOM, it allows us to understand the quality of the images captured by the proposed system more effectively.

Results — Figure 4 shows the droplet images captured by ATOM frontend and collected by the proposed FPGA system. Also displayed in the figure is a similar set of droplets captured by Agilent 86100A Infiniium DCA Wide-Bandwidth Oscilloscope with a sampling rate of $80\,\mathrm{GHz}$, the highest frequency supported by the oscilloscope.

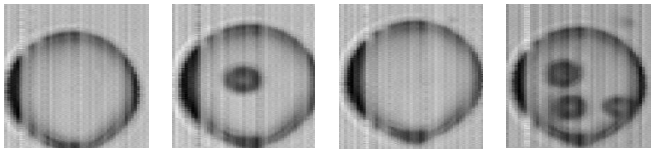
The droplets collected by both systems, regardless of the sampling frequency, are having similar quality where the size, shape and internal features can expressed very clearly. In particular, the beads within the droplets can be recognized unambiguously even with a lower sampling frequency. This result certainly indicates that the proposed FPGA-based system is more than sufficient to perform high-throughput and high-speed bioimaging with ATOM technology.

C. Experiment II: Cancer Cell Imaging

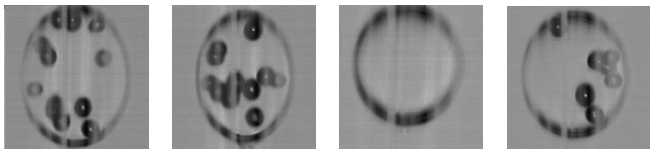
Experimental Set-up — To further understand the design implications of the proposed system in real-life scenarios, two sets of breast cancer cell lines, MCF-7 and MB-231, were individually imaged as the second experiment.

Both MCF-7 and MB-231 are both invasive ductal carcinoma (IDC) that have spread beyond the ducts into other parts of the breast tissue.

Similar to the imaging experiment carried out for microemulsion droplets, each cell was injected into a separate microfluidic device, which could spatially separate the cells along the flow direction. The cells within the microfluidic channel were also accelerated with a constant speed around $1\,\mathrm{m/s}$.



(a) Images captured by the proposed FPGA system with $4\,\mathrm{GSamples/s}$



(b) Images captured by the oscilloscope with $80\,\mathrm{GSamples/s}$.

Fig. 4: Images of microemulsion droplets captured by the FPGA-based data acquisition system and the oscilloscope.

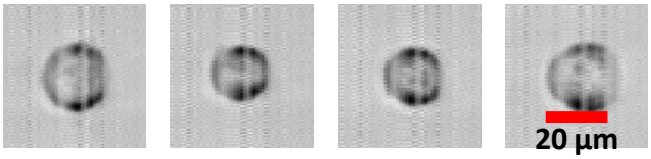
Results — As shown in Figure 5 and Figure 6, the cell images obtained using the FPGA-based system are comparatively less informative. The internal features of each cell are more apparent in the images captured by 86100A. Such discrepancy is due to the fact that the oscilloscope was sampling with a much higher frequency ($80\,\mathrm{GHz}$) than that of the ADC in ROACH-2, which consequently digitized more spatial information (a few nm) about each of the cells.

In spite of this, we believe that the quality of the images retrieved by the FPGA solution is sufficient for many medical analysis such as cell detection or cell type classification with machine-learning algorithm. While there is a merely reduction on the fined-grained details inside each cell, the final classification accuracy would not remarkably be affected, especially when the shape of each type of cell is relatively different.

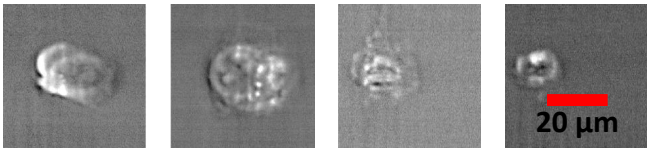
On the other hand, as shown in Table I, the total volume of continuous data-stream that can be captured by the proposed framework is $10\times$ more than that by the oscilloscope. In the experiments above, the maximum amount of data-stream that the FPGA system can collect is around $3.99\,\mathrm{GB}$, which is equivalent to a time-span of $0.56\,\mathrm{s}$. As there exist other processes and the Linux operating system running on the computing nodes, the `n2disk` software is unable to fully exhaust the entire memory space. The recording capacity can be further expanded by 2 times by incorporating $64\,\mathrm{GB}$ of memory on each computing node. In the oscilloscope, however, approximately $350\,\mathrm{MB}$ of data-stream can only be taken with a sampling frequency of $4\,\mathrm{GHz}$ which is simply equivalent to a time-span of $0.05\,\mathrm{s}$.

D. Resource Consumption

Table II summarizes the resource consumption of the proposed FPGA-based data acquisition system. On the FPGA

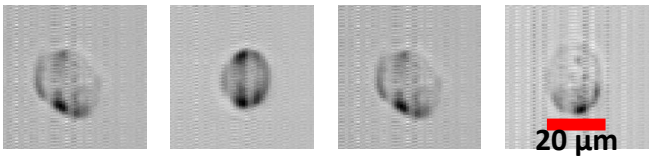


(a) Images captured by the proposed FPGA system with 4 GSamples/s.

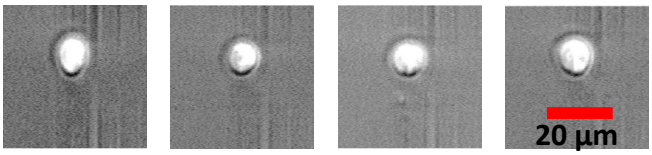


(b) Images captured by the oscilloscope with 80 GSamples/s.

Fig. 5: Images of MCF-7 cancer cells captured by the FPGA-based data acquisition system and the oscilloscope.



(a) Images captured by the proposed FPGA system with 4 GSamples/s.



(b) Images captured by the oscilloscope with 80 GSamples/s.

Fig. 6: Images of MB-231 cancer cells captured by the FPGA-based data acquisition system and the oscilloscope.

TABLE I: Cost and storage capacity of the proposed FPGA-based system vs 86100A oscilloscope.

	FPGA-based System	**Ocilloscope**
Max. Storage Capacity	3.99 GB	350 MB
Total Time-span	0.56 s	0.05 s
Equipment Cost	∼ US$ 14,800	∼ US$ 128,300

TABLE II: Resource consumption of the implementation of the proposed system on FPGA. Percentage values are relative to available resources on Xilinx Virtex-6 XC6VSX475T-1FFG1759C on ROACH-2.

Modules	**Registers**		**LUT**		**BRAM**	
ADC and Calibration	2379	1 %	1495	1 %	159	1 %
Packeter	124	∼1 %	183	∼1 %	0	0 %
10 Gbit Ethernet	2821	1 %	2951	1 %	149	1 %
Overall	14 359	3 %	13 990	5 %	1061	1 %

chosen, the entire implementation consumes less than 4 % of the available resources. In particular, the ADC block together with the four 10 Gbit Ethernet modules consume less than 3 % of Registers, 5 % of LUTs and 1 % of Block RAM. We consider such minimal resource consumption is moderate in the sense that it provides opportunity for future design extensions such as in-situ cellular detection and classification with machine-learning algorithm on FPGA.

V. Conclusions and Future Work

This paper has demonstrated the unique property of FPGAs for high-speed data acquisition with Asymmetric-Detection Time-Stretch Optical Microscopy (ATOM). The presented FPGA-based system is able to continuously retrieve 10× more data when compared with high-end oscilloscope. Furthermore, the proposed implementation can lower the equipment cost by 8× while maintaining high quality images that are adequate for subsequent bioimaging analysis.

In the future, we plan to extend current design to support in-situ cellular detection and classification on FPGA. In particular, we intend to apply machine-learning algorithm that allows real-time identifications of cells to exploit the full potential of ATOM. This will not only be useful for high-throughput cellular bioimaging but also be beneficial for single-cell analysis for disease diagnosis.

The design of the FPGA-based data acquisition system is available on https://github.com/hku-casr/atom-acquisition.

Acknowledgment

This work was supported in part by the Croucher Innovation Award 2013, the Guangdong-Hong Kong Technology Cooperation Funding Scheme GHP/024/16GD, the Research Grants Council of Hong Kong project GRF 17207715, GRF 17207714, and the Innovation and Technology Support Programme ITS/090/14. The support of Lee Family Scholarship and Lady Emily Yee-Man Mak Scholarship is gratefully acknowledged.

References

[1] D. L. Taylor, J. R. Haskins, and K. A. Giuliano, *High Content Screening: A Powerful Approach to Systems Cell Biology and Drug Discovery*. Springer Science and Business Media, 2007.

[2] T. T. W. Wong *et al.*, "Asymmetric-detection time-stretch optical microscopy (ATOM) for ultrafast high-contrast cellular imaging in flow," *Sci. Rep.*, vol. 4, 01 2014. [Online]. Available: http://dx.doi.org/10.1038/srep03656

[3] J. Xie *et al.*, "Accelerated Cell Imaging and Classification on FPGAs for Quantitative-phase Asymmetric-detection Time-stretch Optical Microscopy," in *2015 International Conference on Field Programmable Technology (FPT)*, Dec 2015, pp. 1–8.

[4] J. A. Jensen *et al.*, "Ultrasound Research Scanner for Real-time Synthetic Aperture Data Acquisition," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 52, no. 5, pp. 881–891, May 2005.

[5] C. Leong *et al.*, "Design and test issues of an fpga based data acquisition system for medical imaging using pem," *IEEE Transactions on Nuclear Science*, vol. 53, no. 3, pp. 761–769, June 2006.

[6] ROACH2 - Casper. https://casper.berkeley.edu/wiki/ROACH2. [Online]. Available: https://casper.berkeley.edu/wiki/ROACH2

[7] H. K.-H. So and R. Brodersen, "A Unified Hardware/Software Runtime Environment for FPGA-Based Reconfigurable Computers using BORPH," *Trans. on Embedded Comp. Sys.*, vol. 7, no. 2, pp. 1–28, 2008.

[8] ADC1x5000-8 - Casper. [Online]. Available: https://casper.berkeley.edu/wiki/ADC1x5000-8