A 64-core Platform for Biomedical Signal Processing

Jordan Bisasky[1], Houman Homayoun[2], Farhang Yazdani[3] and Tinoosh Mohsenin[1]
[1] CSEE Department, University of Maryland Baltimore County
[2] ECE department, George Mason University
[3] BroadPak Corp.

***Abstract*—This paper presents a programmable many-core platform containing 64 cores routed in a hierarchical network for biomedical signal processing applications. Individual core processors are based on a RISC architecture with DSP enhancement blocks. Given the number of conditional program loops in DSP applications such as FFT, additional hardware blocks are added that operate in parallel to each core processor. The two blocks calculate the FFT input addresses and determine if a conditional loop is necessary. Performing these operations in parallel to the main processor greatly reduces the time to completion for a DSP application. Each processor is implemented in 65 nm CMOS using standard cell libraries. The 64-core platform occupies 19.51 mm^2 and runs at 1.18 GHz at 1 V. For demonstration, Electroencephalogram (EEG) seizure detection and analysis and ultrasound spectral doppler are mapped onto the cores. The seizure detection and analysis algorithm utilizes 60 processors and takes 890 ns to execute. Spectral doppler utilizes 29 processors and takes 715 ns to run.**

***Index Terms*—65 nm CMOS, DSP, many-core, biomedical signal processing, seizure detection, ultrasound**

I. Introduction

One of the great challenges of the next decade has been on the integration of information technologies and health care. In this case the quality and cost of services for the patients and health care providers must benefit from reduced misdiagnoses and by providing greater access to advanced modalities for more patients [1], [2], [3], [4], [5]. In addition, wearable biomedical devices are used in inpatient, outpatient, and at home e-Patient care [6] that must constantly monitor the patient's biomedical and physiological signals 24/7.

Several biomedical applications require execution of digital signal, image, and video processing algorithms. For instance, ultrasound and seizure detection both contain different filtering, FFT blocks, up/down sampling and windowing techniques that can be parallelized on DSP processors. On the other hand, such portable devices have extremely small budgets for size and power, which currently use application specific integrated circuits (ASIC) or highly custom SoCs [7], [8].

In this paper, we present a programmable low power many-core processing platform which implements the workloads of biomedical signal processing efficiently for computer aided diagnosis. The paper is organized as follows, Section II discusses background on many core platforms, seizure detection and ultrasound. Then the many-core architecture and its enhanced features are discussed. Finally, the CMOS implementation of the many-core along with the application mapping results are presented.

II. Background

A. Many-core Trend

Many-core architectures have been well studied as a potential rival to reconfigurable fabric based platforms [9]. One common many-core architecture is the multiple instruction and multiple data (MIMD) architecture. In such a design, the processors are independent cores capable of executing their own instruction streams and process data through their input-output (I/O) ports and/or local data memories. These architectures have found popularity in replacing preexisting high-performance FPGA+DSP system-on-a-board solutions with a single homogeneous platform. Examples include the Tilera's 64-core Tile64 [10] (based on the MIT Raw architecture [11]), Picochip's 308-core picoArray [12], [13], Ambric's 336-core Am2045 [14], [15], Intel's Polaris 80-core floating point research chip [16], Intellasys' 24-core SEAForth-24A [17], and the UC Davis 36-core and 167 AsAP array [18], [19].

B. Seizure Detection with EEG

About 50 million people worldwide suffer from epilepsy [20], a neurological disorder characterized by seizures. Posttraumatic epilepsy are developed in over 50% of traumatic brain injury (TBI) victims with penetrating head injury in battlefields [21]. The primary tool for diagnosis of an epileptic seizure is an electroencephalogram (EEG) which measures brain activity. Detection and analysis require the placement of several electrodes on the scalp with each electrode being a sensing channel. Previously proposed algorithms and implementations have targeted low power, portable detection in a non-clinical setting [22] [23] [24] [25]. A low power seizure detection and analysis architecture was initially proposed by the authors [26]. Figure 1 depicts the high level task graph of the EEG seizure detection algorithm. Detection is performed in the time domain by determining EEG high frequency components and comparing their magnitude with predetermined threshold values that are calibrated for each patient [26], [27]. After detection of a seizure, seizure analysis is performed by converting to the frequency domain using an FFT block and separating the energy of the data into four frequency bands: Theta (4–7 Hz), Alpha (8–12 Hz), Beta (13–29 Hz) and Gamma (30–50 Hz).

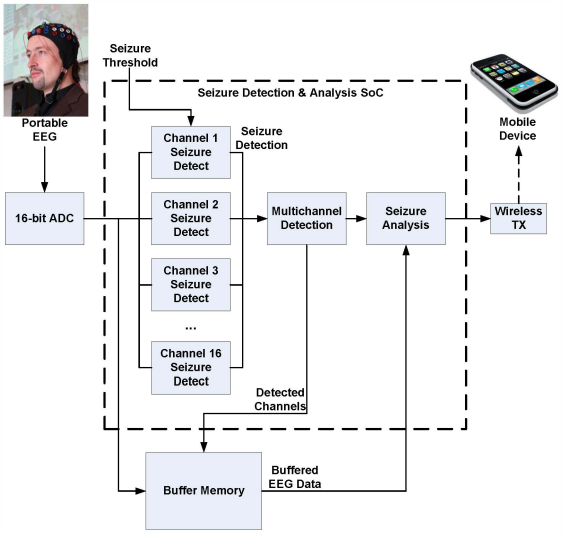

Fig. 1. Proposed multi-channel seizure detection and analysis architecture. 16 single channel detection circuits are instantiated and passed to a threshold detector to confirm a seizure. Subsequently, analysis circuitry is enabled, which perform frequency and energy recording of seizure in certain frequency bands.

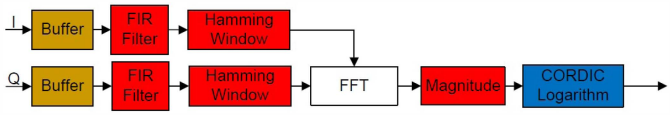

Fig. 2. Spectral doppler mid-end processing task graph

C. Ultrasound Spectral Doppler

Ultrasound imaging uses high frequency sound waves beyond the range of human hearing to generate images within objects of interest. Today, medical ultrasound has become a predominant tool for diagnostic medicine with 159 million exams given annually in the United States and 250 million worldwide. Ultrasound offers several benefits over other imaging modalities including the lack of ionizing radiation from X-rays or radioactive isotopes (from nuclear imaging), and the ability to do real-time 2D and 3D imaging that other modalities struggle to do without high performance computation (for image reconstruction in modalities such as X-ray CT, PET and MRI/NMR).

Within ultrasound processing, there are easily exploitable replicated task pipelines that exploit both temporal and spatial parallelism. Additionally, tasks can be decomposed into smaller blocks which can easily be mapped onto a many-core network. Fig. 2 displays a block diagram of spectral doppler mid-end processing. Many of the blocks can be further decomposed to support task level parallelism.

III. Proposed Architecture

The proposed many-core architecture consists of in-order processors with a RISC-like DSP instruction set, 16-bit integer datapath, and minimal instruction and data memory suitable for task level parallelism [27]. The network is designed as a 4-ary tree architecture, where each single processor in a cluster of four can communicate to any other core within its own cluster through direct connections. The network is designed to support processors operating on asynchronous clock signals in Globally Asynchronous Locally Synchronous (GALS) fashion [19], [28]. Therefore processors can halt or change their clock based on the workload requirement to achieve the maximum energy efficiency.

Each core processor is based on a RISC architecture with a 6 stage pipeline. Figure 3 shows a single processor core diagram. The pipeline includes data forwarding and can perform arithmetic, branch, and inter-core communication instructions. In addition to the pipeline, each core processor contains memory and FIFO blocks. Instructions are assembled using 30-bit words loaded into the instruction memory. Instructions consist of a 6-bit instruction opcode, two 8-bit source inputs, and one 8-bit source output. The instruction memory can store up to 128 words onto the SRAM. Data is loaded from and stored onto the 128 word 16-bit SRAM. Also, 16 registers are used for quick data storage and retrieval. The FIFO blocks are used to queue incoming and outgoing data to other cores and routers on the network.

A. DSP Enhanced Architecture

The primary purpose will be to map and run biomedical applications which are heavily focused on digital signal processing. DSP applications have a predictable runtime compared to an application on a general purpose processor (GPP), so this can be leveraged to decrease the runtime. As a result, the architecture has been designed to improve upon existing GPP's. The goal becomes decreasing processing overhead, so arithmetic operations become the primary runtime component. Three features are added to each core block in order to reduce overhead: parallel loop control, parallel FFT processing, and pointers.

Conditional branch instructions in a GPP must be executed one time per iteration in a program with a loop. For DSP applications, programs commonly iterate tens to thousands of times. Unlike programs on GPPs, the instruction runtime is repetitive and predictable in DSP programs. To take advantage of the predictable runtime of DSP applications, an additional tiny hardware block is included on each core processor. Instead of serially performing decrement and branch instructions each time the program reaches the bottom of an iterating loop, the loop control block performs the operations in parallel to the pipeline. When the block calculates a loop to be performed, an overwriting instruction address is sent from the block to the pipeline.

To aid in the calculation of FFT which is common in biomedical applications, each core block has a tiny FFT block which assists the main pipeline processor with a 4, 8, 16, and 32-point FFT. The radix-2 Cooley-Tukey algorithm computes the FFT in O(NLogN) time and simplifies the arithmetic computations to addition and multiplication. However, the data in memory must be carefully managed in order to correctly calculate the transform. For each iteration, two input data points and the corresponding twiddle factor must be addressed and

read in from memory before the data can be manipulated. In other processors, the addresses for the data and twiddle factor must be calculated in sequence before the FFT calculations. The additional calculations to generate the data and twiddle factor addresses can be a costly time penalty for the FFT. Instead, our proposed core includes an FFT block to perform these calculations in parallel to the FFT calculations. The addresses are generated using simple data reordering and bit reversal and then passed to the decode stage in the pipelined architecture.

To take advantage of both registers and data SRAM, pointer support has been added to reference a memory address using the value of a register. Pointers are another component built into the hardware to perform the calculation with no additional cycles to execute. This is possible since register values are read one cycle before memory reads. The value of the register can then be set as the address of the memory read. Pointers are useful for applications with simple data structures such as lists. The register value can then be incremented or decremented to reference a different address in memory.

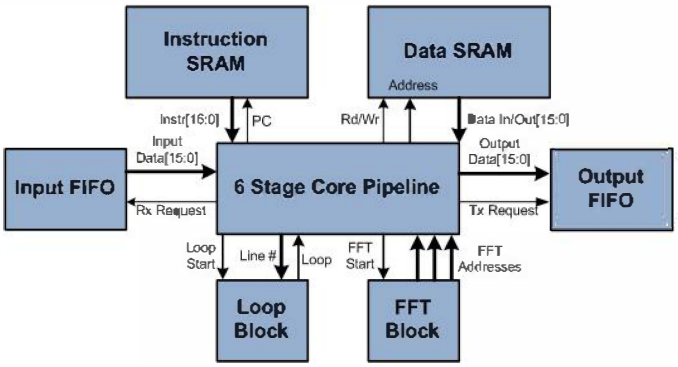

Fig. 3. Single core block diagram consisting of the 6 stage pipeline, instruction and data memory, and input/output FIFO's.

IV. CMOS Implementation and Application Mapping Results

Each processor was synthesized and placed and routed in the 65 nm TSMC CMOS process which occupies 0.070 mm^2 and runs at 1.18 GHz at 1V . We used a standard-cell RTL to GDSII flow using synthesis and automatic place and route. The hardware was developed using Verilog to describe the architecture, synthesized with Cadence RTL Compiler, and placed and routed using Cadence SOC Encounter. Fig. 5 shows the layout of a single core. Table III summarizes the post-layout results. The prototype design routes 64 cores divided into 16 clusters. Every four clusters is connected to a router and every four routers connects to a router in a similar fashion. A single core occupies 0.070 mm^2 and each router occupies 0.0734 mm^2 (shared by 4 cores) and the entire prototype design results in a total area of 19.51 mm^2. The total area of a many core platform is $[2^n \times L_{core} + (2^n - 2) \times L_{router}]^2$, where $n \geq 1$ (hierarchy level), L_{core} is length of one side of a core, and L_{router} is the length of one side of a router.

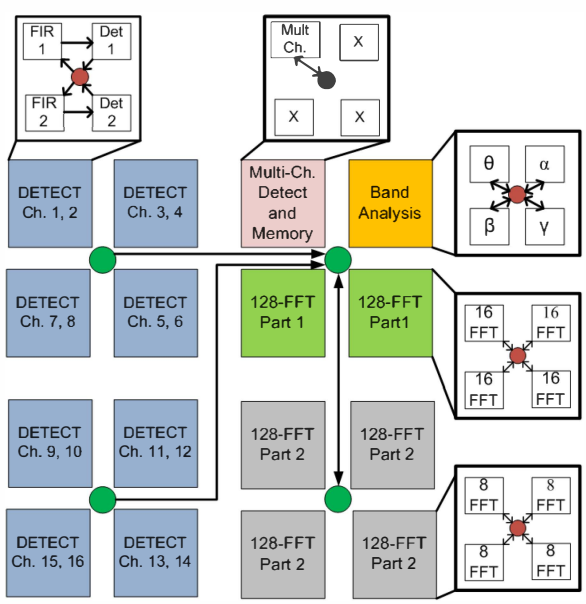

Fig. 4. Mapping of the seizure detection and analysis hardware block onto the many core platform. The implementation supports 16 EEG channels mapped onto 61 cores.

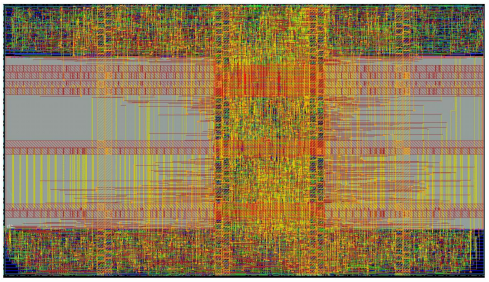

Fig. 5. Layout of a single processor core. The blocks on the left and right side are the instruction and data SRAM memories, respectively.

TABLE I
Seizure Detection and Analysis Memory and Cycle Breakdown Results

Program	IMem	DMem	Time (ns)	Cores
16 tap FIR	50	32	59.5	16
FIR and Detect	53	35	47.6	16
16 pt FFT	75	48	382.5	8
8 pt FFT	35	24	151.3	16
Theta Power	7	2	22.9	1
Alpha Power	7	2	28.0	1
Beta Power	17	2	90.9	1
Gamma Power	22	2	107.1	1
Total	266	147	890	60

TABLE II
Ultrasound Spectral Doppler Memory and Cycle Breakdown Results

Program	IMem	DMem	Time (ns)	Cores
16 tap FIR	50	32	59.5	1
16 tap FIR	50	32	59.5	1
Downsampler	21	0	21.25	1
Hamming Window	11	128	15.3	1
16 pt FFT	75	48	382.5	8
8 pt FFT	35	24	151.3	16
Mag Log	99	0	25.5	1
Total	341	264	715	29

TABLE III
IMPLEMENTATION RESULTS

Technology	65 nm, 1 V
Logic Utilization	96%
Processor Area	0.07 mm^2
Router Area	0.073 mm^2
Processor Speed	1.18 GHz
Router Speed	1.40 GHz

The EEG seizure detection and analysis and ultrasound spectral doppler were programmed and mapped onto the many-core architecture. The mapping of seizure detection and analysis onto 60 cores is depicted in Fig. 4. The detection stage reads the 16 EEG channel inputs where each channel detection occupies two cores each. The 128-point FFT is decomposed into 8 and 16-point FFT blocks and uses 14 cores. As a result, the FFT is calculated in parallel across many cores on the network. Finally, the FFT outputs go to one of four cores to calculate the energy across the four frequency bands.

The results demonstrate performance in area, speed, and power. Despite an instruction and data memory size of only 128 words, the programs mapped onto the cores fit within these constraints. The low memory size minimizes the area of individual core processors. Compared to the previous many-core design from [27], the DSP hardware enhancements discussed in the previous section greatly reduces the number of clock cycles to complete each program. The largest improvement is in the FFT where the loop control and FFT address generation block greatly reduce the number of serialized instructions. Due to an 8-point and 16-point FFT performing 12 and 32 radix-2 calculations respectively, the additionally cycles without the DSP hardware enhancements become magnified. Additionally, the use of pointers simplified the calculation of magnitude in the ultrasound spectral doppler. Combining the improved cycle completion time and the processor clock speed, the two applications are completed in 890 ns for EEG and 715 ns for ultrasound. Table I and table II summarize the results for mapping and running the EEG and ultrasound applications.

V. CONCLUSION

This paper presents the design and implementation of a many-core platform capable of performing biomedical signal processing applications. The low area, low power, high speed single-core processors are optimized to perform the DSP computations. The core processors are connected in hierarchical clusters which are networked by routers to support parallel processing. The individual core processors are additionally aided with hardware blocks that perform FFT calculations in parallel at a minimal area and power increase. As proof of concept, seizure detection and analysis and ultrasound spectral doppler are mapped onto 60 and 29 cores, which take 890 ns and 715 ns to run, respectively. Each processor is synthesized and placed and routed in 65 nm CMOS, it occupies 0.07 mm^2 and runs 1.18 GHz at 1 V supply voltage.

REFERENCES

[1] B. J. Hillman and J. C. Goldsmith, *The Sorcerer's Apprentice: How Medical Imaging is Changing Health Care*, 1st ed. New York, New York: Oxford University Press, Inc., 2011.

[2] A. Al-Kadi, D. Dyer, C. G. Ball, P. B. McBeth, R. Hall, S. Lan, C. Gauthier, J. Boyd, J. Cusden, C. Turner, D. R. Hamilton, and A. W. Kirkpatrick, "User's perceptions of remote trauma telesonography," *Journal of Telemedicine and Telecare*, vol. 15, no. 5, pp. 251–254, Jul. 2009.

[3] P. B. McBeth, T. Hamilton, and A. W. Kirkpatrick, "Cost-Effective Remote iPhone-Teathered Telementored Trauma Telesonography," *The Journal of TRAUMA® Injury, Infection, and Critical Care*, vol. 69, no. 6, pp. 1597–1599, Dec. 2010.

[4] S. A. Dulchavsky, A. E. Sargsyan, K. M. Garcia, S. L. Melton, D. Ebert, and D. R. Hamilton, "Intuitive ultrasonography for autonomous medical care in limited-resource environments," *Acta Astronautica*, vol. 68, no. 9-10, pp. 1595–1607, May-Jun. 2011.

[5] B. W. Dickson and P. C. Pedersen, "Wireless Image Streaming in Mobile Ultrasound," *TELEMEDICINE and e-HEALTH*, vol. 16, no. 2, pp. 161–166, Mar. 2010.

[6] T. F. Budinger, "Biomonitoring with Wireless Communications," *Annual Review Biomedical Engineering*, vol. 5, pp. 383–412, Aug. 2003.

[7] N. Verma, A. Shoeb, J. Bohorquez, J. Dawson, J. Guttag, and A. Chandrakasan, "A micro-power eeg acquisition soc with integrated feature extraction processor for a chronic seizure detection system," *Solid-State Circuits, IEEE Journal of*, vol. 45, no. 4, pp. 804–816, Apr. 2010.

[8] S. Sridhara, M. DiRenzo, S. Lingam, S.-J. Lee, R. Blazquez, J. Maxey, S. Ghanem, Y.-H. Lee, R. Abdallah, P. Singh, and M. Goel, "Microwatt Embedded Processor Platform for Medical System-on-Chip Applications," *Solid-State Circuits, IEEE Journal of*, vol. 46, no. 4, pp. 721–730, Apr. 2011.

[9] S. Borkar, "Thousand core chips: a technology perspective," in *Proceedings of the 44th annual Design Automation Conference*, ser. DAC '07. New York, NY, USA: ACM, 2007, pp. 746–749. [Online]. Available: http://doi.acm.org/10.1145/1278480.1278667

[10] A. Agarwal, "The Tile Processor: A 64-Core Multicore for Embedded Processing," in *High-Performance Embedded Computing, 2007. 11th Annual Workshop on*, Sep. 2007.

[11] M. B. Taylor, J. Kim, J. Miller, D. Wentzlaff, F. Ghodrat, B. Greenwald, H. Hoffman, P. Johnson, J.-W. Lee, W. Lee, A. Ma, A. Saraf, M. Seneski, N. Shnidman, V. Strumpen, M. Frank, S. Amarasinghe, and A. Agarwal, "The Raw Microprocessor: A Computational Fabric for Software Circuits and General-Purpose Programs," *IEEE Micro*, vol. 22, pp. 25–35, Mar. 2002. [Online]. Available: http://dl.acm.org/citation.cfm?id=623304.624515

[12] G. Panesar, D. Towner, A. Duller, A. Gray, and W. Robbins, "Deterministic parallel processing," *Int. J. Parallel Program.*, vol. 34, pp. 323–341, August 2006. [Online]. Available: http://dl.acm.org/citation.cfm?id=1182756.1182758

[13] R. Baines and D. Pulley, "The picoArray and Reconfigurable Baseband Processing for Wireless Basestations," in *DSP enabled Radio, 2003 IEE Colloquium on*, Sep. 2003.

[14] M. Butts, B. Budlong, P. Wasson, and E. White, "Reconfigurable Work Farms on a Massively Parallel Processor Array," in *Proceedings of the 2008 16th International Symposium on Field-Programmable Custom Computing Machines*. Washington, DC, USA: IEEE Computer Society, 2008, pp. 206–215. [Online]. Available: http://portal.acm.org/citation.cfm?id=1488728.1489188

[15] M. Butts and A. M. Jones, "TeraOPS hardware and software: A new massively-parallel, MIMD computing fabric IC," in *IEEE HotChips Symposium on High-Performance Chips (HotChips 2006)*, Aug. 2006.

[16] S. Vangal, J. Howard, G. Ruhl, S. Dighe, H. Wilson, J. Tschanz, D. Finan, A. Singh, T. Jacob, S. Jain, V. Erraguntla, C. Roberts, Y. Hoskote, N. Borkar, and S. Borkar, "An 80-Tile Sub-100-W TeraFLOPS Processor in 65-nm CMOS," *Solid-State Circuits, IEEE Journal of*, vol. 43, no. 1, pp. 29–41, Jan. 2008.

[17] C. Bailey, "Embedded Arrays Venture Forth: IntellaSys 24-Core SEAforth Chips Target Low-Power Multimedia," *Microprocessor Report: the Insider's Guide to Microprocessor Hardware*, Aug. 2006.

[18] Z. Yu, M. Meeuwsen, R. Apperson, O. Sattari, M. Lai, J. Webb, E. Work, D. Truong, T. Mohsenin, and B. Baas, "AsAP: An asynchronous array of simple processors," *Solid-State Circuits, IEEE Journal of*, vol. 43, no. 3, pp. 695–705, Mar. 2008.

[19] D. Truong, W. Cheng, T. Mohsenin, Z. Yu, A. Jacobson, G. Landge, M. Meeuwsen, C. Watnik, A. Tran, Z. Xiao, E. Work, J. Webb, P. Mejia, and B. Baas, "A 167-processor computational platform in 65 nm CMOS," *Solid-State Circuits, IEEE Journal of*, vol. 44, no. 4, pp. 1130–1144, Apr. 2009.

[20] "Epilepsy," Jan. 2009. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs999/en/

[21] "American epilepsy society," Aug. 2007. [Online]. Available: http://www.aesnet.org/go/press-room/consensus-statements/va-statement

[22] S. Raghunathan, S. K. Gupta *et al.*, "A hardware-algorithm co-design approach to optimize seizure detection algorithms for implantable applications," *Journal of Neuroscience Methods*, vol. 193, no. 1, pp. 106 – 117, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165027010004504

[23] S. Sridhara, M. DiRenzo *et al.*, "Microwatt embedded processor platform for medical system-on-chip applications," *Solid-State Circuits, IEEE Journal of*, vol. 46, no. 4, pp. 721 –730, april 2011.

[24] N. Verma, A. Shoeb *et al.*, "A micro-power EEG acquisition SoC with integrated feature extraction processor for a chronic seizure detection system," *Solid-State Circuits, IEEE Journal of*, vol. 45, no. 4, pp. 804 –816, april 2010.

[25] N. Salleh, K. Lim *et al.*, "Ar modeling as eeg spectral analysis on prostration," in *Technical Postgraduates (TECHPOS), 2009 International Conference for*, dec. 2009, pp. 1 –4.

[26] J. Chander, J. Bisasky, and T. Mohsenin, "Real-time multi-channel seizure detection and analysis hardware," *IEEE Biomedical Circuits and Systems (Biocas) Conference*, Nov. 2011.

[27] J. Bisasky, J. Chander, and T. Mohsenin, "A many-core platform implemented for multi-channel seizure detection," *IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2012.

[28] M. Krstic, E. Grass, F. Gurkaynak, and P. Vivet, "Globally asynchronous, locally synchronous circuits: Overview and outlook," *Design Test of Computers, IEEE*, vol. 24, no. 5, pp. 430 –441, sept.-oct. 2007.