

# Design Methodology: Edgeless 3D ASICs with complex in-pixel processing for Pixel Detectors

Farah Fahim<sup>\*ab</sup>, Grzegorz W. Deptuch<sup>a</sup>, James R. Hoff<sup>a</sup>, Hooman Mohseni<sup>b</sup>

<sup>a</sup>Fermi National Accelerator Laboratory, P.O Box 500, Batavia, IL, USA 60510;

<sup>b</sup>Electrical and Computer Sciences Dept., Northwestern University, 2145 Sheridan Rd., Evanston, IL USA 60208-3118

## ABSTRACT

The design methodology for the development of 3D integrated edgeless pixel detectors with in-pixel processing using Electronic Design Automation (EDA) tools is presented. A large area 3 tier 3D detector with one sensor layer and two ASIC layers containing one analog and one digital tier, is built for x-ray photon time of arrival measurement and imaging.

A full custom analog pixel is  $65\mu\text{m} \times 65\mu\text{m}$ . It is connected to a sensor pixel of the same size on one side, and on the other side it has approximately 40 connections to the digital pixel. A  $32 \times 32$  edgeless array without any peripheral functional blocks constitutes a sub-chip. The sub-chip is an indivisible unit, which is further arranged in a  $6 \times 6$  array to create the entire  $1.248\text{cm} \times 1.248\text{cm}$  ASIC. Each chip has 720 bump-bond I/O connections, on the back of the digital tier to the ceramic PCB. All the analog tier power and biasing is conveyed through the digital tier from the PCB.

The assembly has no peripheral functional blocks, and hence the active area extends to the edge of the detector. This was achieved by using a few flavors of almost identical analog pixels (minimal variation in layout) to allow for peripheral biasing blocks to be placed within pixels. The 1024 pixels within a digital sub-chip array have a variety of full custom, semi-custom and automated timing driven functional blocks placed together. The methodology uses a modified mixed-mode on-top digital implementation flow to not only harness the tool efficiency for timing and floor-planning but also to maintain designer control over compact parasitically aware layout. The methodology uses the Cadence design platform, however it is not limited to this tool.

**Keywords:** 3D edgeless detector, design automation, high speed data transfer, priority encoder for zero suppression

## 1. INTRODUCTION

The Vertically Integrated Photon Imaging Chip, (VIPIC), is a large area, small pixel ( $65\mu\text{m}$ ), photon counting ASIC with zero-suppressed data readout, and the capability of registering photon hits with a time resolution of  $\sim 10\mu\text{s}$  which can be externally defined by changing time frames (frameClk). Additionally, it has a high data throughput of 14.4 Gbps per chip. The analog and digital ASIC tiers of VIPIC consist of  $192 \times 192$  pixel array and is  $1.248\text{cm} \times 1.248\text{cm}$  in size. A 1Mper pixel camera module is developed by arranging a  $7 \times 7$  array of 3D VIPIC's bonded to a large area silicon sensor on the analog side and a readout board on the digital side as shown in Figure 1. The readout board consists of a bank of FPGA's, one per VIPIC to allow processing of up to 0.7 Tbps of raw data produced by the camera. This tiered assembly is a flat module which allows easy assembly of cooling plates for thermal dissipation.

The 3D integration procedure starts with post processing the foundry ASIC wafers. An extra metal layer (Metal 9) is added to create metal bonding posts buried in oxide which results in the required highly planar surface topography [1]. Next, face-to-face fusion bonding of the analog and digital ASIC tiers at the wafer-to-wafer level is achieved. One side of this assembly (digital-tier) is subsequently thinned, planarized and small diameter ( $1\mu\text{m}$ ) backside through-silicon vias (B-TSV's) are inserted, followed by patterning of back-metal (Al) for routing and pads which are later used for connection to the readout board. A new layer of oxide is then deposited, which after planarization is bonded to a thick silicon handle wafer. This results in temporarily burying the pads, to allow further processing for connection to the sensor. The same process of thinning, planarization, B-TSV insertion and back metal patterning is carried out on the other side (analog-tier). At this stage the wafer is diced and known good dies (KGD) are identified for connection to the sensor. An array of KGD's are fusion bonded at the die-wafer level, to a large sensor wafer with minimum gaps. The handle wafer is finally removed

\*farah@fnal.gov; phone 1 630 840-8423; fax 1 630 840-2345; [www.fnal.gov](http://www.fnal.gov)

Fermilab is operated by Fermi Research Alliance, LLC under contract No. DE-AC02-07CH11359 with the U.S. Department of Energy.

to expose the back-side pads on the digital tier. The resulting sensor/ASIC hybrids are then bump-bonded to a ceramic readout board as shown in the Figure 1.

The major advantages of this assembly include complete separation of digital activity from low-noise analog parts, large active area with minimal gaps, uniform distribution of power supplies and I/O pads on the back side. Furthermore the ASIC's can be integrated with sensors without bump-bonds. These fusion bonded devices yield a lower equivalent noise charge compared to their bump-bonded counterparts [2].

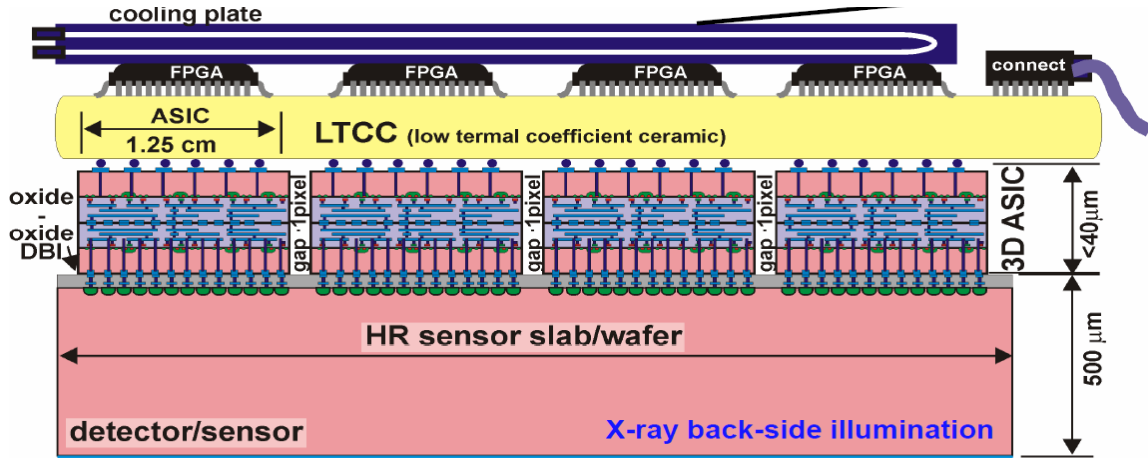


Figure 1. Single module 3D integrated edgeless camera

## 2. ANALOG & DIGITAL PIXELS AND INTER-CONNECTIONS

Each pixel in the analog tier is  $65\mu\text{m} \times 65\mu\text{m}$ , and is designed using full custom analog layout. It contains a signal processing chain which includes a charge sensitive amplifier (CSA) with sensor leakage current compensation, followed by a two stage shaping filter and a window discriminator. Two, 7-bit trimming Digital to Analog Converters (DAC) are used to remove systematic offsets in the comparators in every pixel. Each digital pixel consists of a hit processor, 7-bit counter, 21-bit configuration register and a priority encoder for zero-suppressed readout. Each digital and analog pixel exchange approximately 40 electrical signals between them. This requires the digital pixel to have the same bonding interface as the analog pixel (mirror image).

To easily assemble a large ASIC with an area greater than  $1\text{ cm}^2$ , the  $192 \times 192$  pixel matrix is sub-divided into 36 smaller sub-chips, each containing an array of  $32 \times 32$  pixels. The choice of the number of pixels in a sub-chip, is determined by several factors such as number of data I/O's required per chip, length of pixel address which defines the length of an output data packet, etc.

Each analog pixel is an indivisible unit. These are arranged in a  $32 \times 32$  array to create a sub-chip. A few flavors of analog pixels with minimal variations are created to allow different analog biasing transistors to be placed within different analog pixels. On the other hand, the digital functionality is not physically confined within a pixel but is distributed across a  $32 \times 32$  array to create an indivisible sub-chip shown in Figure 2. This is an essential, unprecedented implementation step, for the edgeless floorplan of the detector, as apart from the pixel logic, the digital tier also needs a high speed output serializer, several differential line drivers and receivers and other additional chip-level functional blocks. Additionally, all global analog signals including power and biases from the readout board have to be conveyed through the digital tier, requiring the use of all metal layers for connectivity to the analog tier, hence areas within the sub-chip are reserved for distributing these signals.

The analog and digital tiers are face-to-face connected using a uniform fusion bonding interface. Metal 9 is added to create metal bonding posts, embedded in oxide. This process results in a highly planar surface topography, required for fusion bonding. The bonding interface is used for exchanging electrical signals while also providing mechanical support. The metal bond post is an octagonal PAD  $2.5\mu\text{m}$  in diameter arranged in a  $5\mu\text{m}$  pitch. The pads which are used for electrical connectivity have vias from metal 9 to metal 8 (last foundry metal layer), which are then routed to the relevant circuitry.

In VIPIC, approximately 25% of the pads in the bond interface are used for electrical connectivity and the rest only provide mechanical support as shown in Figure 3.

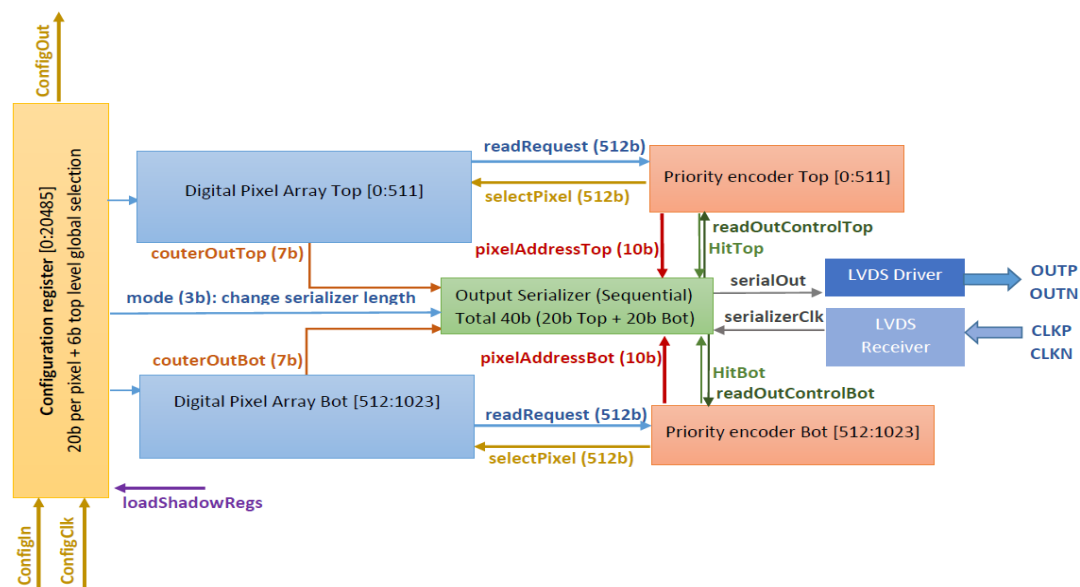


Figure 2. Sub-chip digital functionality

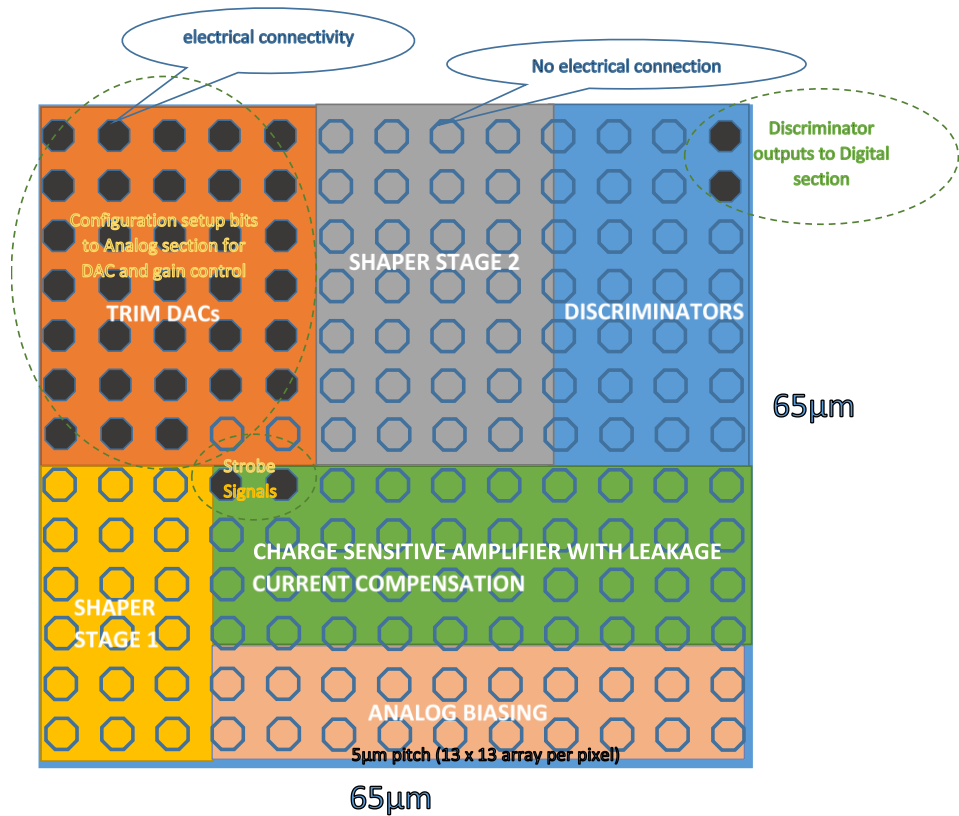


Figure 3. Analog Pixel layout block diagram, showing the bonding interface and approximately 40 interconnections to the digital section

### 3. DIGITAL FLOOR PLANNING & PCB CONNECTIONS

Each sub-chip contains 20 bump-bond pads for external I/O's, they are created using back metal connected to Metal 1 in the ASIC via multiple B-TSVs (more than 100 per group). These bump-bond pads are  $60\text{ }\mu\text{m} \times 60\text{ }\mu\text{m}$  in size, and are placed with a horizontal and vertical pitch of  $520\text{ }\mu\text{m}$  and  $416\text{ }\mu\text{m}$  respectively. Routing of higher density bump-bond pads on the readout board would require very aggressive sizes and separation of traces. A total of 720 bump-bond pads are used per VIPIC. A readout board contains  $7 \times 7$  array of VIPIC's and FPGA's the interconnection translates to an extremely complex layout. To minimize complexity, global signals are shared between two sub-chips as shown in Figure 4 a and b. The 14 analog power and bias signals are distributed on the top and bottom of the sub-chip, these need to be connected from Metal 1 to 9 on the digital tier, which is then electrically connected to Metal 9 of the Analog tier. Shared signals also include digital signals used for analog calibration, StrobeN and StrobeP, digital reset and frameClk. Configuration register clock and I/O and serializer differential I/O are dedicated signals for a sub-chip.

The bump-bond pad sizes are approximately the same size as a pixel. The analog bump-bond pads, irrespective of their placement within a sub-chip, will partially overlap with inter-pixel electrical connectivity at fixed locations every 65 $\mu$ m. Hence, these need to be custom designed to make sure that the inter-pixel connections are not shorted to global signals. These bump-bond pads also create routing and placement restrictions in certain areas across the sub-chip. A power and ground grid for digital VDD and VSS is created utilizing top two metal layers (vertical Metal 8 and horizontal Metal 7), approximately 10  $\mu$ m wide at 65  $\mu$ m pitch.

The differential line drivers and receivers are placed close to the I/O pads and occupy approximately 100  $\mu\text{m}$  x 300  $\mu\text{m}$  area. The central area of the sub-chip is blocked and used to place the high speed output serializer.

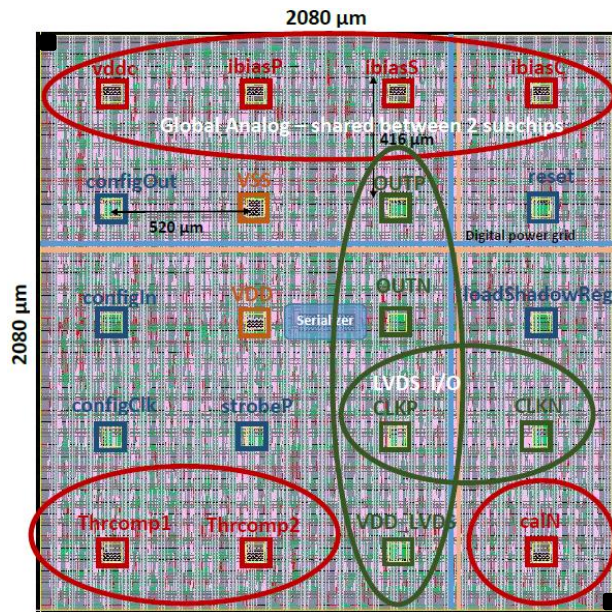


Figure 4 a. Sub-chip 1

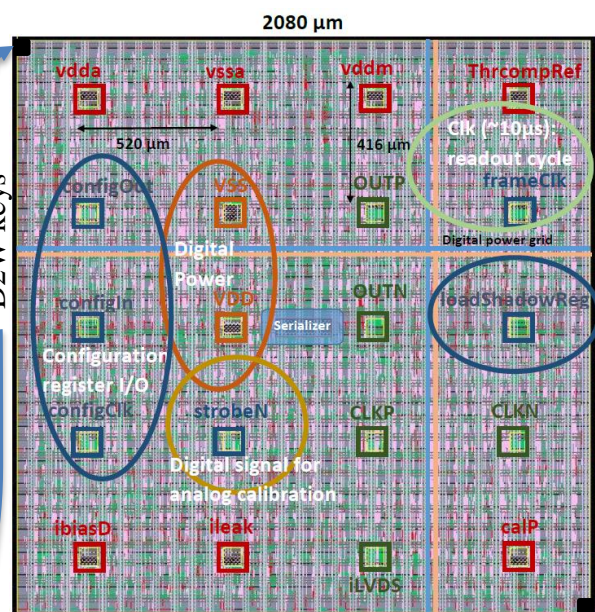


Figure 4 b. Sub-chip 2

The digital tier needs Die-to-Wafer (D2W) alignment keys to be placed at the corners of the chip. Since a sub-chip is an indivisible unit, area to place D2W keys is allocated at the right hand corner top and left hand corner bottom within each sub-chip. This area reserved for markers, should be void and not contain any circuitry.

## 4. DIGITAL DESIGN

Edgeless implementation of the digital tier is challenging from an EDA tool perspective, due to the complexity of functional features required for the application as well as the placement constraints listed earlier. The development stages can be sub-divided depending on the customization required as follows:



### **Full custom digital layout – Hit processor**

The hit processor, accepts the output of the window discriminator from the analog tier and increments a 7-bit gray-counter to register the number of photon hits in the pixel in a given time frame (frameClk). Since the pixel has two 7-bit counters for dead-time less operation, at any given time, one counter is in 'count mode' and the other is in 'read mode' if it had valid data in the previous frame or is 'idle'. At the rising edge of the frameClk, the counters which were in 'read mode' but not yet read out will be reset, while those in 'count mode' with valid data will be swapped and placed in 'read mode'. Counters which were not used will remain in the 'count mode', this feature conserves power in low occupancy detector applications. The logic which checks if the counter is occupied and asynchronously resets it after readout, is extremely sensitive to glitches, hence a gray-code counter was chosen instead of a binary ripple counter to reduce the number of switching bits. Additionally, the choice of gray-code counter reduces power consumption. Spill over protection logic eliminates the data recorded in the previous frame to be allocated to the wrong frame if the entire array was not fully read. Several user defined functions are also added such as using only single comparator instead of a window discriminator etc. The design is asynchronous, without the requirement of a high speed clock tree distribution, as the data is generated by photon arrival. Full analog simulations were performed which clearly indicated that the design is sensitive to parasitics and a full custom layout is needed.

1024 hit processor full custom layout blocks are strategically located across the sub-chip, close to the comparator outputs from the analog tier.

### **Semi-custom digital placement – Priority Encoder**

The priority encoder is used for zero-suppression of data. This also increases the data throughput by only reading those pixels which received photon hits during the time frame (exposure window) defined by the frameClk. The basic data is a simple list of counter value and pixel location. The priority encoder is a binary tree and also generates the address of the pixel [3]. The priority encoder is divided into two parts each generating a 9-bit address for 512 pixels. This allows access of the two parts in an interleaved manner, providing enough time for the address bus to settle when a pixel is selected. The priority encoder heavily relies upon symmetry, its placement doesn't need to be as confined as the hit processor but needs to be guided and symmetrically aware. A symmetrical mirror clone-placement methodology is used to place and route.

### **High speed implementation – Configuration register and output serialiser with mode selection**

The configuration register contains a long shift register chain (21,510-bit). It is the serial communication for programming of the ASIC. Each pixel contains 21-bits out of which 19-bits needs to be sent to the analog pixel across the bonding interface. An additional 6-bits is used for global programming of the ASIC, such as readout mode selection etc. Once the shift register is programmed, its contents are copied to a shadow register. This block requires clock tree distribution of both the configClk for serial shifting and loadShadowReg for parallel loading of the shadow register.

The output serializer is used for high speed data transfer to the FPGA for further data processing. It uses a high speed (~400 MHz) serializerClk to transfer data off chip. This block is centrally placed in a sub-chip and utilizes the data driven capabilities of the tool for accurate timing and data integrity. For convenience and speed, this part of the sub-chip is routed as a separate block.

The three stages of the digital design flow are iteratively repeated, parasitic annotated standard delay format (.sdf) files are created and the top level is simulated across design corners to verify functional performance. The result of this procedure ensures a parasitically aware optimum floorplan. It also maintains timing integrity for timing critical circuits without overloading the EDA tools.

## **5. MANAGING DATA TRANSFER**

Photons arriving asynchronously at the detector generates charge in the sensor, which is processed by the analog pixel and subsequently events are counted in the digital pixel within a certain time period. This time period is determined externally by the user and defined within the ASIC as one period of the frameClk. The resolution of measure of photon time of arrival information is determined by frameClk, which can typically range from a few hundred nanoseconds to a few tens of microseconds depending on the application. Typically it is set at  $< 10\mu\text{s}$ . The change of frame caused by the rising edge of the frameClk creates a new priority list for pixel readout established by the priority encoder. The full output data packet consists of 3-bit synchronization header, 7 bit counter value and 10 bit pixel address. This data is serially transferred using high-speed differential outputs and an output serializerClk running at ~ 400MHz.

For certain applications, the frameClk rates need to be considerably faster ~200ns, however only 4 valid data packets at a data transfer rate of 50ns / data packet can be read within this time frame. A really short exposure time, results in very few events, hence a 7-bit counter will certainly not be fully occupied. Thus truncating the counter to e.g. 2-bits will be sufficient. Hence various readout modes are developed, to change the length of the data packet, which reduce the time for data transfer/packet.

The high-speed output serializerClk, is independent of the slower ~1µs frameClk, which are generally not aligned to each other. Although synchronizing the two clocks is possible but the application might require independent setup of frameClk and serializerClk. Furthermore, synchronisation still does not guarantee correct alignment of the two signals at the pixel, as the clock tree for the frameClk is different from the path readoutControl (derived from the serializerClk) utilizes, through the priority encoder. The delays of these signals are position dependent and cannot be well controlled for a high speed system. The power penalty from buffering and managing the clock tree of a slow clock with an independent high-speed clock is unnecessary, and practically unfeasible. A novel technique for ensuring that high priority data is not corrupted during frame changes has been developed.

### Readout modes

The data output of the ASIC can be either operated in a zero-suppressed or full-frame imaging format, which results in different data packet lengths. In the zero-suppressed format, the data packet needs to contain the 10-bit pixel address and between 2-7 bits of counter value, furthermore the 3 bit synchronization header is optional (but if found to be essential for debugging). In this case, for example a 20-bit data packet transfers 3-bit start symbol, 10-bit pixel address and 7-bit counter value, requiring 50ns to transfer a single data packet, or a 10-bit pixel address and 2-bit counter value, requires 30ns. In the full imaging format, since every pixel is read out only the 7-bit counter value is required, with 17.5ns for readout per data packet which achieves a 55kfps for a 1Mpixel detector, and this could increase considerably if fewer counter bits are chosen.

Since, the time it takes for the shortest data to be read out requires < 17.5ns, building a one stage pipeline ensures sufficient time for the next valid data to be transferred from a pixel to the serializer ready for serial readout. Hence, the 1024 pixels in a sub-chip have been divided into two banks of 512 pixels (top and bottom), each with its own 9-bit priority encoder. This allows for a reconfigurable output serializer with a maximum 40-bit register length, where the two banks have their own independent 20-bit output registers. Interleaved latching of data from the two 512 pixel banks into two parts of the serializer, operates as a one-stage pipeline. When one bank is being readout the other bank is being latched. The Figure 5. shows the main readout modes.

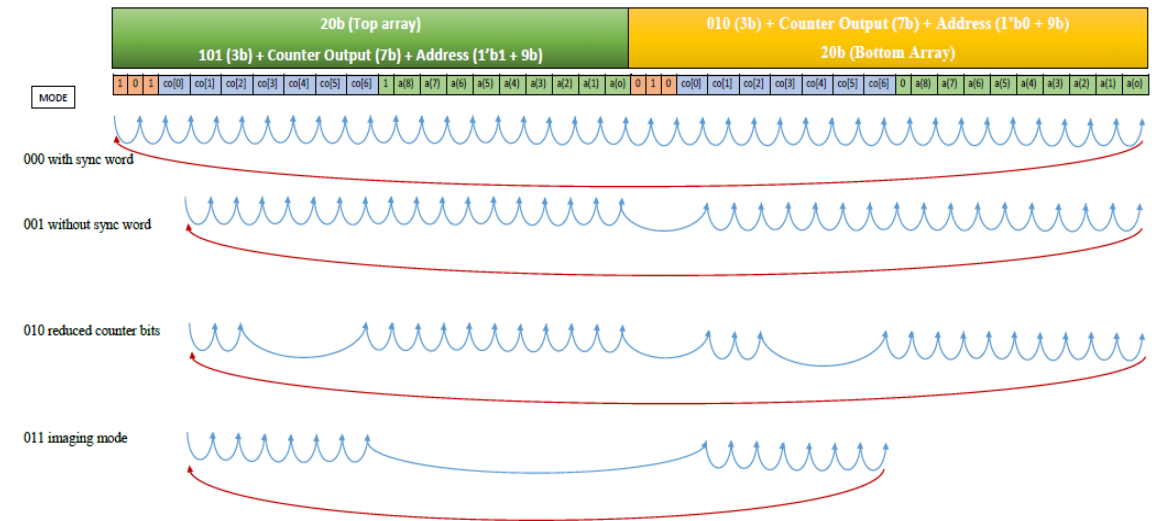


Figure 5. Output serializer readout for different modes of operation

## Maintaining data integrity at frame changes

During the current time frame, each pixel with valid data, sends a request signal for read out, to the priority encoder. The priority encoder establishes the order in which the pixels are allowed to transfer data to the output serializer. The output serializer allows a specific time window for the counter output to be transferred and the address to become available, such that it can be latched in time for off-chip data transfer (by the loadSerializer).

The rising edge of the frameClk, changes the frame. Counters in the 'readout mode' are reset, those in the 'count mode' are changed to 'readout mode' and those that are 'idle' do not change. Simultaneously, the priority encoder creates a new priority list by assigning the order in which multiple pixels are read out.

The following signals are involved in data readout:

- frameClk: rising edge used to indicate change of frame (external slow clock).
- readoutControl: is used to enable data transfer from a pixel to the serializer register, which is generated by the output serializer. This signal is interleaved between the two 512 pixel banks and is alternately broadcasted to the top pixel matrix and then to the bottom pixel matrix for pixel selection. The readoutControl pulse width is 2.5 ns corresponding to the serializerClk of 400 MHz. The time between the pulses of readoutControl is set by the readout mode depending on the number of output bits.
- selectPixel (n): allows for the contents of the counter to be enabled and the pixel address to be established by the priority encoder. Effectively, this signal is the readoutControl signal as seen by the pixel, controlled by the priority encoder. The negative edge of the readoutControl enables the pixel and the positive edge disables it. The next negative edge of readoutControl selects a new pixel, next in the priority list established by the priority encoder.
- loadSerializer: is used to latch data alternately from the top and bottom pixel matrix. It is issued just before the next pixel is selected to ensure that the data has sufficient time to settle before being latched.

It is important to note that if the rising edge of frameClk occurs when readoutControl is high, the current pixel has been disabled but a new pixel has not yet been selected. Hence, no data is corrupted, and the priority encoder can create a new priority list before the next pixel is enabled.

The timing diagrams of three case scenarios are shown in Figure 6 a, b, c. For simplicity, in Case I and II the signals corresponding to only one half of the pixel matrix are shown.

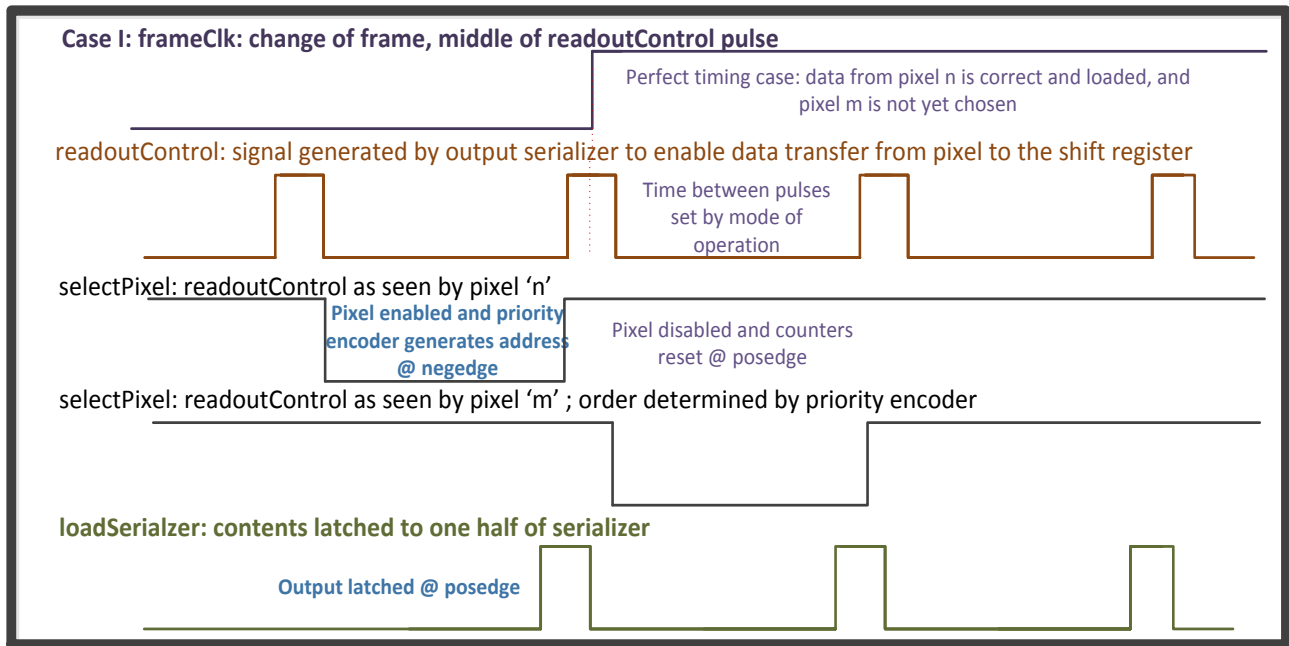


Figure 6 a: Best case scenario when no data is corrupted

Although case I is the ideal scenario for one half of the priority encoder, for the second half, the same frame change does not allow for sufficient time for the address to settle. The timing diagram for the second half is similar to Case II. Case I, also assumes that the frameClk reaches the selected pixel exactly when readoutControl is high. Frame changes can occur at the edge of the readoutControl signal due to different delays of the signals, which can vary from pixel-to-pixel. This can result in glitches in the digital circuit leading to data corruption or losses.

In Case II the change of frameClk can be shifted even further from the edge of readoutControl. This results in disabling the current pixel and hence data it is lost. It simultaneously moves the priority encoder pointer to the first pixel of the new priority list. However, there is not enough time for data to settle, corrupted data is latched by the output serializer. The highest priority data of the new frame is lost.

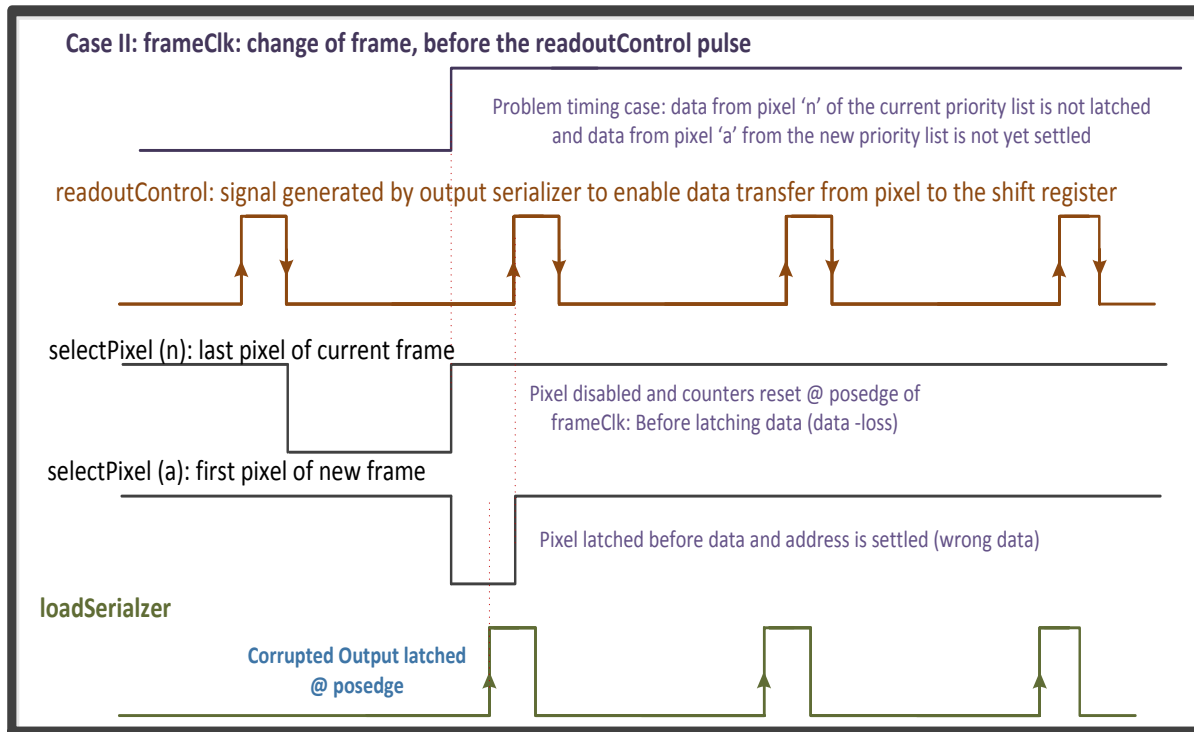


Figure 6 b. Change of frame before readoutControl pulse

Case III shows the change of frame after modification. The correction is based on realizing that the last data in the old priority list is not as important as the first data in the new priority list. It is acceptable to lose the last data but loss of first data in a frame should be avoided.

The rising edge of the frameClk, triggers the rising edge of readoutControl for both the top and bottom halves. This disables the last pixel being readout before the frame changes. The readoutControl is then held high up until at least one complete readout cycle is finished. The frameClk, distributed to the pixels is then delayed to the middle of the high state created on the readoutControl. This ensures the arrival of frameClk edge to any pixel in the matrix when readoutControl is high. Adjusting these signals, leads to a minimal unavoidable dead-time in the readout of data. The following sequence with a time frame change is achieved on the output serial link: two unavoidably corrupted last data outputs, two known data patterns corresponding to frame change and then restarting readout with data from the top of the priority list in the new frame.



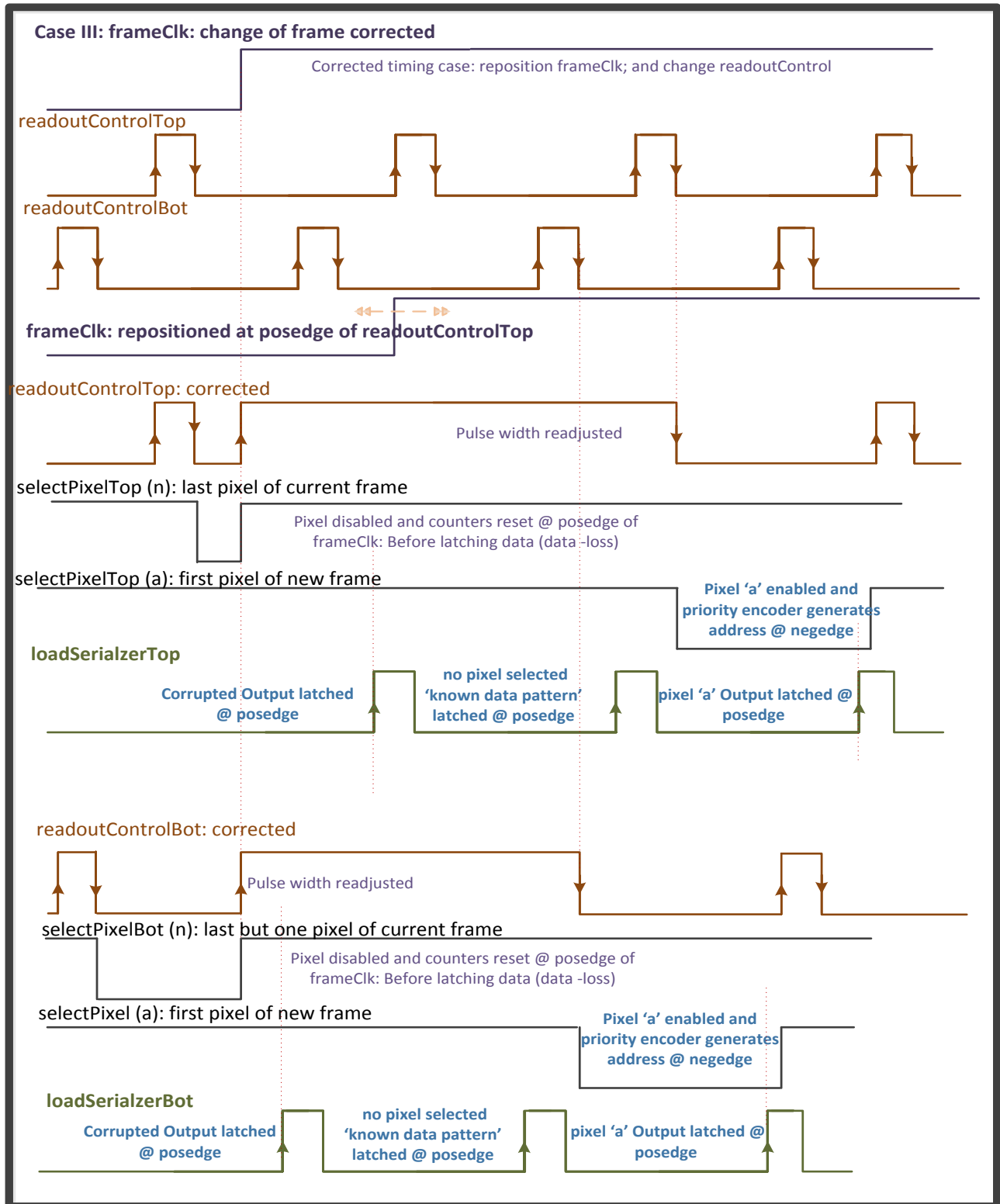


Figure 6 c. Corrected behavior for change of frame

## 6. CONCLUSIONS

The VIPIC large area multi-tier ASIC development with complex analog signal processing and digital data processing is the key part of a large area edgeless camera system with minimum gaps between ROIC's. Grouping a smaller array of pixels into a digital sub-chip allows for adequate area while maintaining indivisible functional features within a repeatable unit. Resources shared between a double sub-chips contain all I/O pads connected to the readout board, which are independent from other double sub-chips on the large ASIC. Their placement and connectivity are optimized both from the aspect of ASIC development as well as the PCB development.

The digital sub-chip implementation required the use of EDA tools in a specialized approach, due to the various placement and functional constraints. An iterative strategy allows for the placement, routing and timing of full-custom, semi-custom and high-speed circuitry such that the user has control over the routing, parasitics and placement of the full-custom block and the tool has control over the clock distribution and timing of high speed blocks.

Several readout modes have been implemented to allow the user to redefine a data packet and change the output data rate/packet. Additionally a novel technique for ensuring that high priority data is not corrupted during frame changes, which results from different propagation delays across the ASIC of the readoutControl and the frameClk has been developed. It allows for minimal unavoidable data loss.

## REFERENCES

- [1] P. Enquist, "Scalable direct bond technology and applications driving adoption," in *Proc. IEEE Int. 3D Syst. Integr. Conf. (3DIC)*, Jan./Feb. 2012, pp. 1–5.
- [2] G. W. Deptuch *et al.*, "Fully 3-D integrated pixel detectors for X-Rays", accepted *IEEE Trans. Elec. Devices*, Jan 2016
- [3] G. W. Deptuch *et al.*, "Design and tests of the vertically integrated photon imaging chip," *IEEE Trans. Nucl. Sci.*, vol. 61, no. 1, pp. 663–674, Feb. 2014.