

A Vertically Integrated Pixel Readout Device for the Vertex Detector at the International Linear Collider

Grzegorz Deptuch, *Senior Member, IEEE*, David Christian, James Hoff, *Member, IEEE*, Ronald Lipton, Alpna Shenai, Marcel Trimpl, *Member, IEEE*, Raymond Yarema, *Life Member, IEEE*, and Tom Zimmerman

Abstract—Tracking and vertexing in future High-Energy Physics (HEP) experiments involves construction of detectors composed of up to a few billions of channels. Readout electronics must record the position and time of each measurement with the highest achievable precision. This paper reviews a prototype of the first 3D readout chip for HEP, designed for a vertex detector at the International Linear Collider. The prototype features $20 \times 20 \mu\text{m}^2$ pixels, laid out in an array of 64×64 elements and was fabricated in a 3-tier $0.18 \mu\text{m}$ Fully Depleted SOI CMOS process at MIT-Lincoln Laboratory. The tests showed correct functional operation of the structure. The chip performs a zero-suppressed readout.

Index Terms—3D-IC, CMOS, FDSOI, pixel detector, SOI, sparsified readout, TSV.

I. INTRODUCTION

THE scaling down of IC technologies to nanometer dimensions is problematic for future analog design in CMOS processes [1]. Nano-scale CMOS processes, in spite of continuously improving lithography methods, suffer from increased technological dispersions due to local variations of doping concentration, built-in stresses and decreased power supply voltage. This results in narrowed signal to noise separation that is particularly troubling for analog design. Three-dimensional integration may be the best way to continue Moore's Law even in the absence of device scaling. The "Handbook of 3D Integration" [2], which is among first compendiums in this domain, defines a three-dimensional integrated circuit (3D-IC) as a chip with two or more layers of active electronic components, integrated horizontally that are thinned and integrated vertically into a single circuit. 3D-IC technology is being pursued in many different forms. 3D circuits can be manufactured by independently fabricating 2D circuits corresponding to different vertical levels (called tiers) on separate wafers, then bonding the wafers together after precise alignment and thinning and interconnecting

them through deep metal vias known as through silicon vias (TSV). TSVs may be added as the last step after wafer bonding in the areas free of active devices or may be an integral part of the foundry process, being formed before or after front end of line (FEOL) processing. The first approach is called via last and the second one is via first [3]. The interstrata vias are typically filled with tungsten or less often with polysilicon. A variety of bond techniques are under investigation. Presently, the dominant four major bonding and interstrata interconnection approaches, two for via last and two for via-first techniques, are: oxide-to-oxide bonding and adhesive polymer bonding for via last, and copper-to-copper bonding and metal/adhesive redistribution layer bonding for via first. The last described technique includes a method called Direct Bond Interconnect (DBI) [4], [5]. It is a combination of oxide-to-oxide bonding and contact metal compression. High reliability, high mechanical endurance and pitches as dense as a few micrometers can be achieved with DBI.

The design described in this work was realized in the 3DM2 Multi-Project Wafer (MPW) run organized by MIT-Lincoln Laboratory, Boston, MA, USA, within the DARPA funded development program in 2006. The 3D-IC process by MIT-LL is based on fabrication of three regular 6" wafers in a $0.18 \mu\text{m}$ fully depleted (FD) SOI CMOS single poly triple metal process that are later stacked together and vertically interconnected using the via-last oxide-to-oxide approach. The choice of SOI for 3D stacking was influenced by the existence of the SOI buried oxide (BOX) layer, which provides a uniformly flat etch stop for wafer thinning [6], [7].

Wafer-to-wafer stacking may be seen as a prelude to the strict-sense 3D-IC methods relying on growing literally an integrated circuit by interlacing new crystalline silicon layers with interconnect metals and inter-layer dielectrics [8]. The crystalline layers can, for example, be attached in a bonding process from a silicon-on-insulator (SOI) wafer or by laser heating recrystallization of amorphous or polycrystalline silicon films deposited on the existing stack. The dopants are included and their electrical activation is achieved in the recrystallization process. The biggest concern is the thermal budget, as fabrication of the devices on the top layer can affect the dopant distribution in the lower layers, as well as affect the reliability of the metal wires below the top layer. Catalysts are used for amorphous silicon deposition, and tungsten-filled vias are used to provide tolerance to high temperatures during the processing. As the semiconductor industry is intensely working on different forms of future 3D-IC technology, 3D circuit designs can be currently realized only by using fully processed 2D wafers, and by using existing process characterization and extracted transistor models.

Manuscript received October 05, 2009; revised November 27, 2009 and January 18, 2010. Current version published April 14, 2010. Fermi National Accelerator Laboratory is operated by Fermi Research Alliance, LLC under Contract DE-AC02-07CH11359 with the U.S. Department of Energy.

G. Deptuch, J. Hoff, A. Shenai, M. Trimpl, R. Yarema, and T. Zimmerman are with the Department of Particle Physics Division, ASIC Microelectronics Group of Electrical Engineering, Fermi National Accelerator Laboratory, BP 500, Batavia, IL 60510 USA (e-mail: deptuch@ieee.org; jhoff@fnal.gov; shenai@fnal.gov; trimpl@fnal.gov; yarema@fnal.gov; zimmerman@fnal.gov).

D. Christian and R. Lipton are with the Particle Physics Division, Fermi National Accelerator Laboratory, BP 500, Batavia, IL 60510 USA (e-mail: dcc@fnal.gov; lipton@fnal.gov).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TNS.2010.2042301

The advantage of the via-last approach is that different 3D-IC tiers can be individually optimized. Theoretically, heterogeneous wafers, i.e., from different foundries or even different process families may be combined, optimally distributing tasks like photon sensing, analog amplification, and digital processing. The sensor layer may be tailored specifically to the needs dictated by the radiation to be detected. Factors affecting sensor optimization include material, pixel granularity, and the choice of front or back side illumination. Each pixel can then be equipped with readout electronics comprising tens or hundreds of transistors distributed on separate tiers.

A fully processed sensor layer (made with high-resistivity silicon for operation in depletion) can be attached to a 3D multi-tier readout circuit in a separate step. Thus, a side benefit of the 3D-IC development is emergence of die-to-wafer or die-to-die fusion bonding techniques that may be used as a replacement of In or PbSn bump bonding. The use of the DBI or similar technique is assumed for attachment of a dedicated pixel detector, fabricated separately on high resistivity n-type silicon, to the readout circuit described in this work.

Future High-Energy Physics (HEP) experiments involve construction of tracking detectors composed of up to a few billion channels (pixels), with square or rectangular pixels of dimensions range from $20\ \mu\text{m}$ to a few hundreds of micrometers. In this paper we describe the first readout chip designed for HEP with the 3D-IC approach. It is a prototype for a vertex detector in the International Linear Collider (ILC). The chip is named VIP1, which stands for Vertically Integrated Pixel.

A. 3D-IC SOI Via-Last Technology by MIT-LL

The 3D-IC 3DM2 MIT-LL run included stacking of three 6" wafers, called tiers A, B and C, fabricated in a $0.18\ \mu\text{m}$ fully depleted SOI (FDSOI) process. Each wafer features a 400 nm thick BOX, 50 nm thick Si islands for transistors, lateral diodes, gate oxide linear capacitors, doped resistors, a single polysilicon layer, and 3 routing metal layers. Tiers B and C have additional back metal layers $2\ \mu\text{m}$ and $0.6\ \mu\text{m}$ thick, respectively, that are deposited on BOX after thinning. This provides enough material for wire bonding pads and helps with power distribution. The total thickness of the structure after completion of the 3D assembly is about $700\ \mu\text{m}$, while the thickness of the 3 active tiers is only about $22\ \mu\text{m}$. The basic configuration for NMOS and PMOS transistors is four terminal devices, each with a floating bulk. One of the terminals is the common SOI "back-gate". The process allows use of 5 terminal devices with an individual bulk contact realized by extension of the bulk material in combination with the polysilicon gate, arranged in a so called h-gate shape. The use of source-bulk-connected transistors is another available option for obtaining transistors with defined bulk potential. This configuration is obtained by placing small inserts of opposite doping of n-type and p-type into sources of PMOS and NMOS transistors, respectively. Transistors with bulk contacts were used in the analog design, in places where better matching is required. In the digital section, floating bulk transistors allow much higher circuit density. In practice, since the thickness of silicon islands is only several tens of nm, the resistance of bulk connections obtained with the methods described above may be

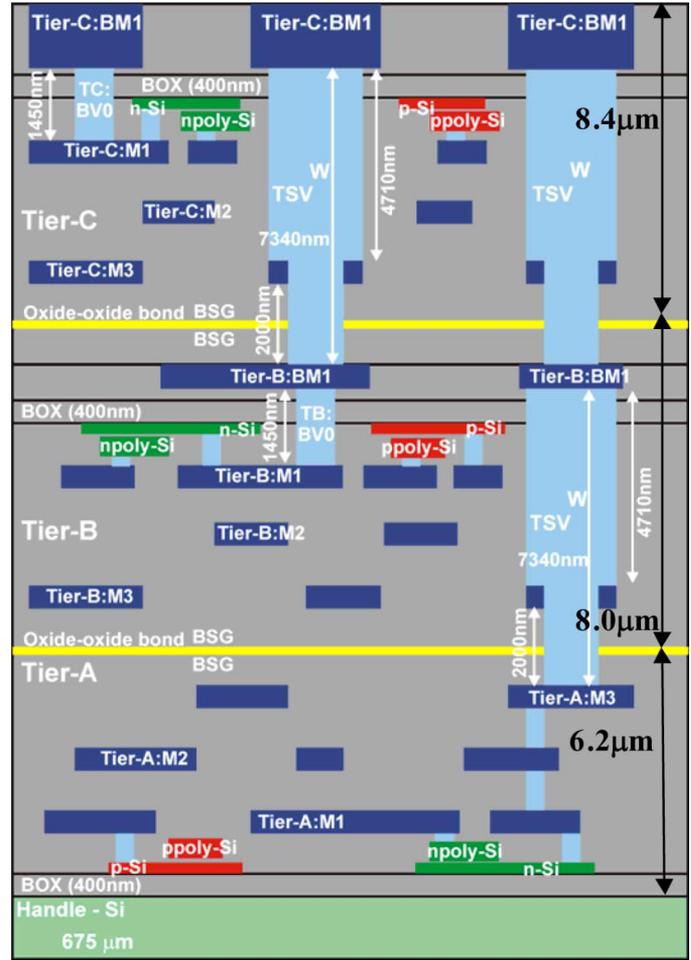


Fig. 1. Cross-section of the 3-tier structure in the MIT-LL via-last 3D-IC process [11].

quite high. The resistance may be made even larger by depletion of silicon occurring as a result of a net charge of ions accumulated in the BOX. Accumulation of contaminant ions, like K, and Na that may travel in ubiquitous oxides, in combination with trapped holes and electrons in oxides, may result in significant transistor threshold voltage shifts. These effects are known from the literature [9], [10], and may be limiting factors for (primarily analog) circuit performance in FDSOI processes.

The cross-section of the 3D stack of 3 wafers after assembly and insertion of TSVs is shown in Fig. 1. The actual area required for each TSV is about $5 \times 5\ \mu\text{m}^2$, including clearance from neighboring circuitry. The vias are done in two phases, i.e., TSVs are inserted between Tier-A and Tier-B after bonding these two layers, and the same operation is then performed after bonding Tier-B to Tier-C. Stacked TSVs from Tier-A to Tier-C are allowed, avoiding excessive consumption of active area. Bonding of Tier-B to Tier-A is done face-to-face, while Tier-C to Tier-B is done face-to-back. Several dimensions are shown in the drawing to give a sense of scale of the structural arrangement of all layers.

B. ILC Experiment Constraints and Requirements

An extensive worldwide R&D program is being carried out on detectors for the ILC. However, a decision to build the ILC

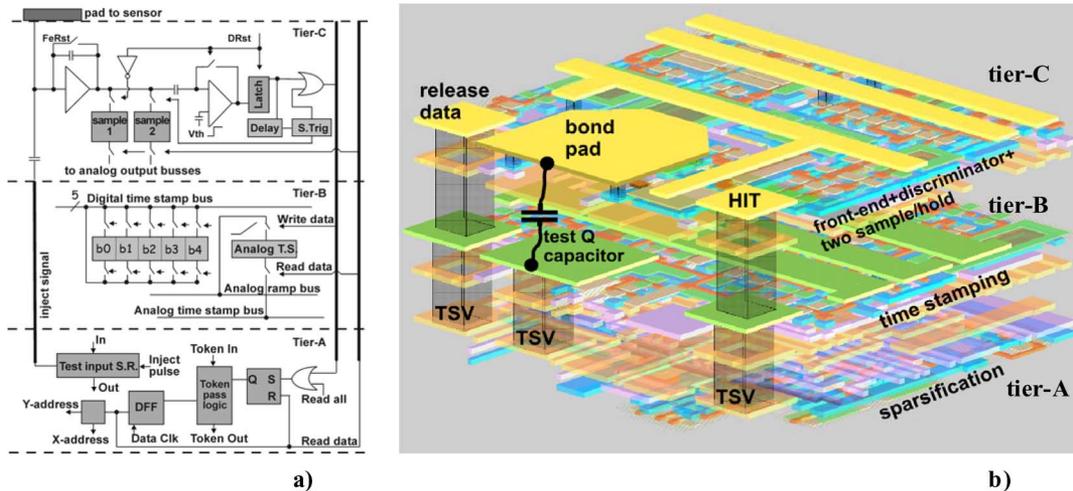


Fig. 2. a) Partition of function between tiers; b) corresponding 3D view of a single pixel (MicroMagic).

is still several years away. The ILC e^+ / e^- machine will have a beam structure of 2820 beam crossings in a 1 ms beam train, occurring 5 times/sec. The current concept of the vertex detector postulates 5 small concentric cylinders of pixel detectors around the beam crossing point. The total number of pixels varies from 0.3×10^9 to 0.8×10^9 , depending on the design scenario. The requested resolution of the impact parameter for correct identification of secondary and tertiary vertices translates to a spatial resolution on the plane of the pixel detector of better than 5 micrometers. Simple binary readout of information would require pixels of $17 \times 17 \mu\text{m}^2$. Adding analog information about the signal amplitude would allow larger pixels, where precision is enhanced by means of centroid weighting. Sparsification, also referred to as zero-suppressed readout, is highly desirable in order to reduce the volume of data being transmitted off the chip, thus reducing the digital power dissipated in the chip.

The data is to be transferred after each beam train. As heat is an issue, various schemes to reduce power dissipation are under investigation. Power in the analog front-end circuitry may be switched off at the end of the beam train, and power in the data transmission circuits can be switched off until a short time before the next beam train is due to arrive. Time stamping with a time binning of 30 μs or better within each beam train is required for unambiguous identification of hits belonging to events within the train. Physics simulations of background, mostly “beamstrahlung” electrons, indicate a flux of 0.03 particles passing through each square millimeter on the surface of the inner-most cylinder for every beam crossing. To allow for charge spreading, hits between pixels, and magnetic field effects in fully depleted sensors, it is assumed that there are 3 hit pixels for every particle. Thus the hit rate on the innermost cylinder is 252 hits/beam train/ mm^2 [12]. Assuming a pixel size of $20 \times 20 \mu\text{m}^2$, dictated by the spatial resolution goal, the average occupancy in a pixel is 0.1-hit/train. Thus, the probability of a cell being hit twice or more in a train is less than 0.5% following the Poisson statistics. The possibility of occurrence of multiple hits in pixels leads to a decrease in the detection efficiency. It can be shown with for example a simple Monte-Carlo simulation that storage of a single hit per pixel

is enough for recording unambiguously 95% of incoming hits. The extension of the storage up to two hits per pixel active for the train time increases the detection efficiency to 99.8%. The known limit of the detection efficiency is not considered as a serious issue for the VIP1 “demonstrator” chip. It is believed that additional pixel buffers can be added without substantial changes in the architecture in the future design. The final depth of the pipeline must be chosen following simulations of the occupancy at the ILC.

The detector must be very thin, i.e., on the order of 50 μm to 75 μm in order to prevent degradation of spatial resolution due to multiple scattering. A thin detector translates to a small charge signal, so the noise requirement is less than 50 e^- of equivalent noise charge (ENC), referred to the input of the in-pixel amplifier. The expected signal from a 50 μm thick detector is about $4 \times 10^3 e^-$.

II. DESIGN AND OPERATION OF THE PROTOTYPE VIP DEVICE

A. General View of the Chip Design

The VIP1 prototype features $20 \times 20 \mu\text{m}^2$ pixels, laid out in an array of 64×64 elements. There are a total of 4096 pixels. There are 175 transistors per pixel, including both the analog and digital sections. Since a detector layer was not available as a part of the 3D-IC run with MIT-LL, the intention was to mate the prototype chip to a detector at a later time [13]. Thus, each pixel was designed with a polygon pad of about 7 μm diameter. Regular $85 \times 85 \mu\text{m}^2$ pads, suitable for wire bonding, were placed on the chip periphery for regular operation of the chip. All pads were located on tier-C (the top tier). The architecture of the chip was chosen to be extendable to sizes in the range of 1024×1024 pixels to be compatible with the actual needs of the application. The partition of functionalities between tiers is shown in the conceptual block diagram in Fig. 2. The 3D view of a complete single pixel is shown next to the block diagram in Fig. 2. The VIP1 chip includes all the major features needed for a vertex detector readout chip at the ILC: 20 μm pixels, readout between ILC bunch trains, high speed data sparsification, digital and analog time stamping, and analog outputs from each pixel

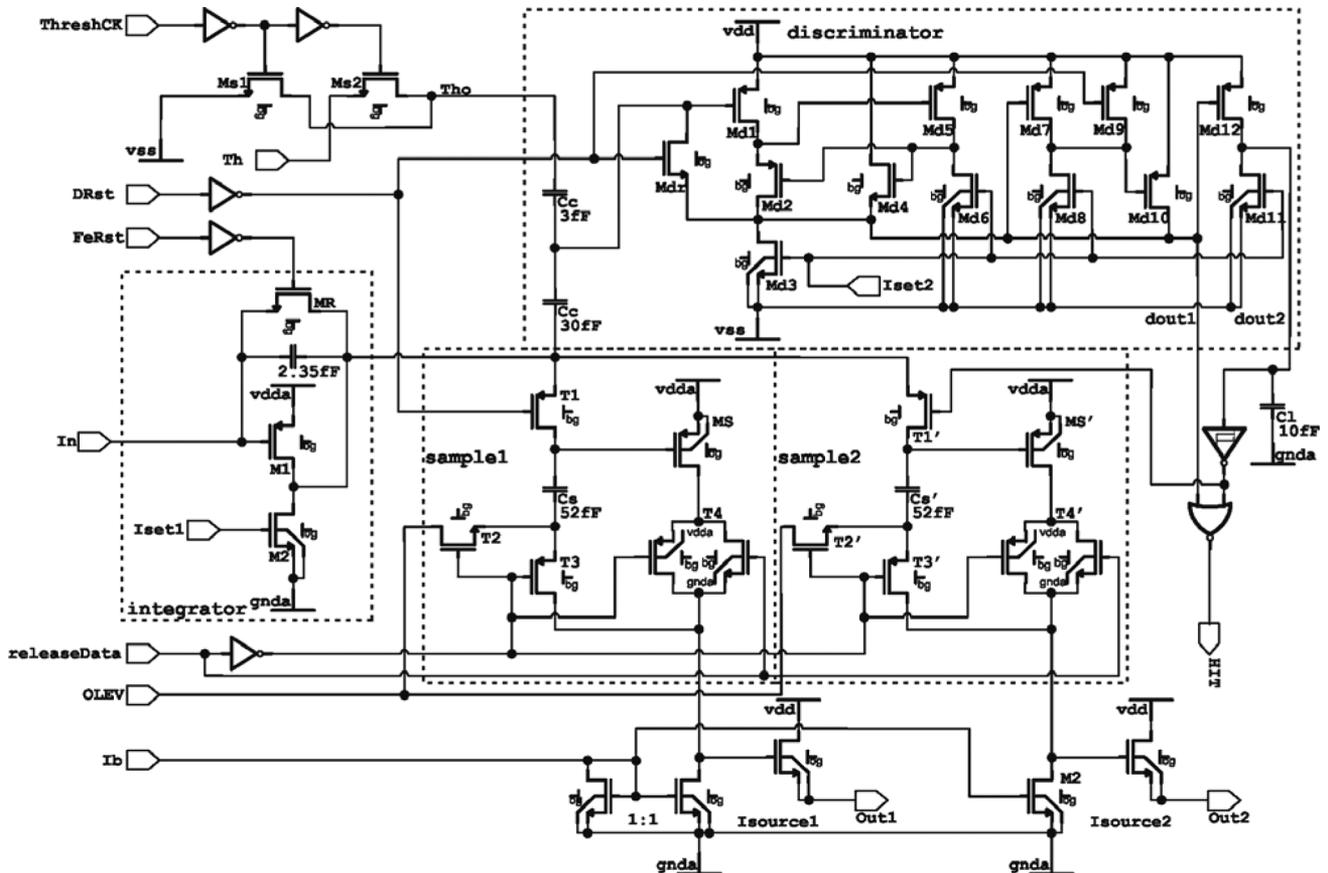


Fig. 3. Detailed schematics diagram of the tier C circuitry: Integrator, autozeroed single ended discriminator, two-cell sample/hold with voltage followers.

for improved spatial resolution. A number of design choices have been made which demonstrate the functionality that can be contained in a small pixel cell. The original design of the chip assumed a binary readout, i.e., only addresses of the hit pixel would be available externally. The analog signal readout from the hit pixels is a feature that was added in the course of the design to allow enhancing the spatial resolution by weighting signals from neighboring pixels. The circuit tasks were distributed between the three tiers such that the critical analog functions are on tier-C (closest to the detector), and the digital readout is on tier-A (farthest from the detector). Tier-B was used for implementing the time stamping circuitry. Analog and digital power supply lines and grounds were separated.

The operation of the chip is divided into two phases synchronized with the cycle of the ILC machine. The acquisition phase, in which signals from particles traversing the detector are being acquired, corresponds to the length of a beam train. The second phase is reporting, in which the collected information is shipped off the chip using a serial link.

B. Integrator, Discriminator, Sample and Hold (Tier-C)

The front end full schematic diagram, shown in Fig. 3, consists of an integrator, a correlated double sampler, an auto-zeroed discriminator and timing circuitry for generation of the “hit” pulse. The signal chain is unipolar, and was designed to accept holes from the detector. The integrator is built with transistors M1 and M2 as a gain stage and a current source, re-

spectively. A positive input charge from holes produces a negative-going integrator output signal. The high gain of the integrator is provided by the small value of the feedback capacitor of approximately 4 fF, including parasitic capacitances between the input and the output of the gain stage. The bias current of the integrator is set in the range of a few hundreds of nanoamperes to a few microamperes. The feedback capacitors are reset in all pixels at once by means of an external reset signal FeRst. The discriminator is built as a single ended high gain stage, which is capacitively coupled to the integrator and accepts a negative-going signal. The (negative) threshold of the discriminator is chip-wide settable. The threshold level is capacitively injected to the input of the discriminator after the DRst signal is removed. The coupling capacitance and the threshold injection capacitance are 30 fF and 3-fF, respectively, resulting in attenuation of the external threshold level step by a factor of 11. The sequence of operation is as follows: first both FeRst and DRst are asserted active, then FeRst is deactivated (the release of the integrator reset), followed by deactivation of DRst (the release of the discriminator reset). The result is cancellation of the integrator and discriminator offsets at the discriminator input. The two sample/hold cells are built with 52 fF capacitors (C_s and C_s'), transistors to provide gain (MS and MS'), and a system of switches. When reading out, the sample/hold circuits operate as unity gain buffers. Nearly rail-to-rail sampling is achieved if the DC reference voltage (OLEV) is properly set. The first sample of the integrator output level is taken when

the discriminator reset is released, and the second one is taken after the discriminator triggers in response to the input signal passing over the threshold. The output of the discriminator is connected to a simply timing circuit. The timing circuit forms a short “hit” pulse used in the time stamping on tier-B and the sparsified readout on tier-A.

C. Time Stamping (Tier-B)

The ILC bunch train is divided into 32 time slices, resulting in an approximate time resolution of 30 μs . A five bit Gray code counter, placed at the periphery of the array of pixels, is used to generate numbers corresponding to each time slot. The counter is reset at the beginning of the bunch train. The code words are distributed simultaneously to all pixels. There are bi-directional buffers with tri-state outputs at the bottom of each column that distribute counter values to the pixels during acquisition phase and allow readout of stored time stamps during the reporting phase on the same lines. The time slice information is stored within the pixel cell latched by the “hit” pulse. The time stamping information is stored in 5 static-type single bit memory cells placed in each pixel. There are 10 transistors per bit in the digital time stamping. At the beginning of the bunch train, all cells are preset to the high logic state. During the acquisition state, the value of an individual bit cell can be flipped only when the “hit” pulse is high. Only transitions from the high logic state to the low logic state are allowed. In addition to the digital time stamping, there is an analog sample/hold block. This analog cell is similar in topology to that used in the double correlated sampler on tier-C. Integration of the digital and analog time stamp was done for test purposes only in the VIP1 chip. The analog time stamping latches the temporal value of a slowly rising (or falling) voltage ramp, which is sent to all pixels from the periphery, at the assertion of the “hit” pulse. After the beam train, i.e., during the reporting phase, the stored analog voltage is read out on a separate bus to an external ADC.

The circuitry storing hits and orchestrating the sparsified readout is located on tier-A. Tier-A also contains one D flip/flop per pixel that is a part of the shift register used to enable pixels for injection of test pulses. The output of the flip/flop drives a transmission gate that allows a voltage step to pass through a small test capacitor to the input of the integrator in pixels that are selected for test pulsing. The test capacitor was realized using parasitic coupling between backside metal on tier-B (Tier-B:BM1) and Metal3 (TierC:M3) on tier-C. The expected value of the test capacitance was about 0.25 fF for the geometry involved. The test capacitance was intended only to check functioning of individual channels, thus mismatch of its value was tolerable. The D flip/flops in the injection shift register were dynamic to save on real estate. However, the choice of dynamic flip/flops was not optimum, since later tests showed difficulty of operation of the injection chain due to extremely high transistor leakage current in the MIT-LL process.

D. Sparsified Readout With Look-Ahead Token (Tier-A)

The pixel sparsification logic is shown in Fig. 4. The hit signal from the discriminator is fed through an OR gate to the hit latch realized with an S-R flip/flop. The second input of the OR gate is used for programming pixels for reading regardless of the

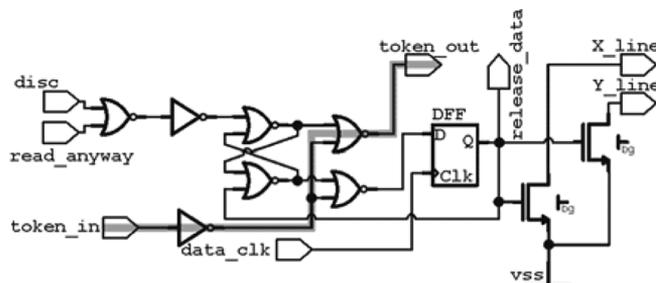


Fig. 4. Pixel sparsification logic on tier-A.

input signal. There are two options available in the VIP chip, i.e., reading all pixels in the array or reading the first column only. The S-R flip flop waits until the signal, called token, is active on the token_in line, and with the next data_clk pulse its content is transferred to the D flip/flop. If the S-R flip/flop was set, two wire-OR X_line and Y_line signals go to the periphery to generate X and Y coordinates of the pixel being read out. Simultaneously, the release_data signal is sent vertically to the tier-B time stamping circuitry and the tier-C front-end circuitry in order to make stored analog and digital values available for readout. The release_data resets the S-R flip flop allowing the transcribing of token_in into token_out. Pixels with S-R flip/flops unset are transparent to the token. Since the conditioning of data for readout consumes some time (addresses generated at the periphery as well as digital time stamping are multi-bit), the propagating token has enough time to either reach the next hit pixel or to leave the matrix. The data clock must account for the maximum delay the token may experience. The worst case is when the matrix is least occupied and the matrix reaches megapixel sizes. Thus, it may be required to force certain pixels be always output to avoid losing the token for matrices of larger sizes. This operation may add some overhead in data sent out. However, it may be seen helpful in optimization of the sparsification circuit leading to avoidance of sending both cartesian coordinates for each hit. The adopted readout scheme is referred to as a look-ahead token passing to sparsify the hit data on the chip. It is similar to the solutions used for the first time in the past in the DELPHI experiment at CERN [14] and based on the development of the FPIX chip for the BTeV experiment at FERMILAB [15]. The pixel addresses are stored on the periphery of the array of pixels in order to reduce the size of the pixel. All digital bits are shipped out through a serial link. The analog information from the correlated double sampler and the analog time stamp are available on separate analog lines at the same time as the digital data.

E. Scalability of the Readout Architecture

A simplified block diagram of a VIP-like $m \times n$ pixel array with zero-suppressed readout is shown in Fig. 5. During the acquisition phase, a hit latch gets set in each pixel that is hit. A sparse readout is performed with a token travelling row by row during the reporting phase. To start readout, all hit pixels are disabled except the first hit pixel to which token arrived. The pixel being read out points to the X and Y addresses that are stored on the perimeter, and the digital time stamp and all the analog information stored in the cell is sent to the edge of the

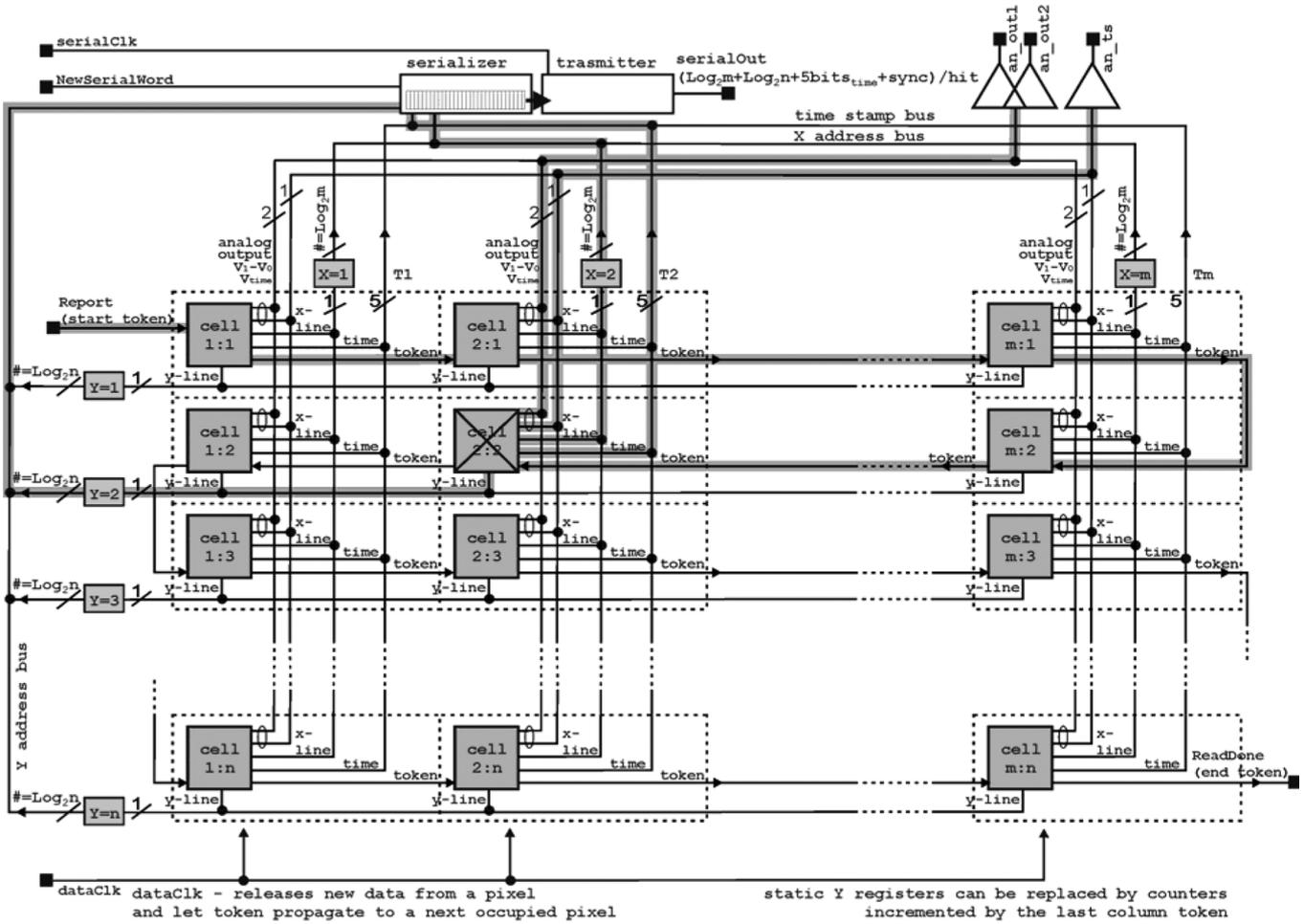


Fig. 5. Simplified block diagram of the VIP1 chip depicting zero-suppressed readout.

chip. While a pixel is being read, the token scans ahead, looking for the next pixel to read. To assure finding the next hit pixel before the readout of the current pixel is finished or the end of the matrix is reached, the maximum time needed by the token to propagate before it stops at the hit pixels has to be shorter than the time needed for shipping the full information from one pixel off the chip. Assuming the token passing speed to be 0.2 ns per empty pixel, the maximum time to reach the next cell to be read is $m \times n \times 0.2$ ns. All the digital information is serialized on the perimeter of the chip for transmission off the chip. When it is assumed that the serializer runs at a very reasonable 50 MHz frequency allowing $t_m = \text{trunc}(\text{Log}_2 m) + 1$ bits for the X and $t_n = \text{trunc}(\text{Log}_2 n) + 1$ bits for the Y addresses, along with 5 to 7 bits for the digital time stamp and status, a hit pixel is read out in $(t_m + t_n + 7 + 1) \times 20$ ns, which may turn out to be less than the time required to find the next hit pixel. Finding the next hit pixel or alternatively reaching the end of the matrix by the token has to be absolutely assured before attempting the next readout. For large matrices, the chip may need to be set to always read out at least a certain number of pixels per row, (for example, the first pixel in each row), regardless of the hit pattern.

There is 199 ms of time between bunch trains to read out the hits in the ILC. Substituting credible values for $m = 1000$ and $n = 1000$ and $20 \mu\text{m}$ pixels, the maximum number of hits in the

hottest part of the detector is calculated to be about 100×10^3 . If one extra pixel is read out in each row, the maximum number of pixels to be read is 101×10^3 . 1×10^3 is a small overhead but it allows reliable operation of the readout chip with one propagating token through the matrix of pixels. To read 101×10^3 of pixels with 30 bits/pixel at 50 MHz takes less than 61 ms over one serial link, therefore the readout time is far less than the ILC inter bunch train time of 199 ms.

III. VERIFICATION OF THE VIP1 DEVICE OPERATION

A. Fabrication and Test Environment

The fabrication of the 3DM3 run by MIT-LL may be seen as basically production of 3 independent runs with different mask sets. In addition to the fabrication of 3 full wafers, the operations required in 3D assembly are complex. The VIP1 design was submitted for fabrication in October 2006. The first lot of 17 diced devices was delivered in November 2007 (lot L1). The time required for completion of the run reflects the complexity of the 3D endeavor undertaken by a non-commercial VLSI line.

When the tests were started it turned out that the yield was very low. Measurements performed on tests structures, which were placed between the actual array of pixels and the wire

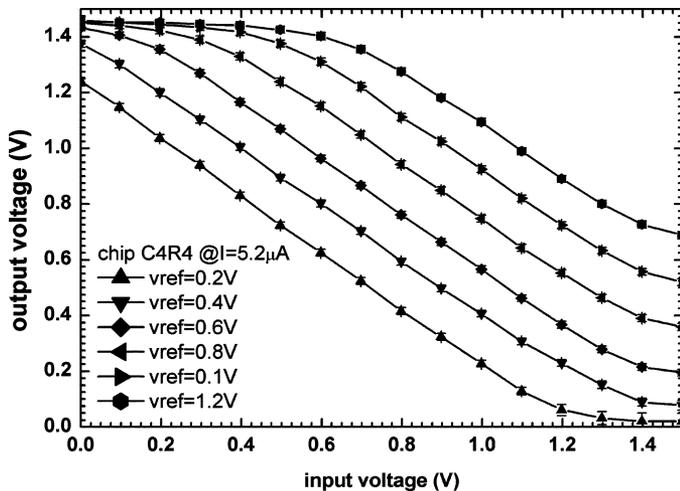


Fig. 6. Static transfer characteristics of the high swing voltage follower circuit for different reference voltages measured on C4R4 chip.

bonding pads, showed the existence of problems. These problems were numerous; however none of them was traced down to any imperfection or omission in connections realized by TSVs. The observed problems could rather be attributed to flaws in the actual processing of individual tiers, such as open circuits in traces of metal on the same tier, shorts between different parts, very high leakage currents in the range of 100 nA through the ESD protection diodes, and fluctuations over time of voltage or current levels in active devices. Tests performed later on the whole matrix of pixels showed that large and variable leakage currents through active devices in the process caused dire problems. Over different chips, the total current consumed from the digital power supply spanned in a very large range from 200 μ A up to 76 mA. The expected draw of current should not virtually be exceeding a few tens microamperes from the digital supply. The only noteworthy current is in the pull-up section of hit pixel address generation.

The second lot of 13 devices was delivered in April 2008 (lot L2), after discovering that important shifts of parameters of active devices occurred in the 3D assembly of the L1 lot. There were reported almost 50% (200 mV) shifts of threshold voltages for transistors on tier-B in the L1 lot. This could be due to the oxide regions in SOI processes that are very vulnerable to charging from exposure to plasma, ion contamination and movement, ionization, etc.

The initial tests were performed on peripheral test structures that included portions of pixel circuitry from each tier. The sparsification logic showed good operation; however, the dynamic D flip/flops retained their states for only a few tens of microseconds. The operation of digital time stamping was demonstrated and the analog time stamping test structure was used for measurement of the static transfer characteristics of the sample/hold circuit with high swing voltage follower. The measurements were carried out by sampling rail-to-rail voltage input signals for different values of the DC reference voltages. The results, presented in Fig. 6, demonstrated operation of the circuit with gain very close to -1 for all reference voltages tested. The negative slope means inverting operation of the voltage follower. The

TABLE I
SLOPE OF THE TRANSFER FUNCTION OF THE FOLLOWER CIRCUIT

Reference voltage Equal to OLEV in Fig.3	slope coefficient	offset coefficient
0.2 V	-1.011 V/V	1.233 V
0.4 V	-0.996 V/V	1.396 V
0.6 V	-1.003 V/V	1.566 V
0.8 V	-0.977 V/V	1.725 V
1.0 V	-0.953 V/V	1.872 V
1.2 V	-0.92 V/V	2.006 V

value of the reference voltage, that can be referred to as potential of the OLEV node in Fig. 3, can be selected in a such way that input voltages separated from both rails by only 200 mV can be processed with very good linearity, despite low power supply used. The slopes of the transfer function of the follower circuit, measured for six different reference voltages, are listed in Table I

The range of the input voltage in consecutive fits was adaptively chosen in correspondence to the reference voltage rising from 0.2 V to 1.2 V in order to stay in the linear region of the transfer function. The fitting ranges of the input voltage were 0 V–1.1 V, 0.1 V–1.3 V, 0.2 V–1.3 V, 0.4 V–1.4 V, 0.6 V–1.4 V, 0.7 V–1.4 V, spanning from the smallest to highest reference voltage.

The discriminator was observed to operate correctly. The integrator response was also measured for typical expected signal amplitudes. For two different bias currents of 0.25 μ A and 0.75 μ A, the integrator time constants were measured at $\tau = 268$ ns and $\tau = 116$ ns respectively, with both sample capacitors connected (before discriminator reset release). With only one sample capacitor connected (after discriminator reset release, discriminator armed), the time constants were $\tau = 216$ ns and $\tau = 95$ ns. An input referred noise from 25 e^- to several tens of e^- of equivalent noise charge (ENC) was measured on an analog front-end test structure, depending on bias currents and capacitive load. All results obtained on working circuits were consistent with prior simulation.

The tests on the matrix of pixels were carried out after initial measurements on test structures. The actual functionality of the VIP1 chip was demonstrated through the following series of tests: token propagation in the two cases of an empty array and an array with all pixels hit, full sparsified data readout, operation of the digital and analog time stamping circuits and readout, threshold scan, input test charge scan, and fixed pattern and temporal noise. The performance was strongly compromised by very low process yield. With a total of 23 tested chips, only a few were usable for assessment of the full functionality. There was no difference observed between chips from lots L1 and L2.

B. Tests of Token Passing

The tests of token propagation were done by measuring the delay time of both edges of the square pulse injected to the token input to the matrix. The matrix was emptied beforehand from any hits by performing a full readout procedure until the matrix was transparent for the token. The token propagation delay calculated per pixel for 3 different chips is shown in Fig. 7 as a function of the digital power supply. The propagation delay time

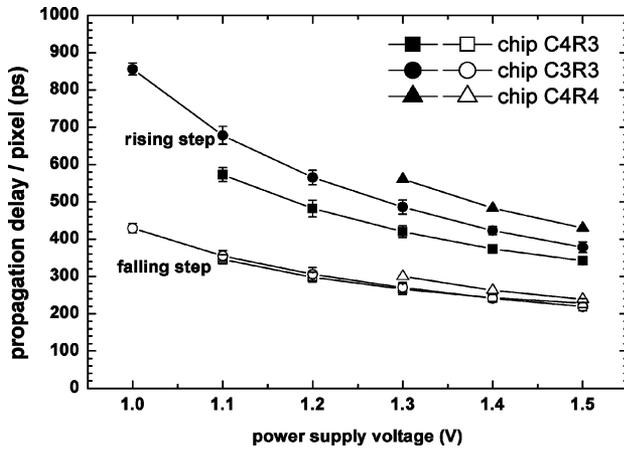


Fig. 7. Taken propagation delay per pixel for different chips at different power supply voltages.

is significantly longer than the assumed 200 ps, but readout of a 1000×1000 pixel detector is still possible between bunch trains of the ILC at a reasonable readout clock frequency of 50 MHz.

C. Tests of the Analog Chain

Low pixel-to-pixel dispersions of the threshold level of the discriminator are crucial as there is no threshold adjustment on an individual basis per pixel in the chip. When the discriminator is armed, charge injection from the reset switch automatically sets a negative threshold (a negative threshold is required since the discriminator input signal is negative-going). The external threshold injected through the capacitor moves the threshold of the discriminator further below the baseline or in the opposite direction, depending on the polarity of the voltage step applied. The measurements of the threshold scan were done using the full sparsified readout mode. After arming of discriminators in each pixel, a threshold was injected from outside. A series of measurements were done scanning the threshold in decreasing order, i.e., from the point where no pixels were triggering up to the point where almost all pixels were present for readout. Tests were carried out for two states of the analog front-end circuits. In the first part of the tests, both the integrator and the discriminator were armed, while in the second group of tests, the integrator was reset (disabled) and the discriminator was armed. Threshold injection levels were measured at 4.4 mV with a one sigma spread of 4.8 mV for the integrator armed, and 21.3 mV with a one sigma spread of 1.6 mV with the integrator kept reset. The intrinsic threshold injected by arming the discriminator can be extracted from the measurement with the integrator kept reset. Its value, referred to the input of the integrator, is about $530 e^-$ with a dispersion of $40 e^-$ when using the designed gain of the integrator. Operation with the integrator reset released shifts the effective threshold by about $425 e^-$ closer to the baseline and increases threshold dispersions to about $120 e^-$. After a careful examination of possible coupling paths and polarities of signals, it was concluded that capacitive coupling through the air between the DRst signal and the input pads of pixels could result in the observed effect. The increase of the dispersions may be explained by different coupling distances due

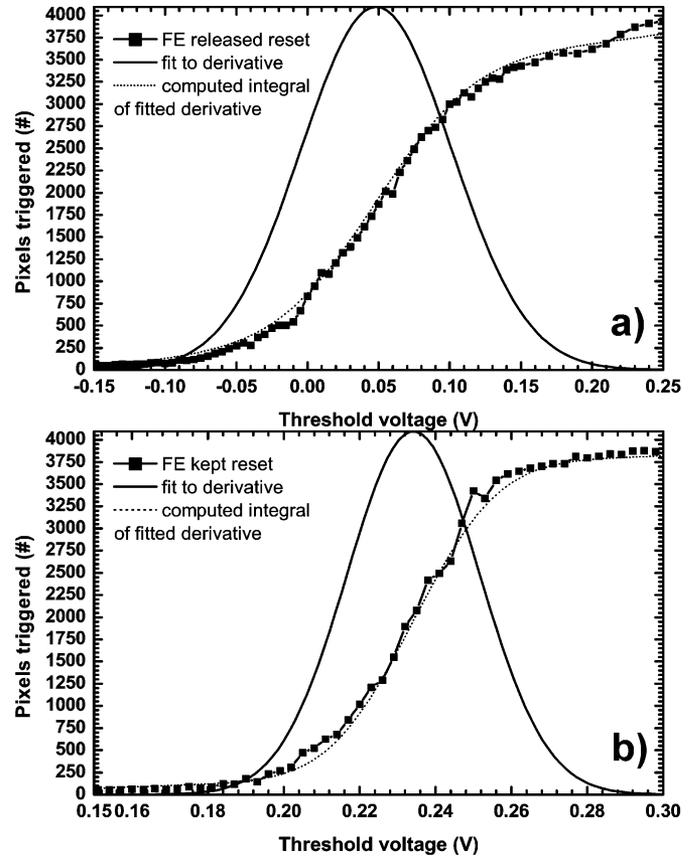


Fig. 8. Triggering threshold dispersions of the analog channels for the state of the front-end a) with both integrator and discriminator armed and b) when integrator is reset.

to the geometry of the VIP1 chip. The measured threshold levels seem to be acceptable for the first prototype. The corresponding plots showing the measured s-curves and Gaussian fits to the derivatives are presented in Fig. 8. All 4096 pixels were subjected to this test. The Gaussian model of threshold dispersions was assumed and fits were obtained with coefficients of determination $R^2 = 0.86$ and $R^2 = 0.94$ for released reset and kept reset, respectively. The mean and sigma values of the Gaussian curve were $\mu = 0.048 \text{ V} \pm 0.003 \text{ V}$ and $\sigma = 0.053 \text{ V} \pm 0.004$ for the release reset and $\mu = 0.2342 \text{ V} \pm 0.0007 \text{ V}$ and $\sigma = 0.0175 \text{ V} \pm 0.0007 \text{ V}$ for the kept reset conditions. The fits were performed to smoothed numerically calculated derivatives. The computed integrals of derivatives are also shown in Fig. 8 in order to show the deviations from original s-curves. Although, threshold dispersions match well with the assumed Gaussian distribution in the central part of the s-curves, a slight slope is visible in the region of saturation. Linear dependence was added in the fitting model to account for this underlying wide tail part of the threshold voltage distribution. Unfortunately, more detailed studies of this secondary effect were not considered practical due to the little statistics of chips available for test.

Next after the analysis of threshold dispersions, a test charge was injected into 119 pixels to simulate a hit pixel pattern. The arbitrarily selected pattern of pixel positions which had their test charge signals enabled is shown in Fig. 9. The sequence of bits

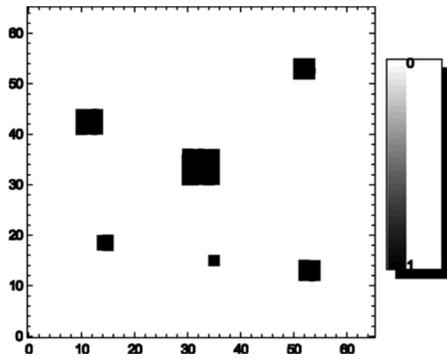


Fig. 9. Arbitrarily selected pattern composed of 119 pixels; positions at which test charge signal was injected.

corresponding to the selected pattern was shifted into the injection shift register, and then the integrators and the discriminators were enabled in all pixels. The external threshold was initially set to 0. Then a voltage step was applied to the test charge capacitors. The step was achieved by switching between two externally defined voltages swapping the control signal applied to the switches. The readout was done and addresses of the hit pixels and the corresponding signal amplitudes were recorded. The number of pixels detected as 'hit' in the sparsified readout and the mean value of the analog response are shown in Fig. 10. Almost all pixels from the programmed pattern are above the threshold at an injected voltage step of about 0.35 V. Clearly linear dependence of the measured mean response on the injected signal starts to be visible above 0.45 V. By knowing the slope of the response, the designed gain of the integrator, and that 100 ADC units equals 35 mV, the value of the test capacitor can be estimated at 0.29 fF. Consequently the 0.35 V threshold level corresponds to about 580 e⁻, which is in a very good agreement with the value calculated from the 3D geometry for the test capacitor and the mean threshold obtained in the threshold scan. There were some pixels sporadically popping up in other places outside the programmed pattern; however they were not included in the current analysis.

D. Tests of Time Stamping and Sparsified Readout

The unmodified pattern of 119 pixels was used for tests of time stamping and multiple sparsified readouts. The tests were a bit complicated since the dynamic flip/flops in the injection chain could not hold the programmed value for time longer than a few tens of microseconds. The cleanness of the programmed injection pattern depended strongly on the digital power supply voltage, showing best operation for voltage from 0.2 V to 0.3 V below the nominal 1.5 V power supply. The goal was to obtain unbiased results, so it was decided to set the Gray time stamping counter before each test, empty the whole matrix from any accidental hits, then shift in the injection pattern and arm all integrators and discriminators. After this setup, the analog test charges were injected immediately. The voltage ramp used for analog time stamping, was started with an appropriate delay to achieve the desired value at the moment of injection of the test charges. The duration of the ramp was shortened from the default 1 ms to about 100 μ s for the sake of these tests. The number of tests was equal to 32, allowing examination of all possible digital

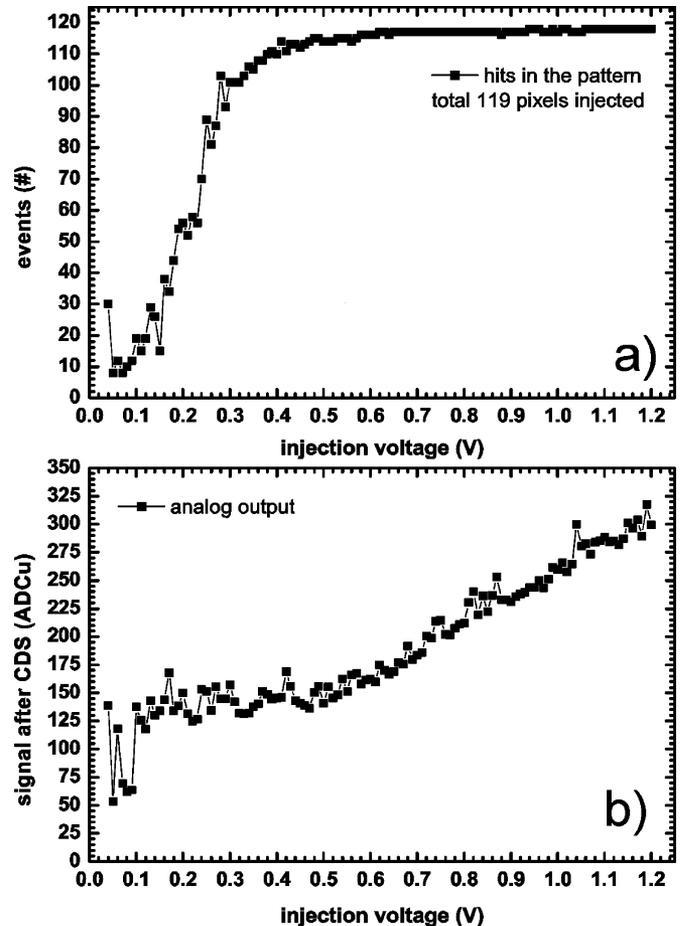


Fig. 10. Response to the test charge injected to the preselected positions in the pixel matrix; a) number of pixels detected as 'hit' in the readout, b) mean analog signal from each measurement. The pattern of position for injection shown in Fig. 9 was used.

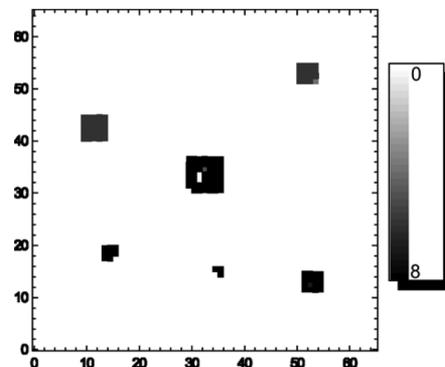


Fig. 11. Positions obtained from 8 sparsified readouts when charge was injected at different states of Gray code time stamp counter. The pattern shown in Fig. 9 was used for test charge injection.

time stamps. The tests were performed in 4 groups of 8. The results of one set of tests are shown in Fig. 11. Pixels triggering in all tests are marked in black. The white color corresponds to the non-responding pixels; pixels responding only in some tests are marked in gray. The same test charge was injected in each test. It is noticeable that the results are not ideal. However the main source of problems is attributed to the shift register used for programming the injection pattern. The shift register was

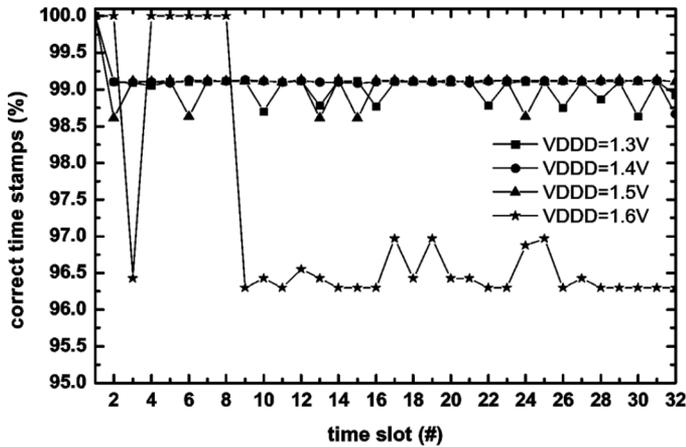


Fig. 12. Correctness of digital time stamping information for pixels with injected test charge from the 119 pixel test pattern for different digital power supply voltages.

built with dynamic D flip/flops that were not holding patterns due to high leakage currents of transistors in the off state. The results of these comparisons between the acquired values and the states set in the counter are plotted in Fig. 12. For the codes latched in the acquisition, only 1 to 2 pixels show faulty time stamping codes when the digital power supply is set to 1.3 V or 1.4 V. The performance of sample/hold circuits was compromised by strong leakage currents on tier-B. The stored values in the analog time stamping were strongly affected by the time elapsed between sampling and readout. As a consequence of the leakage, the stored values that were measured are attenuated by a factor of close to two, and large dispersions from cell-to-cell were observed. The results of the analog time stamping tests are not shown in this paper.

IV. LESSONS LEARNED AND FUTURE PLANS

FD-SOI processes are best suited for fast low-power digital circuits. Their appropriateness for analog circuits is questionable due to typically poor transistor matching. The reasons include several effects, like thermal separation of transistors, mechanical stresses present in silicon islands seated on the thick layer of BOX, the flow of ions and their accumulation in places where transistor may be affected, charging of oxide volumes that may result from accumulation of doses of ionizing radiation or may happen in the processing when, for example plasma etching is extensively used. Obtaining reliable body contacts for transistors is also problematic, resulting in problems like hysteresis or noisy fluctuations of the channel current as the channel is modulated by varying bulk potential. It is believed that there are strong limits on the maximum resolution of analog-to-digital conversion circuits that can be obtained in the SOI technology, see for example [16], [17]. In 3D-IC assembly, the tiers are separated by oxide, resulting in the possibility of strong detrimental interstrata capacitive coupling.

The via-last approach in 3D-IC assembly can be advertised as opening an outlook on heterogeneous integration of different materials, processing technologies and functional components. Since the wafer bonding is not supposed to provide interstrata electrical connection, the bonding surfaces can be prepared only

for mechanical attachment. Nevertheless, introduction of bulk process wafers in the 3D stack represents already some difficulty as it requires development and/or use of isolation between the inserted TSVs and the walls of cavities etched in wafers. Additionally, since the 3D-IC assembly is a complex process, it is unlikely that 3D assembly houses will be willing to accept mixed wafers, unless certain criteria for 3D integration are met. It will be harder to agree between users upon the use of 3D stack components satisfying needs that can be different for each user. It may turn out that the optimized stacking will be available only through dedicated engineering runs, excluding MPW-type options. Thus the cost of the heterogeneous 3D stacking in prototyping may be unacceptably high. The 3D-IC process run by MIT-LL included 3 wafers fabricated in an identical 0.18 μm FDSOI process. This allows design of a full electronics processing chain; however the detector wafer has to be attached separately.

Commercial CMOS bulk processes offering via first TSVs as one of the fabrication steps seem to be an attractive alternative. We plan to transfer the VIP chip design from the MIT-LL FDSOI process to the Tezzaron/Chartered 0.13 μm bulk CMOS process with TSV formed after front end of line processing in the near future [18], [19]. The Tezzaron/Chartered process is well suited to analog circuit design, offering deep N-wells, MiM capacitors, and multiple threshold voltage transistors. Another advantage is a full set of commercial tools to support the design in this process. The TSVs are only 6 μm deep in the Tezzaron/Chartered process and the wafers are aggressively thinned to expose TSVs that are initially buried. The 3D stacking is done using a Cu-Cu bond on a prepared surface. However, in the meantime, an improved version of the VIP1 chip was submitted to the 3DM3 run at MIT-LL in October 2008. The process was virtually downgraded by intentionally designing circuits with transistors several times larger than the minimum dimensions allowed in the process. Both width and length of transistors were scaled. Some other steps were also undertaken in order to improve the analog performance. Implementation of these self-imposed yield improvement rules resulted in the increase of the pixel size to $30 \times 30 \mu\text{m}^2$.

V. CONCLUSIONS

Industry is making rapid progress in developing 3D integrated circuits. The HEP community is beginning to respond with new initiatives to explore this technology. The demonstrator VIP1 chip is a multifunctional device serving as proof of the 3D-IC principle for HEP. It is a 64×64 array whose architecture allows an easy expansion to 1000×1000 . There are 175 transistors in each $20 \mu\text{m}$ square pixel. The active thickness of the 3 combined layers is only 22 μm . A choice will be made for future applications that will select analog or binary readout and one of the two time stamping approaches. The current chip has provision for a test input signal which can be expanded to include a pixel disable circuit with little extra circuitry. The power dissipated by a full scale version of this chip is consistent with the air cooling requirements of the ILC pixel vertex detector. The support logic around the perimeter of this chip is small. In future designs, this can be reduced further.

The VIP1 chip was submitted for fabrication in the 3DM2 0.18 μm FD-SOI multi-project run at MIT-LL in October 2006. This VIP1 chip was extensively tested. The tests showed correct functional operation of the structure. Although the immediate successor of the VIP1 chip was submitted in the FD-SOI process to the next run opened by MIT-LL, it is strongly believed, seen current results and exploring available expertise, that the FD-SOI approach has the shortcomings for mixed-mode design. The problems are linked inherently to the technology aspects not being limited to this or another vendor. Better performance, on the side of the analog circuitry, can be expected from bulk CMOS processes. The translation of the VIP1 design into the commercial bulk CMOS process is underway. The successive submissions are planned in a commercial 3D bulk 0.13 μm CMOS with integrated TSVs.

All pads were located on tier-C in the VIP1 chip. Future effort will address generation of the pads designated for the fusion bonding of the detector on one side of the 3D stack while the pads providing all electrical connection for the operation of the readout circuit will be patterned on the opposite side of the device. This would allow fabrication of a real 4-side buttable device.

3D-Integrated Circuit technology enables higher densities of electronic circuitry per unit area without the use of nano-scale processes. Ideally, heterogeneous wafers (different foundries or different process families) may be combined with some 3D integration methods, leading to the optimization of each tier in the 3D stack. It is advantageous for mixed mode design with precise analog circuitry because processes with conservative feature sizes typically present lower process dispersions and tolerate higher power supply voltages, resulting in larger separation of a signal from the noise floor. The economy driven criteria will show if this path can be enabled in future.

ACKNOWLEDGMENT

The authors would like to acknowledge the valuable contributions of M. Turqueti and R. Rivera from the Computing Division of FERMILAB for the development of the data acquisition system and help in tests of the VIP1 chip. Thanks go to MIT-LL, particularly to B. Tyrrell, for fruitful discussions and for revealing details of the 3D-IC integration and FDSOI process. We would like to thank A. Dyer from the ASIC testing group of FERMILAB for his dedicated and excellent technician support. The VIP1 chip was designed and fabricated in MIT Lincoln Laboratory's 3D-IC process, funded under the DARPA Advanced Microelectronics Technology Development Program managed by Dr. D. Radack.

REFERENCES

- [1] M. Vertregt, "The analog challenge of nanometer CMOS," in *Proc. IEEE Int. Electron Devices Meeting Tech. Dig.*, San Francisco, CA, Dec. 11–13, 2006, pp. 1–8.
- [2] P. Garrou, C. Bower, and P. Ramm, *Handbook of 3D Integration, Technology and Applications of 3D Integrated Circuits*. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co./KGaA, 2008.
- [3] M. Puech, J. M. Thevenoud, J. M. Gruffat, N. Launay, N. Arnal, and P. Godinat, "Fabrication of 3D packaging TSV using drier," in *Proc. Symp. Design, Test, Integration and Packaging of MEMS/MOEMS*, Nice, France, Apr. 9–11, 2008.
- [4] Q.-Y. Tong, "Room temperature metal direct bonding," *Appl. Phys. Lett.*, vol. 89, pp. 182101–182103, 2006.
- [5] Q.-Y. Tong, G. Fountain, and P. Enquist, "Room temperature SiO₂/SiO₂ covalent bonding," *Appl. Phys. Lett.*, vol. 89, pp. 042110–042112, 2006.
- [6] J. A. Burns, B. F. Aull, C. K. Chen, C.-L. Chen, C. L. Keast, J. M. Knecht, V. Suntharalingam, K. Warner, P. W. Wyatt, and D.-R. W. Yost, "A wafer-scale 3-D circuit integration technology," *IEEE Trans. Electron Devices*, vol. 53, pp. 2507–2516, Oct. 2006.
- [7] C. K. Chen, N. Checka, B. M. Tyrrell, C. L. Chen, P. W. Wyatt, D. R. W. Yost, J. M. Knecht, J. T. Kedzierski, and C. L. Keast, "Characterization of a three-dimensional SOI integrated-circuit technology," in *Proc. IEEE Int. SOI Conf.*, Hudson River Valley, NY, Oct. 6–9, 2008, pp. 109–110.
- [8] B. Rajendran, R. S. Shenoy, D. J. Witte, N. S. Chokshi, R. L. De Leon, G. S. Tompa, and R. Fabian, "Low thermal budget processing for sequential 3-D IC fabrication," *IEEE Trans. Electron Devices*, vol. 54, no. 4, pp. 707–714, Apr. 2007.
- [9] M. Connell, M. Grady, P. Oldiges, D. Onsongo, M. Passaro, W. Rausch, P. Ronsheim, and D. Siljeborg, "Impact of mobile charge on matching sensitivity in SOI analog circuits," in *Proc. IEEE/SEMI Advanced Semiconductor Manufacturing Conf.*, Stresa, Italy, Jun. 11–12, 2007, pp. 6–10.
- [10] Y.-C. Tseng, W. M. Huang, D. J. Monk, P. Welch, J. M. Ford, and J. C. Woo, "AC floating body effects and the resultant analog circuit issues in submicron floating body and body-grounded SOI MOSFET's," *IEEE Trans. Electron Devices*, vol. 46, pp. 1685–1692, Aug. 1999.
- [11] MITLL Low-Power FDSOI CMOS Process Design Guide, 2006.
- [12] J. E. Brau, M. Breidenbach, C. Baltay, R. E. Frey, and D. M. Strom, "Silicon detectors at the ILC," *Nucl. Instrum. Methods Phys. Res. A*, vol. A579, pp. 567–571, 2007.
- [13] R. Yarema, "3D circuit integration for vertex and other detectors," in *Proc. 16th Int. Workshop Vertex Detectors*, Lake Placid, NY, Sep. 23–28, 2007, p. 017.
- [14] K. H. Becks *et al.*, "The DELPHI pixels," *Nucl. Instrum. Methods Phys. Res. A*, vol. A386, pp. 11–17, 1997.
- [15] J. R. Hoff, A. Mekkaoui, D. C. Christian, S. Zimmerman, G. Cancelo, P. Kasper, and R. Yarema, "PreFPIX2 core architecture and results," *IEEE Trans. Nucl. Sci.*, vol. 48, no. 3, pp. 485–492, Jun. 2001.
- [16] B. M. Tenbroek, M. S. L. Lee, W. Redman-White, C. F. Edwards, M. J. Uren, and R. J. T. Bunyan, "Drain current mismatch in SOI current mirrors and D/A converters due to localised internal and coupled heating," in *Proc. 23rd Euro. Solid-State Circuits Conf.*, Sep. 16–18, 1997, pp. 276–279.
- [17] W. Redman-White, B. M. Tenbroek, M. S. L. Lee, C. F. Edwards, M. J. Uren, and R. J. T. Bunyan, "Analogue design issues for SOI CMOS," in *Proc. IEEE Int. SOI Conf.*, Sanibel Island, FL, Sep. 30–Oct. 3 1996, pp. 6–8.
- [18] R. S. Patti, "Three-dimensional integrated circuits and the future of system-on-chip designs," *Proc. IEEE*, vol. 94, no. 6, pp. 1214–1224, Jun. 2006.
- [19] R. S. Patti, "3D scaling to production," in *Proc. 3D Architecture for Semiconductor Integration and Packaging Conf.*, San Francisco, CA, Oct. 31–Nov. 2 2006.