The Front-end ASIC for the ATLAS Pixel Detector

K. Einsweiler, LBNL

Overview of FE specifications and design

History of ATLAS Pixel FE ASIC

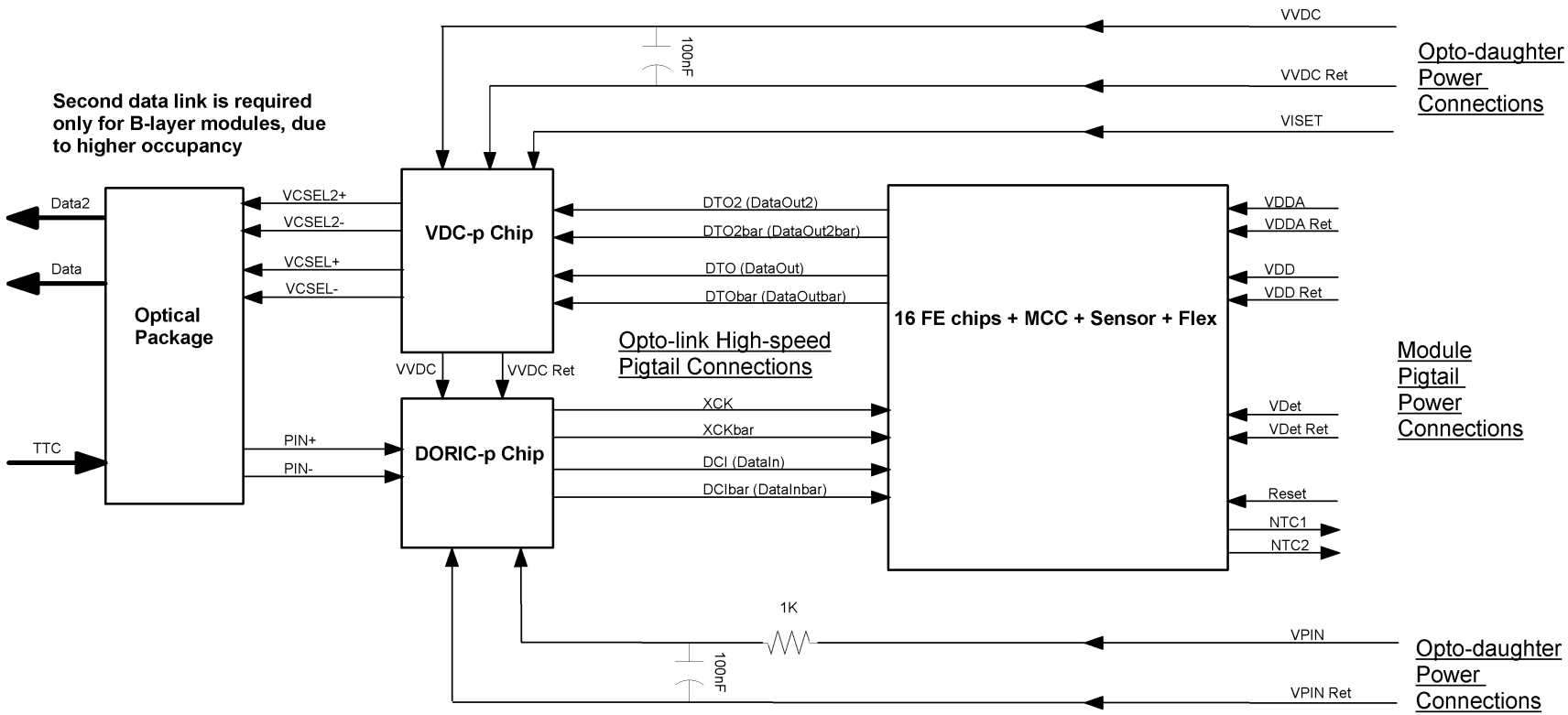
The first 0.25μ generation of the FE ASIC, FE-I1

Wafer probe results and yield issues

Second generation 0.25μ FE ASIC, FE-I2

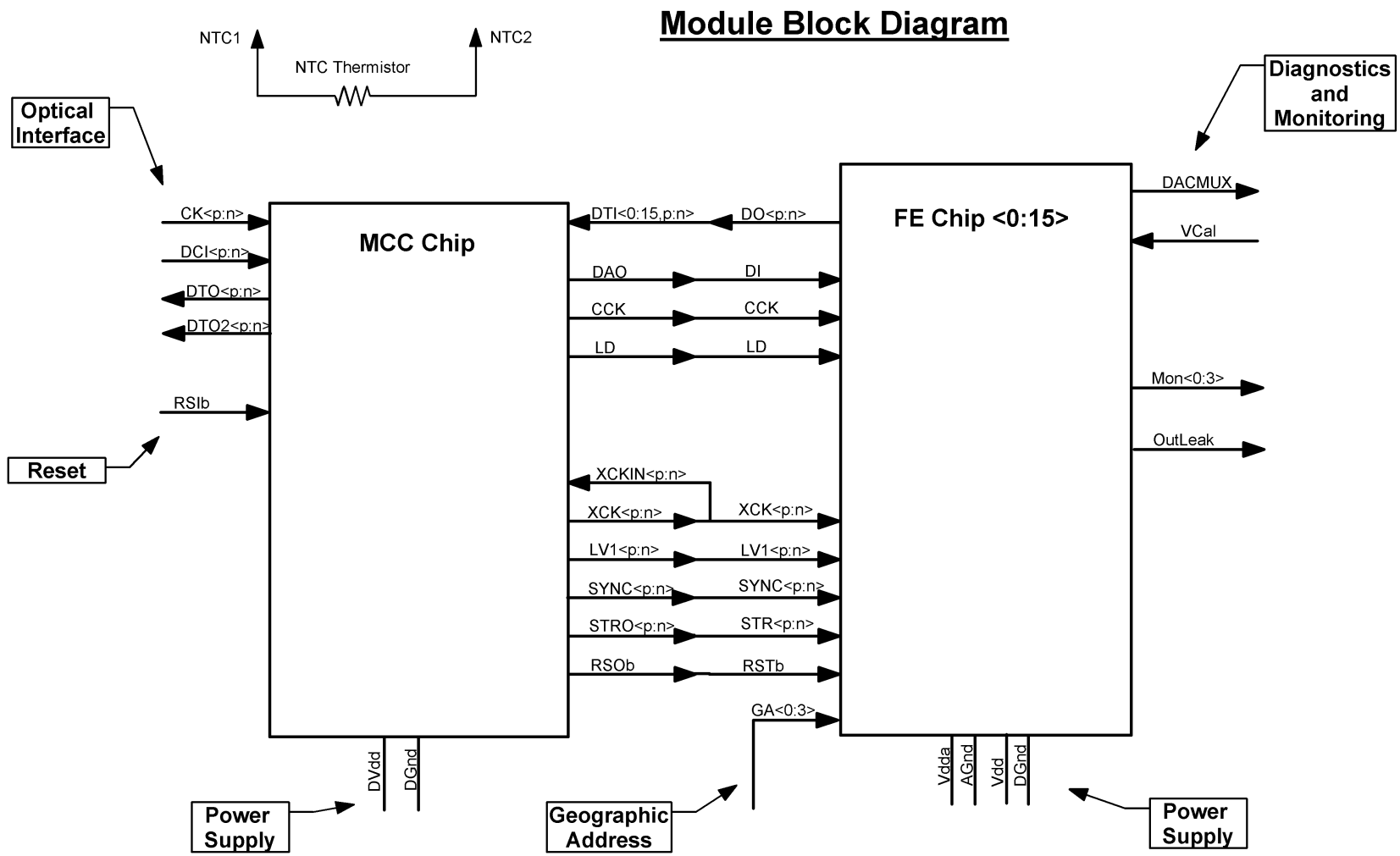
System Design of Pixel Module

Overall system architecture, including interconnections to opto-link and all power supplies:



- Optical package and DORIC/VDC mount on separate opto-card, up to 1m away.
- Module itself uses two LV supplies (analog and digital) and one HV bias supply.
- Communication between module and opto-card uses 3mA LVDS I/O

Block diagram of module itself:



- Two chip design, including a single controller and event-builder chip (MCC), and 16 front-end chips bumped to a single silicon substrate.
- Flex hybrid is used to provide interconnections above.

Features:

- Basic interface to the outside uses a 3-wire protocol (SerialIn, SerialOut, XCK), which maps onto the SCT opto-link protocol
- Basic interconnections between FE and MCC use bussed signals. Slow control will not operate when recording events, so it uses full-swing CMOS. All fast signals use low-swing differential “LVDS-like” signaling. Point-to-point signals use 0.5 mA drivers (FE chips only), external or bussed signals use 3.5 mA drivers.
- To provide enhanced speed and robust module design, the serial output lines are connected from the FE to the MCC in a star topology (16 parallel inputs on MCC).
- There are no remaining analog signals between MCC and FE at this time. All FE chips have internal current references and adjustment DACs to control analog operating points, as well as calibration.
- Architecture is “data-push” style: each crossing for which LV1 accept is present causes all FE chips to autonomously transmit back hit information for the given crossing. LV1 signal may remain set for many contiguous crossings to allow readout of longer time intervals. MCC merges such events together.
- Synchronization signal available to ensure FE chips label LV1 properly.

Requirements Summary

Power budget (actually current budget for 0.25μ):

- Present design of services and cooling based on typical and worst-case current and power analyses for DMILL chips. The FE-I should fit the same total current budget for the VDDA (analog) and VDD (digital) supplies.
- The present budgets for the digital supply are 25mA typical, 40mA worst-case. For the analog supply, they are 60mA (21μA/pixel) typical and 80mA (28μA/pixel) worst-case. The design of the low-mass power distribution system is very challenging, and we must stay safely within these budgets. With the reduced voltage used for the 0.25μ process, the total power and cooling are much less difficult issues than they were in the past.

Geometry:

- The active die area for the FE chip is 7.2 x 10.8 mm, of which 7.2 x 8.0mm is sensitive area for particle detection. The sensitive area of the FE chips must extend to the edge of the die along 3 sides, with all additional logic and I/O concentrated on the remaining side.
- Physics studies indicate that the pixels should be as narrow as possible in one dimension, and a 50μ pitch has been chosen as reasonably achievable. In the long direction, adequate resolution is obtained with a dimension of 300μ - 400μ. The present prototyping program has frozen the length at 400μ.

Basic FE Chip Geometry

- Agreement on pixel size was struck in Sept 96, in order to allow compatible, parallel detector and electronics development.
- The geometry adopted was 50μ x 400μ for the pixel size, with pixels arranged in 18 columns of 160 pixels per column.
- The geometry was mirrored between columns, so that inputs for pixels in column 0 and 17 are on the outside. All other inputs are paired. This gives us 9 column pairs, with a common digital readout in the center, and analog cells on the sides.
- The input pad geometry in the inner column pairs is then a double row of 50μ pitch pads. The metal pad is specified to be 20μ octagon, with a 12μ opening in the passivation for the bump-bonding.
- The cut die size must not extend beyond 100μ from the edge of the active area on three sides of the die. Hence, nothing outside of the pixel circuitry is allowed on three sides of the chip, to allow module construction.
- The bottom of the chip (all peripheral logic and I/O pads) are allowed 2800μ, making the total active die region 7.2mm x 10.8mm.
- An I/O pad structure of 48 pads, each consisting of a 100μ x 200μ wire-bond pad, and a group of 4 bump-bond pads for MCM-D applications, was frozen.
- For final modules, only the central 30 bond pads are available for connections due to mechanical envelopes. Other 18 pads are available for diagnostics.

Brief History: FE-B, FE-D, and FE-H

- Rad-soft prototyping delivered functional chips in 98 (FE-B, FE-A/C). FE-B (HP 0.8μ) demonstrated all basic ATLAS pixel performance goals in lab and testbeam.
- Submitted FE-D1 run, containing FE-D1 front-end chip, DORIC and VDC chips, and prototype MCC-D0 chip (plus test chips). Submission went out in July 99.
- FE-D1 suffered from minor design errors, and very poor yield in two circuit blocks. After considerable investigation, the low yield was related to technology problems (low rate of very leaky NMOS). Foundry never succeeded in isolating the problem, but proposed a series of special corner runs.
- Submitted FE-D2 run in Aug 00, with two versions of FE-D2. In one version, all design errors were fixed, but basic design (including dynamic logic blocks which suffered low yield) was left unchanged. Second version replaced low-yield blocks with static versions, and removed other circuitry (trim DACs) to make room. Run included full MCC-D2 (100mm^2) and new DORIC and VDC chips as well.
- Corner runs gave no new information on yield/technology problems. Yield on static chip was better, but still unacceptable. Work with this vendor was terminated.
- Began work on FE-H in Dec 99. Chip was almost ready to submit when we received notification of massive cost increases from Honeywell. With wafer cost of 20-30K$, effort was abandoned before actually building a complete pixel chip.
- The failure of both traditional rad-hard vendors left us with 0.25μ approach, based on commercial process and rad-tolerant layout. Major effort started in Sept 2000.

Design Methodology for 0.25μ Process

Begin with CERN/RAL design kit for IBM CMOS6SF process:

Similarity between IBM and TSMC design rules:

- Implement common design environment where a single design can be streamed out and submitted to either vendor for fabrication.
- Cost of both is similar, but have frequent access with 8-10 week turnaround to TSMC foundry via MOSIS.
- In case of problems with one vendor, we have a back-up. Although TSMC has not been as thoroughly qualified for radiation as IBM, many tests have been performed by FNAL group, and we have also irradiated prototypes to 60MRad and 120MRad.

Updated CERN/RAL Standard Cell Library:

- Minor layout modifications made to provide compatibility with TSMC rules.
- Low noise mixed-mode design goals suggested separation of digital ground and substrate connection in the cells.
- We have also separated analog ground and substrate connection in individual analog blocks. Two substrate connections are joined at bottom of column, and both are connected to analog ground at the pad frame.
- Prefer to “over-fill” cells in non-routing layers (poly and active), so that problems do not arise later during integration.

Feature List for FE-I

Design is logical evolution from FE-D and FE-H designs.

Analog Front-end (designed for VDDA=1.6V operation):

- The FE uses a DC-feedback preamp design which provides excellent leakage current tolerance, close to constant-current return to baseline for TOT, and very stable operation with different shaping times.
- It is followed by a differential second amplifier stage, DC-coupled to the preamp. The reference level (VReplica) is generated in the feedback block, and should match the DC offset of the preamp with no input. The threshold control is applied using two currents to modify the offsets on the inputs to the second amplifier stage, allowing a large range for threshold control.
- The two-stage amplifier is followed by a differential discriminator which provides the digital output sent to the control logic.
- The control logic provides a 5-bit threshold trim capability in each pixel, plus a 5-bit feedback current trim capability for tuning the TOT response. There are four control bits, including Kill (shut down preamp), Mask (block entry of hit into readout logic), HitBus (enable output to global FastOR) and Select (enable injection of charge for testing). The HitBus bit also controls the summing of a current proportional to the feedback plus leakage current in the preamp, allowing monitoring of the feedback current, and of the leakage current from the sensor.

- A global FastOR net is created using all pixels enabled for this type of readout, and provides a self-trigger and diagnostic capability.
- All critical bias currents and voltages on the chip are controlled by internal DACs. There are 12 8-bit DACs for the analog front-end, and an additional DAC for the charge injection. The current DACs are referenced to an internal CMOS current reference, and the DAC values are loaded from the Global Register, and controlled via the Command Register.

Digital Readout (designed for VDD=2.0V operation):

- It uses an 8-bit Grey-coded 40 MHz differential “timestamp” bus as a timing reference throughout the active matrix. All pixels measure their leading and trailing edge timing by asynchronously latching this reference in RAMs.
- Hits (address plus LE/TE timing) are transferred from the pixels as soon as the trailing edge occurs, using a shared bus structure in the pixel column pair. This bus operates at transfer rates up to 20 MHz in order to meet our requirements. Differential signal transmission and sense amplifiers are used to achieve this.
- Significant buffering is provided in the end of column region for hit storage during the L1 latency (up to 6.4μs in this chip). Sixty four buffers are available for each column pair (one for each five pixels). The coincidence with the L1 trigger is performed in this buffer. Hits from rejected crossings are immediately cleared.

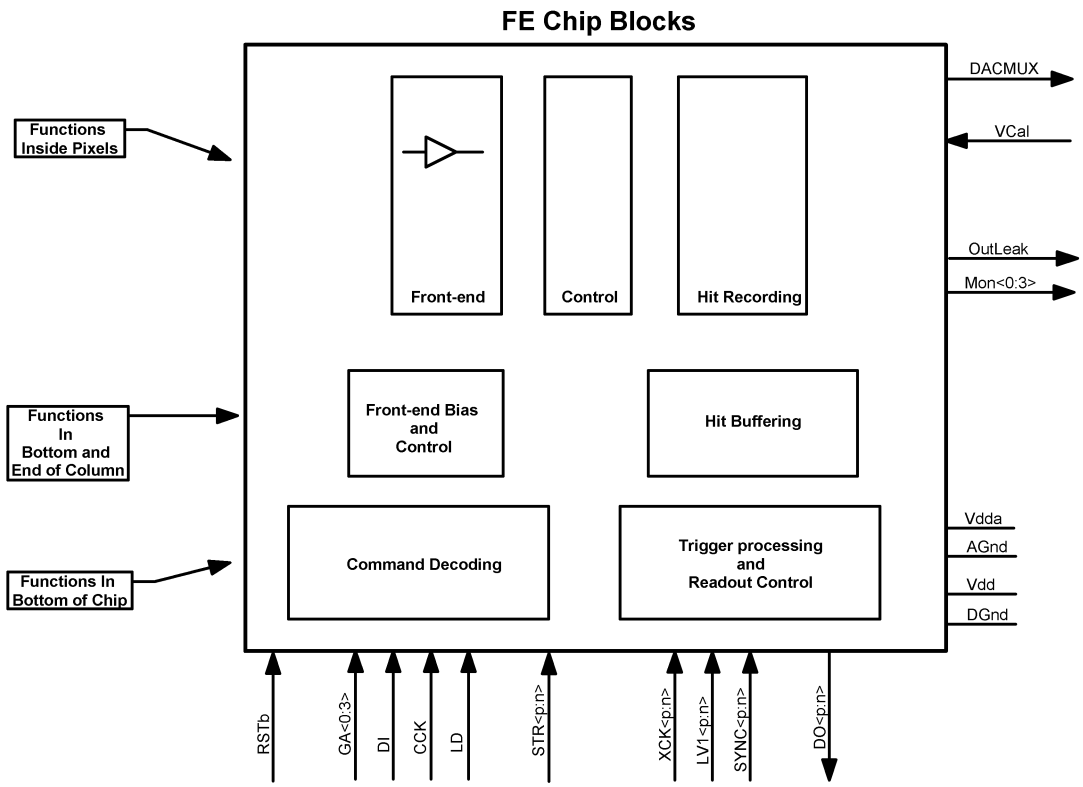
- A readout sequencer stores information on up to 16 events pending readout. As soon as the output serial link is empty, transmission of a new pending event begins. Essentially, sending a L1 trigger corresponds to making a request for the all hits associated with the corresponding beam crossing, which are then pushed off the FE chip to the MCC.

Control Logic:

- Global control of the chip is implemented using a simple command protocol. A Load signal controls whether input bits are interpreted as address and control, or as data. There is a 20-bit Command Register. Individual bits in this register implement specific commands (e.g. ClockGlobal, WriteGlobal, ReadGlobal).
- A Global Register, consisting of 202 bits, controls Latency, DAC values, enabled columns, clock speeds, and several other parameters. This register is implemented as a shift register and a shadow latch with full readback capability. The shadow latch is SEU-tolerant since it contains critical configuration information.
- A Pixel Register which snakes through the active array provides access to the 14 control bits in the pixel (Select, Mask, HitBus, Kill, FDAC<0:4>, TDAC<0:4>). Readback capability is supported by transferring FF information back into the long shift register for readout. The 14 latches in each pixel are SEU-tolerant.
- Each chip on a module is geographically addressed, and its identity is controlled by external wire-bonds to avoid confusion. A broadcast mode is also supported.

FE-I Block Diagram:

Basic FE block diagram, expanded from module diagram:

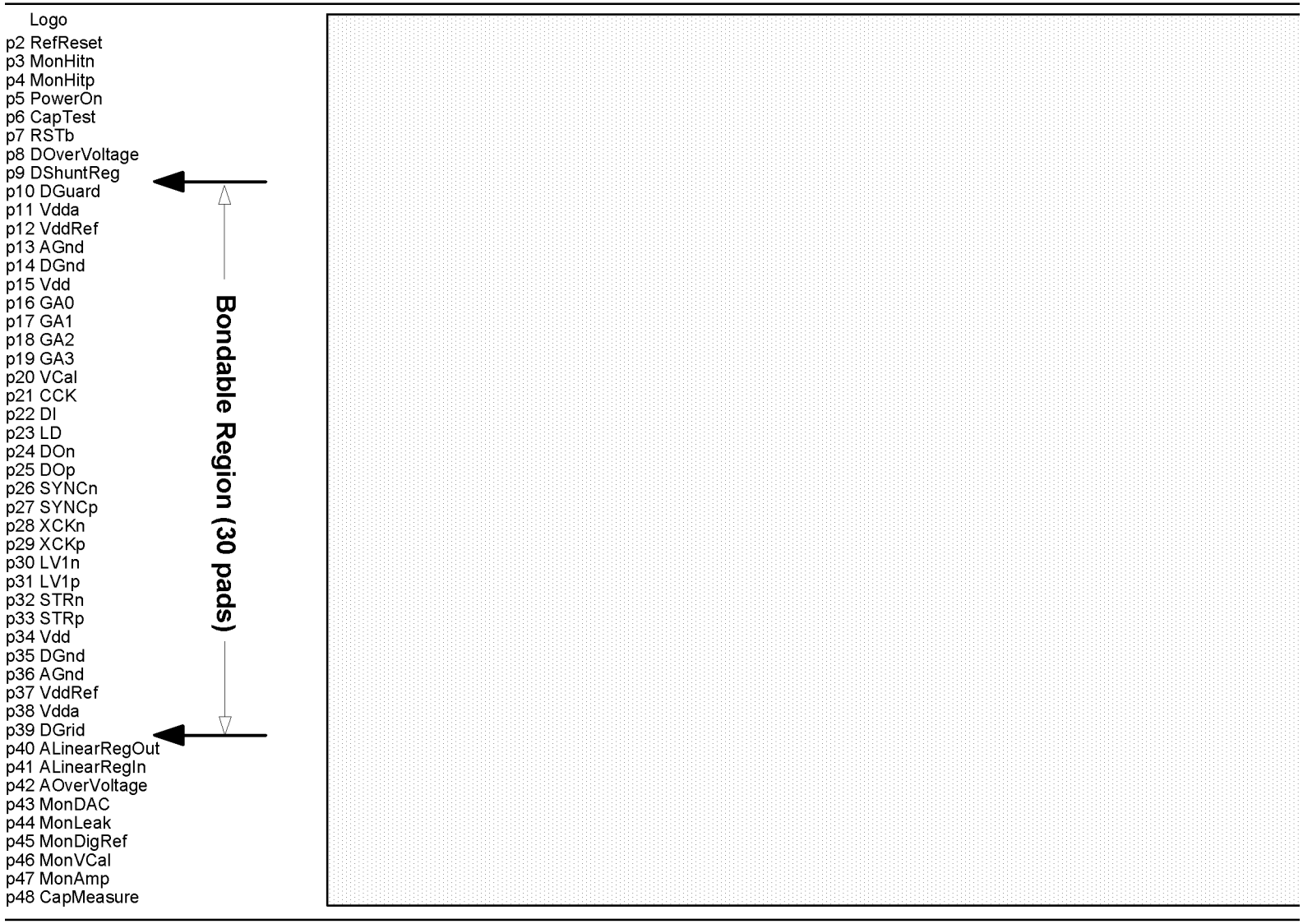


- Within the pixel, there is the front-end (preamp/ discriminator), the control block, and the readout block.
- Just below the active pixel matrix is the biassing and control for the front-end blocks, and the data formatting and buffering for the readout blocks.
- Finally, there is the overall readout control and the command decoding.

- Basic Digital I/O shown on bottom: 4 CMOS inputs for control (RSTb, DI, CCK, LD), and 4 fast, differential I/O's for timing and readout (XCK, LV1, SYNC, DO).
- Calibration and monitoring are shown on the right. A fast, differential strobe (STR) supplies calibration timing. An analog voltage input (VCal) provides external calibration. Monitoring pins include FastOR, DACs and test pixel output.

FE-I Pinout and Geometry

Sketch of pin assignments and overall geometry of die:



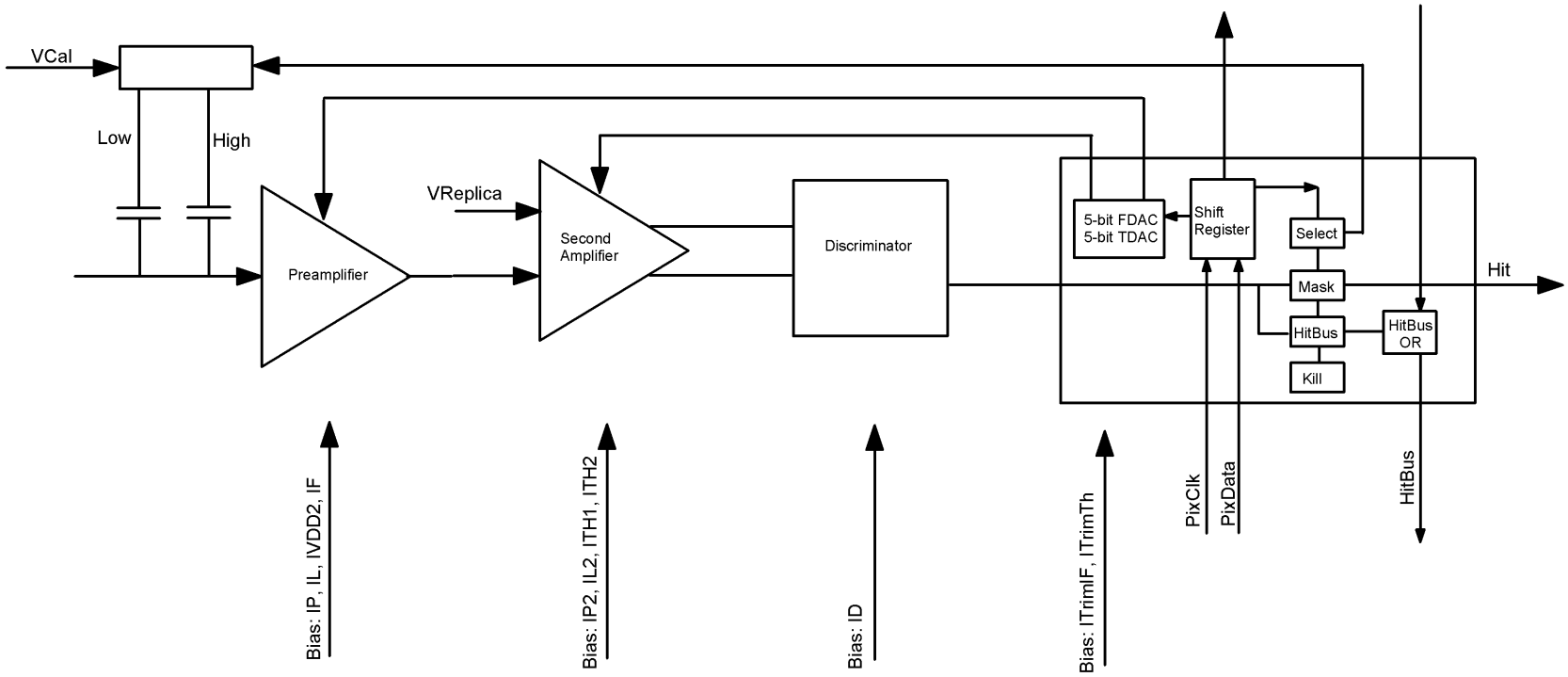
Front-end, biassing and control

Summary of the requirements:

- A nominal capacitive load of 400 fF is expected, roughly half to ground (parasitic) and half to the nearest neighbors (inter-pixel). Good performance should still be obtained with loads above 500fF. The n^+ on n-bulk detectors provide negative signals.
- Pixels are oriented to maximize signal and efficiency (minimize charge sharing).
- The outer layers should be 280μ silicon, and the B-layer should be 200μ.
- The worst-case signal after the lifetime dose of 10^{15} n-equiv/cm^2 is about 10Ke with 600V bias. We propose to operate at the 600V bias at the end of the detector lifetime, and have real prototype experience to show that this works well.
- This leads to an in-time threshold requirement of about 4-5Ke. This requirement is defined using a maximum timewalk relative to a large reference charge (100Ke) of 20ns, in order to allow for additional timing dispersion between channels on a chip and module, as well as between modules. This could be achieved by for example setting a 3Ke threshold, and having the required overdrive for a timewalk of 20ns be less than 2Ke. This is the most challenging requirement for our front-end.
- Noise should be less than 450e and threshold dispersion less than about 200e, leading to an overall threshold “variation” of less than about 500e.

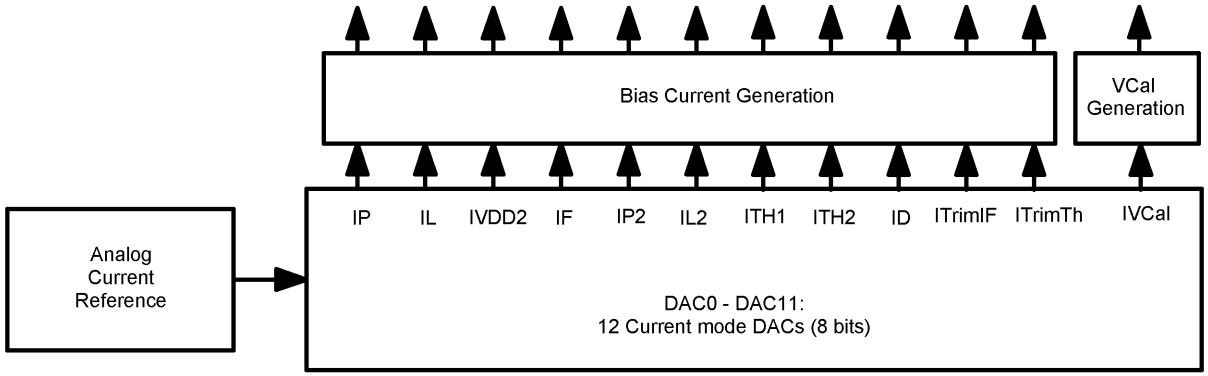
- Leakage current tolerance should be at least 50nA per pixel, without significant changes in operating performance, and independently achieved for each pixel.
- Noise occupancy should be less than 10^{-6} hits/crossing/pixel.
- Crosstalk between neighboring pixels should be less than 5-10%, where this is defined as the ratio between the threshold and the charge which must be injected into a pixel to fire its neighbors.
- A double pulse resolution of 2μs is required for the outer layers, and 0.5μs for the B-layer, in order to achieve our total deadtime requirements.
- It is required to provide binary readout of each pixel, but a modest analog resolution (4-5 bits) is very desirable if it can be achieved without a large impact on the other performance specifications.
- A threshold range of at least 0 - 6Ke is needed.
- Two calibration injection capacitors of roughly 5-10fF and 30-40fF should be included in each pixel, giving a low range for threshold and noise measurements and a high range for charge calibration, timewalk, and crosstalk measurments.
- We have no evidence that real diode input protection is needed for the preamps. In all present chips, no explicit input protection is provided, and no identifiable problems have been observed.

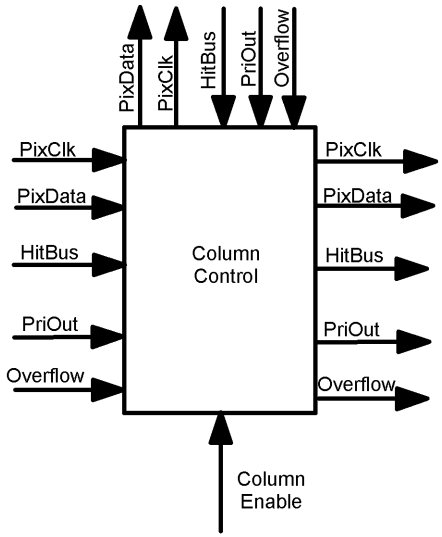
FE and Control Blocks:



- Preamp has roughly 5-10fF DC feedback design, and 15ns risetime. Input transistor is a PMOS with W/L of 25.2/0.6μ, and operates at about 8μA bias.
- Feedback circuit generates reference voltage for a differential second stage (VReplica). Feedback current is 2nA for a 1μs return to baseline and 20Ke input.
- Threshold control operates by modifying the offsets at the input of the second amplifier. Second amplifier bias is about 4μA.
- Discriminator is DC-coupled and differential. It uses a bias current of about 5μA.

FE Biassing and Control Blocks:





- A reference circuit is used to supply a 4μA reference current to the current mode DACs (1μA/bit, but two LSB generated by simple mirrors).
- The current mode DACs are a rad-tolerant 8-bit design with good linearity. A 9-bit version is used for the VCal calibration DAC.
- The biassing circuits are located directly on top of the DACs, and mirror the current down so that 64 DAC counts provides the nominal bias current. The biasses are distributed as V_{gs} voltages, with matching mirrors in each pixel.
- A single column enable bit controls the major operations of a column pair. It allows bypassing a column pair in the pixel shift register chain, bypassing a column pair in the HitBus FastOR net, plus bypassing the sparse scan readout and buffer overflow generation of a column pair.

Digital readout

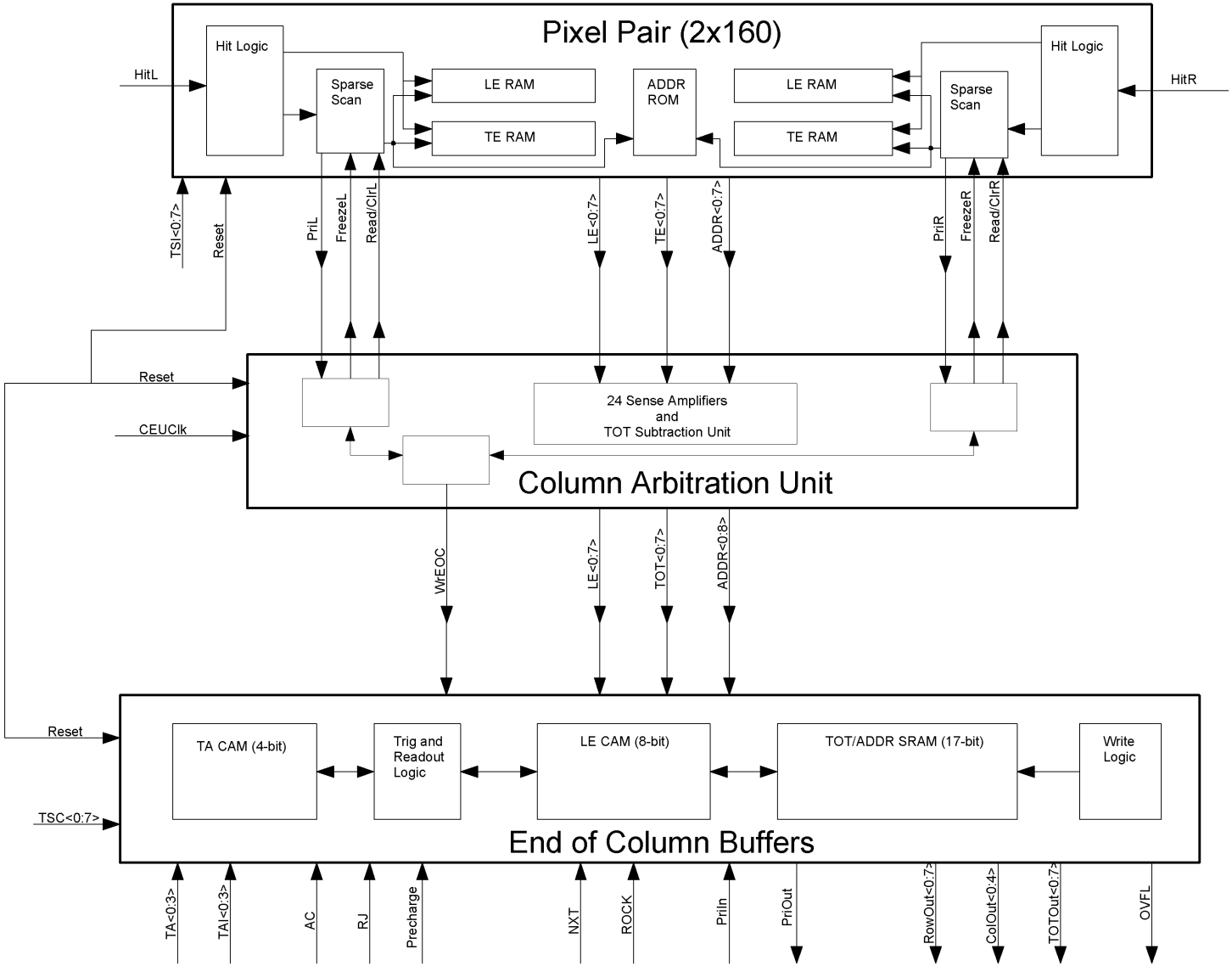
Summary of the requirements:

- Make a unique association of each hit pixel with a 40 MHz beam crossing.
- Store hits in pixel array for L1 latency period, which can extend up to 3.2μs.
- Make a modest TOT measurement by counting time differences between leading and trailing edges in 40 MHz units.
- Simulations for the current architecture exist, driven by the full GEANT simulation of ATLAS. This suggests that the current architecture needs to operate with a 20 MHz column clock rate and have 25 buffers per column pair in order to provide safe operation of the outer layers. The B-layer requirements are more stringent, and require something like 40 buffers.
- There is only a single error condition which occurs, namely overflow of the EOC buffers. In the case where the EOC buffer block in a given column pair overflows, hits are lost until a free buffer exists, and the error condition is stretched to cover a full L1 latency (covering all possible events which could have lost hits due to this condition). The error status is then transmitted in the EOE word.

Specifications:

- Clock duty cycle specified to be between 40% and 60% (high phase lies between 10ns and 15ns, or nominal +/- 2.5ns).

Block diagram of the basic column-pair readout:

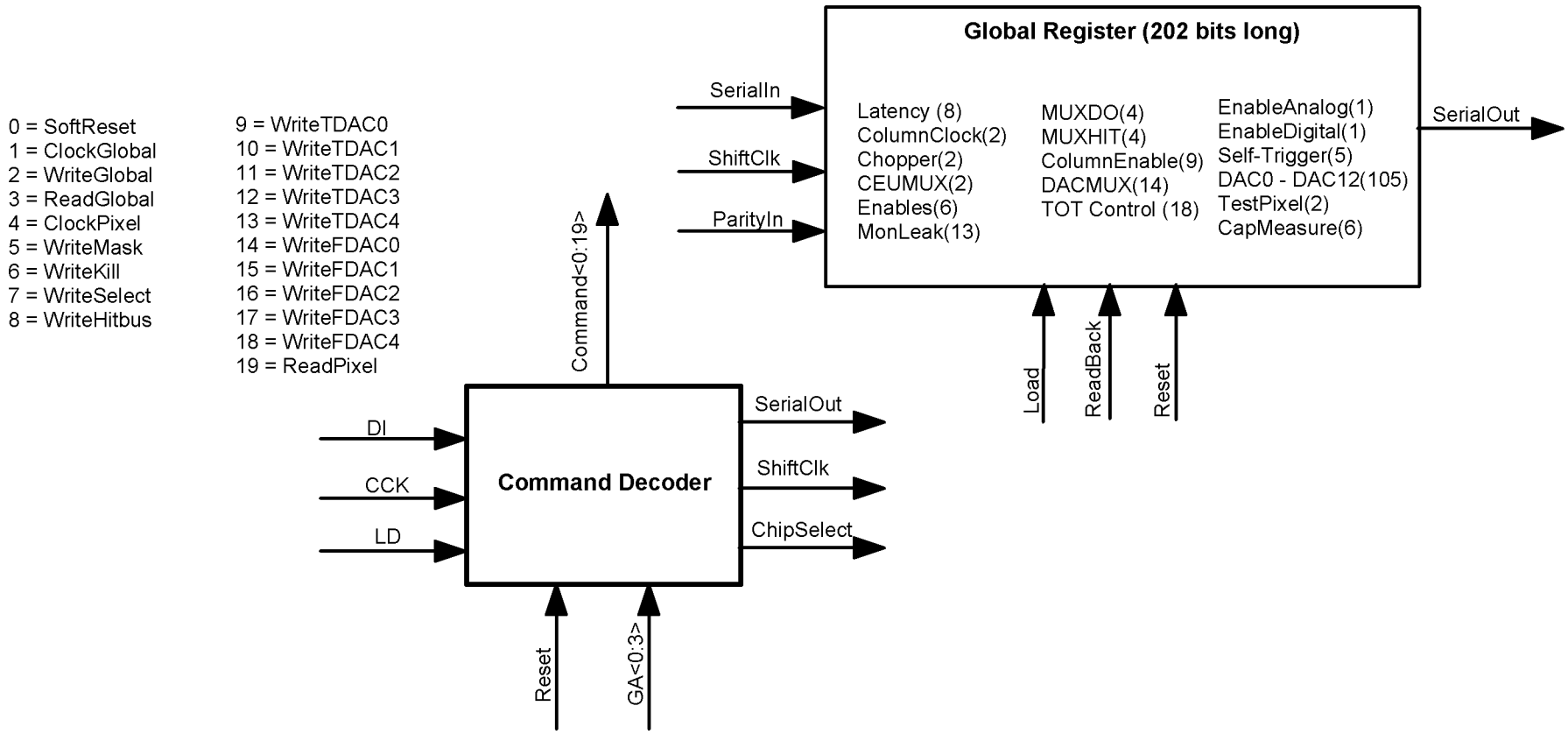


3

Summary of basic steps in readout of pixel data:

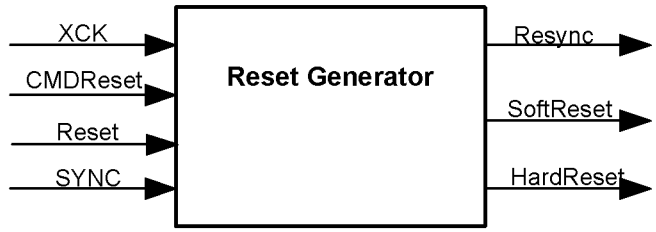
- Transfer hit information (LE and TE timestamp, plus pixel row address) into an EOC buffer. This operation begins when data is complete (after discriminator trailing edge). The transfer of hits from a column pair is synchronized by the CEU in the bottom of column, which operates at a speed of 5, 10, or 20MHz.
- Hit information is formatted by the CEU. Formatting includes TOT calculation (TE-LE subtraction) in all cases. Optionally, a digital threshold may be applied to TOT, a timewalk correction may be applied (write hit twice if below correction threshold, once with LE and once with LE-1), or both. These operation are pipelined to minimize deadtime, but EOC writes cannot occur faster than 20MHz.
- Hit information is written to the EOC buffer, and waits there for a corresponding LVL1 trigger. If a trigger arrives at the correct time (checked using LE timestamp of hit), the data is flagged as belonging to a particular 4-bit trigger number. Otherwise it is reset and the buffer is freed.
- Once the chip has received LVL1 triggers, the trigger FIFO will no longer be empty. This initiates a readout sequence in which the EOC buffers are scanned for the presence of hits belonging to a particular trigger number. If hits are found, they are transmitted to the serializer. After all hits for a given trigger number have been sent, an End-of-Event word is appended to the data stream.
- All of these operations occur concurrently and without deadtime, with all column pairs operating independently and in parallel.

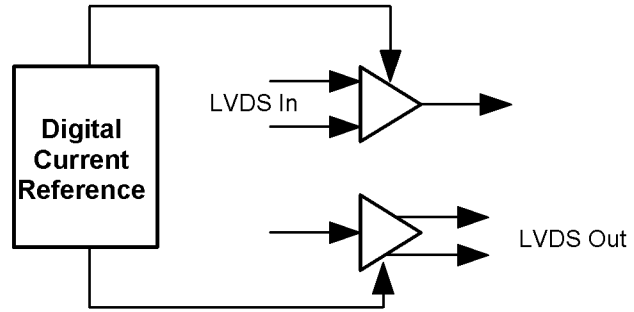
Block diagram of Command Decoder and Global Register:

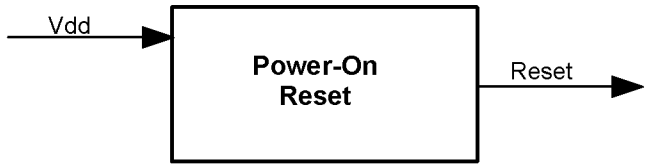


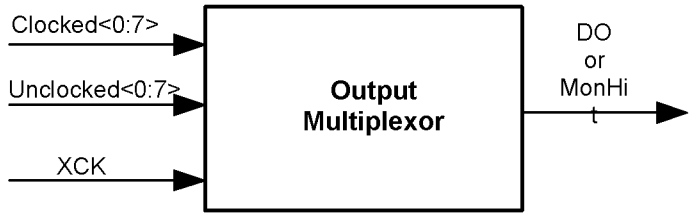
- Simple command protocol, based on a 5+20-bit command field, after which LD goes high and associated data may be transmitted. This supports 20 different, independent commands.
- Global Register controls overall operation of FE chip. Because of the critical importance of its bits, the actual latched values are stored in special SEU-tolerant FF. First measurements with 0.25μ prototypes indicate the SEU flip rate for these bits is less than one flip per bit during the lifetime of ATLAS.

Block diagrams of remaining blocks in the periphery:







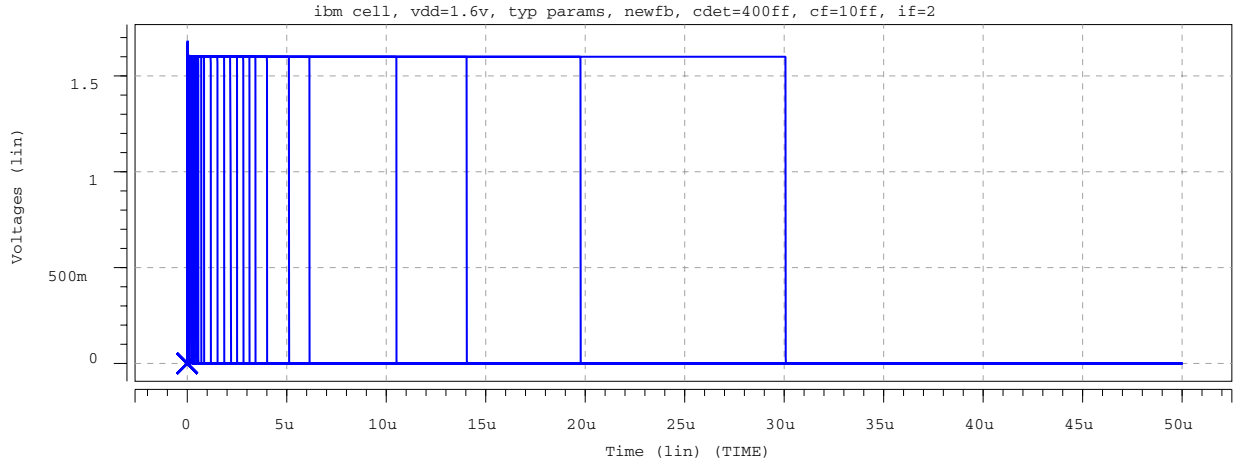


- There is a reset generator that either uses the external RESET pin, or the width of the SYNC input in XCK clocks, to generate three internal reset signals.
- The Resync makes sure that all FE on a module are using the same trigger number (it resets the trigger FIFO). The SoftReset puts the chip into the “empty” state for data, but does not alter any configuration information. Finally, the HardReset also resets all configuration information to zero.

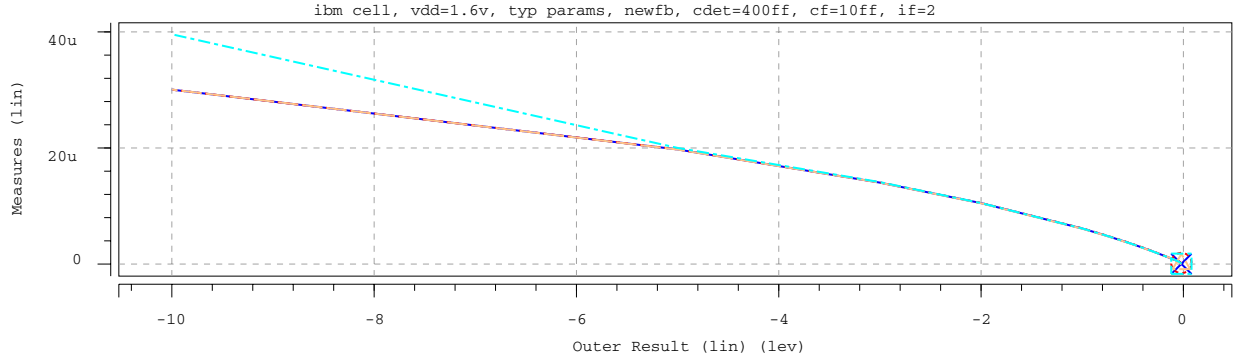
- There is a Power-on Reset which makes sure the digital part of the chip is powered up into the Reset state. This clears the Global Register, and in particular, clears the EnableReadout bit, blocking XCK distribution to digital readout circuitry. The result is low analog and digital power consumption (4mA analog, 8mA digital), suitable for performing basic module connectivity tests without requiring operation of liquid cooling system. Basic connectivity checks do not require analog power.
- There is a self-trigger generator, which either passes the input LV1 through to the trigger processing circuits, or uses the internal FastOR signal to generate its own LV1 signal after a programmable latency. This allows the chip to be used in self-trigger mode with a source, and it will produce output data once it has been armed by a previous LVL1 trigger, and it sees a signal on the internal FastOR.
- There is a dual 8-fold output multiplexor which selects which internal signal or data stream is transmitted off the chip through the serial output. The first eight inputs are synchronized with XCK, while the second eight are not. An identical multiplexor circuit is used both for the standard serial output pin (DO), and for the MONHIT monitoring output pin.
- The LVDS driver/receiver circuits use a second internal current reference to define drive current and receiver bias. The common mode voltage is referenced to the Vdd supply using resistors. Output drive is 0.5mA for low power. Circuits are very stable under process and power supply variations, and easily meet specifications at 2V supply voltage.

Features of new FE-I front-end design:

- TOT performance is almost linear to very large charges. This is not necessarily desirable, as a heavily ionizing particle can kill a pixel for a long time. Below is TOT scan from 0 to 1Me signal, giving TOT of 40μs.

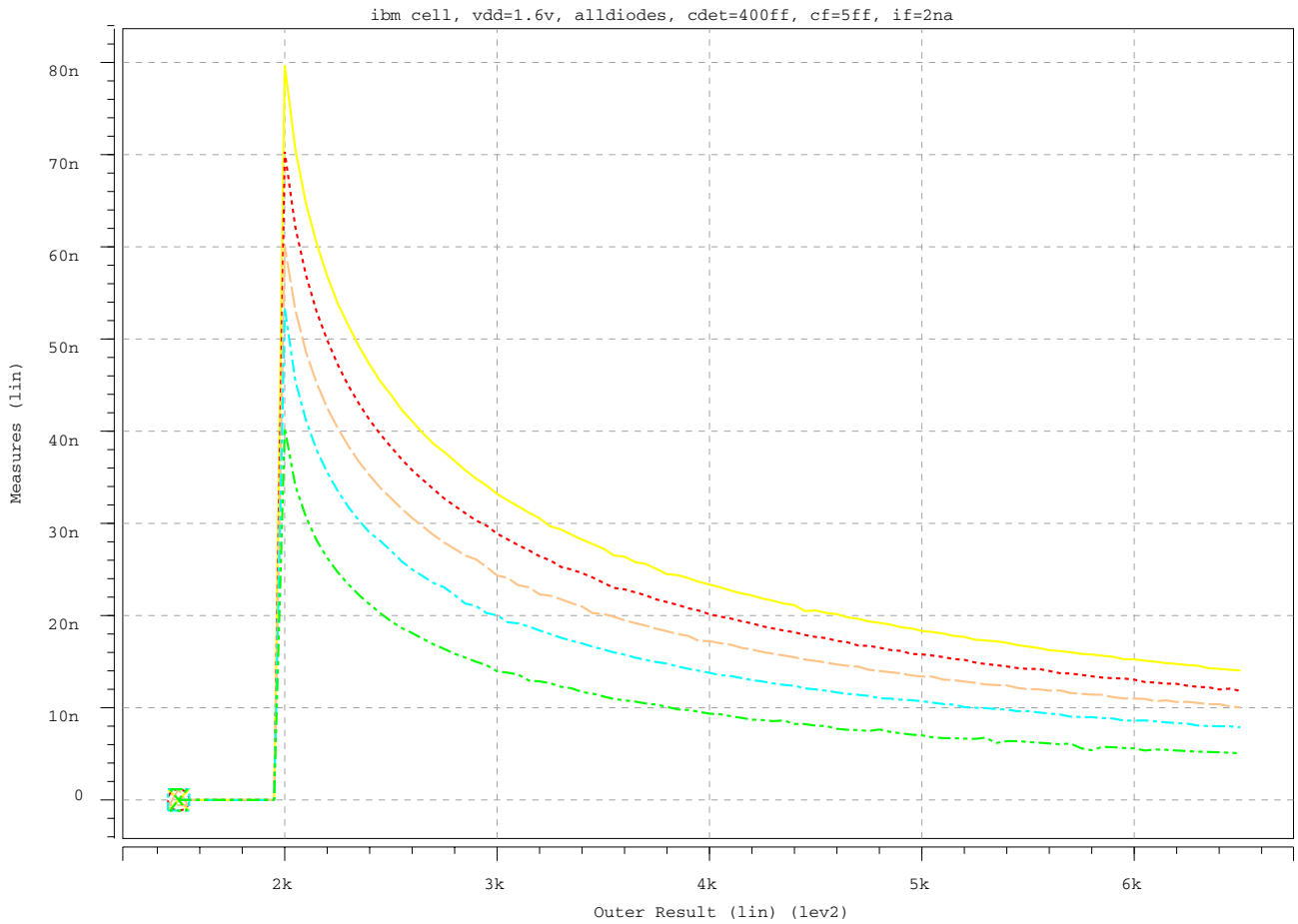


Design	Type	Wave	Symbol
D0: CELL16V_NEWFB_TOT	Transient	D0:A0:v(out)	



Design	Type	File	Wave	Symbol
D0: CELL16V_NEWFB_TOT	Transient	cell16v_newfb_tot.mt0	D0:A0:tot	
D0: CELL16V_NEWFB_TOT	Transient	cell16v_newfb_tot.mt1	D0:A1:tot	
D0: CELL16V_NEWFB_TOT	Transient	cell16v_newfb_tot.mt2	D0:A2:tot	
D0: CELL16V_NEWFB_TOT	Transient	cell16v_newfb_tot.mt3	D0:A3:tot	

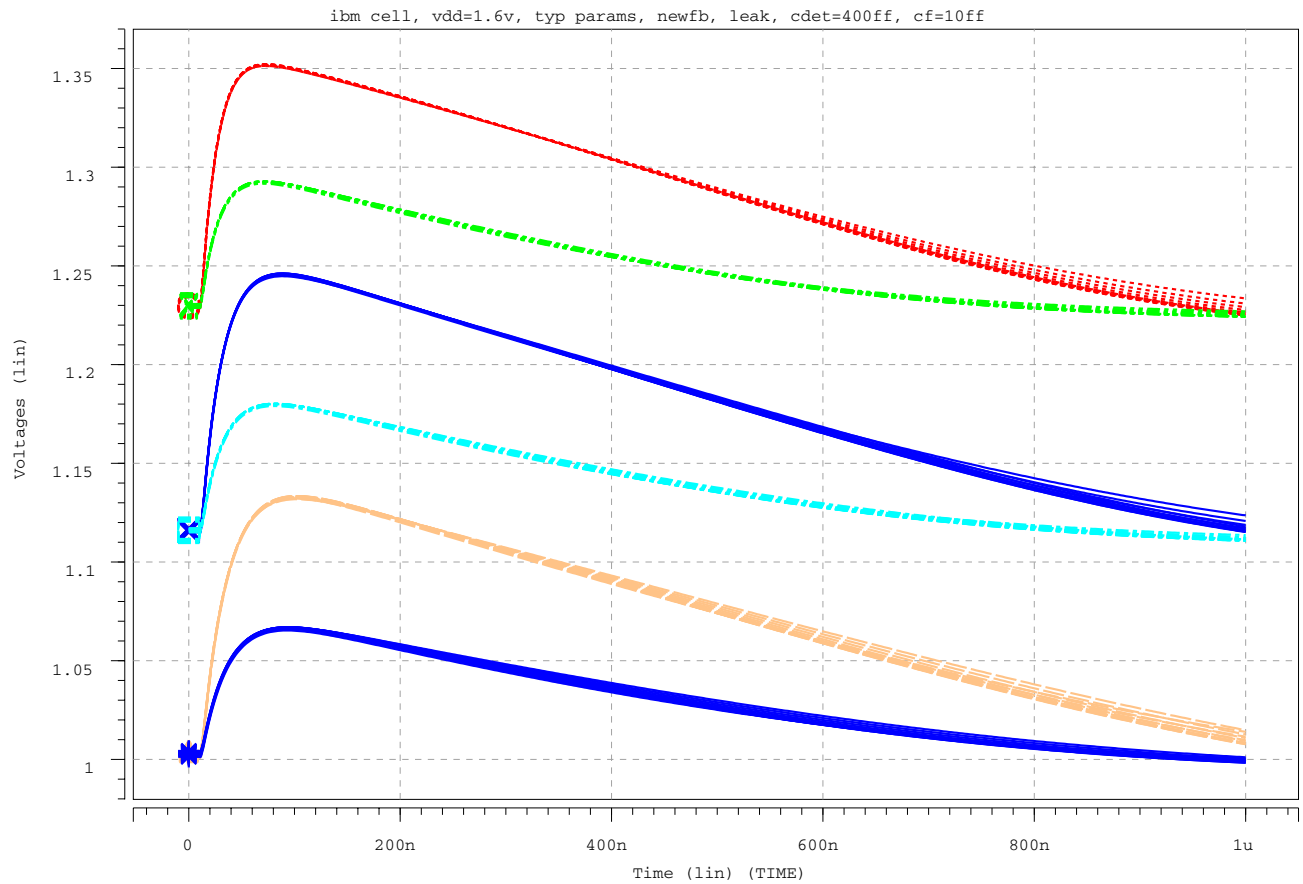
- Timewalk performance (relative to 100Ke) with Cfb=5ff, for CDet from 0fF to 400fF:



Design	Type	File	Wave
D0: CELL16V_TWALK_NEWFET_ALLDIODE_CDET	Transient	cell16v_twalk_newfet_alldiode_cdet.mt8	D0:A8:4:timewalk
D0: CELL16V_TWALK_NEWFET_ALLDIODE_CDET	Transient	cell16v_twalk_newfet_alldiode_cdet.mt9	D0:A9:4:timewalk
D0: CELL16V_TWALK_NEWFET_ALLDIODE_CDET	Transient	cell16v_twalk_newfet_alldiode_cdet.mta	D0:A10:4:timewalk
D0: CELL16V_TWALK_NEWFET_ALLDIODE_CDET	Transient	cell16v_twalk_newfet_alldiode_cdet.mtb	D0:A11:4:timewalk
D0: CELL16V_TWALK_NEWFET_ALLDIODE_CDET	Transient	cell16v_twalk_newfet_alldiode_cdet.mtc	D0:A12:4:timewalk

- Overdrive for Cfb=5fF and CDet=200fF predicted to be only 1500e for 20ns timewalk. For CDet=400fF, this deteriorates to 2500e.

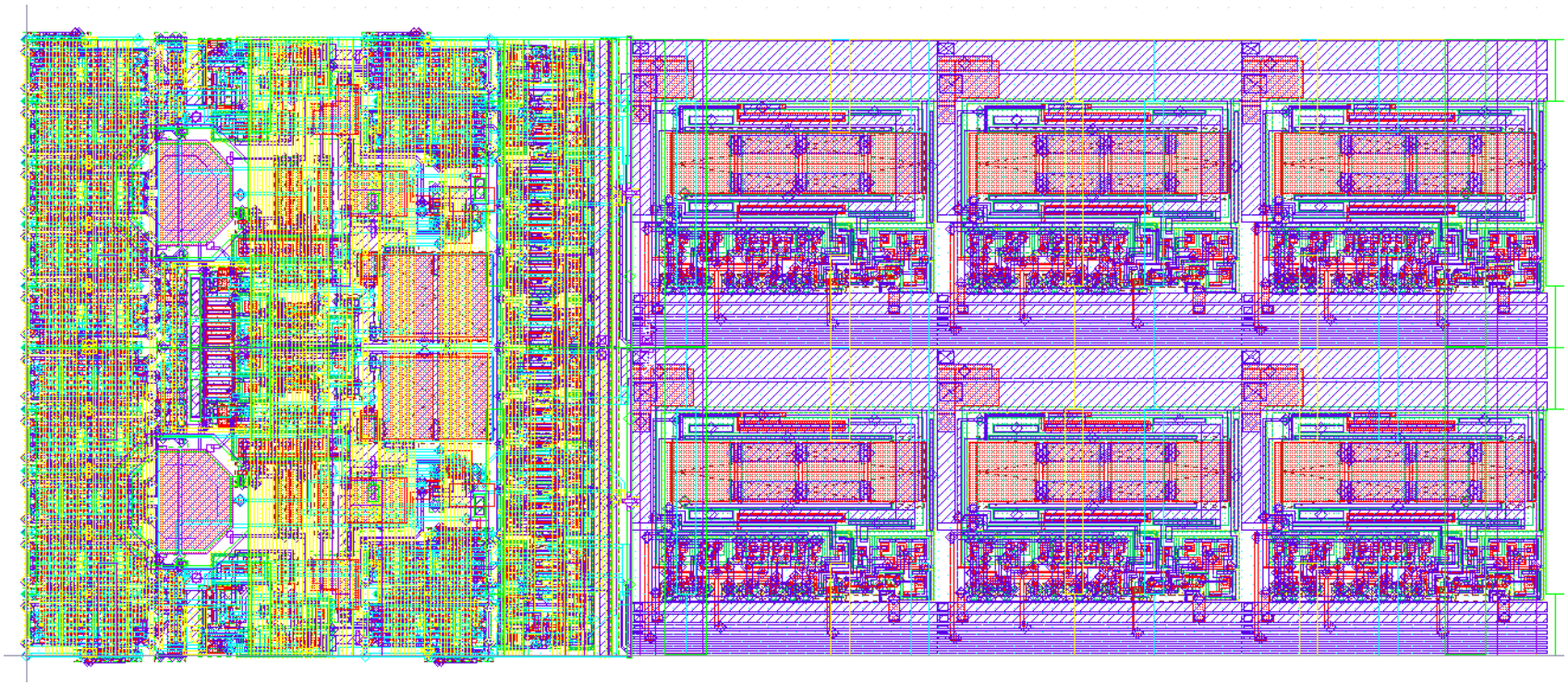
- Variation in shaping as a function of leakage current (from 0 to 50nA):



Design	Type	File	Wave	Symbol
D0: CELL16V_NEWFB_LEAK	Transient	cell16v_newfb_leak.tr0	D0:A0:v(out1)	
D0: CELL16V_NEWFB_LEAK	Transient	cell16v_newfb_leak.tr1	D0:A1:v(out1)	
D0: CELL16V_NEWFB_LEAK	Transient	cell16v_newfb_leak.tr2	D0:A2:v(out1)	
D0: CELL16V_NEWFB_LEAK	Transient	cell16v_newfb_leak.tr3	D0:A3:v(out1)	
D0: CELL16V_NEWFB_LEAK	Transient	cell16v_newfb_leak.tr4	D0:A4:v(out1)	
D0: CELL16V_NEWFB_LEAK	Transient	cell16v_newfb_leak.tr5	D0:A5:v(out1)	

- Previous FE-B shaping was much stronger in presence of leakage, destroying the charge measurement for irradiated sensors. This is no longer the case.

Layout of pixel, showing two FE blocks and two cap groups:



- A total of 3 smart capacitors are placed in each pixel, for a total of 8640 in the chip, giving roughly 15nF of total decoupling. These capacitors are claimed by IBM to have excellent properties up into the GHz region.
- Capacitor size is roughly 40x50μ, allowing the placement of 3 capacitors in the remaining empty space in the pixel.

Control Features:

Command Register:

- FE-I contains 20-bit Command Register made up of SEU-tolerant latches.
- Mostly just strobes for writing all pixel FF. New feature is readback for 14 FF in pixel control block. This was thought necessary since there are now more than 40K configuration bits in the pixels of one FE chip.

Global Register:

- FE-I contains a 202 bit long Global Register made up of SEU-tolerant latches.
- Too many highlights to mention:
- Enable bit for TSI/TSC to reduce digital power when not acquiring data
- EOC MUX control to choose whether LE, TE, or TOT is presented in TOT field of hit output data
- Total of 14 DACs, including new 9-bit VCal DAC and 9-bit MonLeak DAC
- Dual injection capacitors and improved internal chopper for each pixel, plus independent external injection path for good cross-calibration
- Special additional blocks like CapMeasure to measure values of Clo, Clo+Chi, and Cfb directly, and simple DAC-based comparator ADC to measure feedback current and leakage current in each pixel.

Pixel Control bits:

- FE-I contains 14 bits of configuration in each pixel, made up of SEU-tolerant latches.
- Highlights include:
- Kill bit to turn off pixel preamplifier to avoid injection of analog noise
- Separate enables for HitBus and Digital Readout
- Select for calibration which is no longer stored in the shift register, but is a separate FF.
- A 5-bit TDAC for threshold trimming and a 5-bit FDAC for feedback current trimming.

Testability features:

- Digital injection of Str signal into each pixel, allows decoupling of analog and digital parts of the chip, and creation of well-defined data pattens.
- MUX to allow capturing LE and TE RAM data into TOT field of output data for testing purposes.

Digital Readout Features

- All differential timestamp distribution, and differential SRAM and ROM blocks in the pixels, combined with new differential senseamp design with swings of VDD/2.
- Use of 8-bit timestamps to provide 6.4μs maximum L1 latency.
- TOT processor block at bottom of each column. This block calculates the TOT for each hit. It can apply a simple threshold to the TOT to suppress writing small charges into the EOC buffers. It can also apply a one crossing timewalk correction to all hits below a settable TOT threshold. Hits below the threshold are written twice, once with the raw leading edge time, and once with a leading edge time of one crossing earlier. Both features can operate simultaneously if desired.
- Implementation of 64 EOC buffers, exceeding known B-layer buffering requirements even at design luminosity.
- Global control of clocking in column pair (CEU clock control set to 0 disables all operations in column pair), by column enable bit (suppresses clocking of EOC buffers for disabled column pairs), and by TSI/TSC disable bit. By default all of these bits are off when chip is powered on, leading to low initial digital power consumption.
- Replacement of 4-bit trigger number with 4-bit BCID in hit word. BCID increments every crossing, and provides more robust protection against missed/excess L1 triggers. Internal event building always based on trigger number, so no ambiguities are introduced.

New features and pins outside the bonding region:

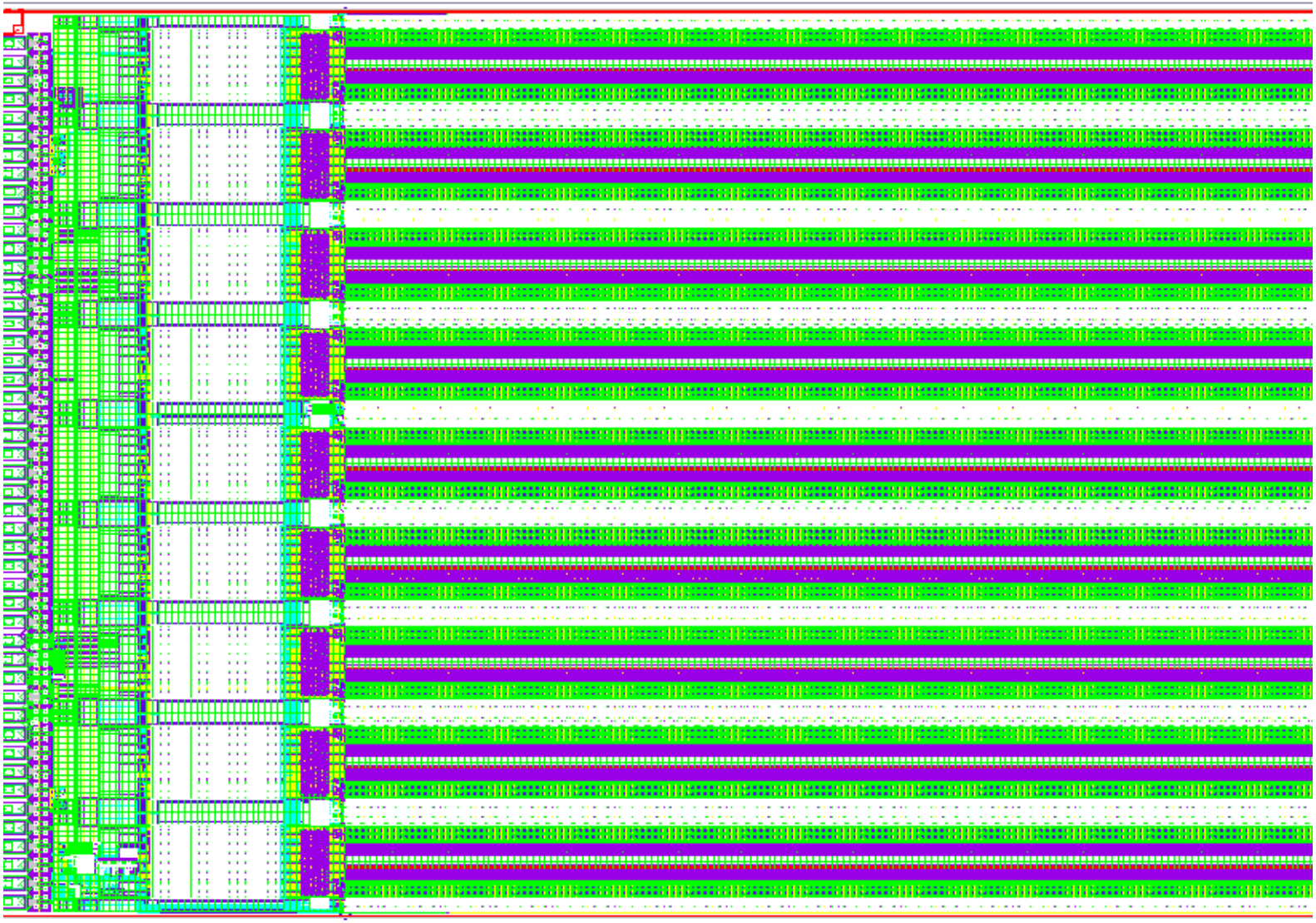
- Power Management features: Two overvoltage clamp circuits are included, one for each power supply. They use a diode and a resistor to set a soft threshold of about 2.7V, after which a large PFET is used to sink excess voltage to ground. Note that the power pads also include the recommended IBM transient clamps, which are designed to protect the chip from sharp spike transients on the power rails when no power is applied to the chip. In addition to the overvoltage clamps, there are two simple regulators. One is a shunt regulator, based on the same circuit as the clamp, but with a threshold of 2.0V. The second is a simple linear regulator using a band-gap reference, and set for 1.6V operation. The shunt regulator is intended for study of powering schemes based on constant current supplies. The linear regulator would generate the analog supply voltage from the digital supply voltage, allowing operation of the FE chip on a single power supply. There is little risk posed by these circuits if the wire-bonds are not connected, and placing them inside the FE chips allows the performance of modules to be compared with and without these circuits, without changes to the Flex design.
- MonDAC provides multiplexed access to all of the internal DACs for characterization during testing.
- MonLeak provides access to a current summing tree (controlled in the same way as the HitBus) that allows a direct measurement of the preamp feedback current and the sensor leakage current: I(OutLeak) = 2*If + ILeak. This has already proven very

useful in chip characterization. A simple internal ADC, based on a 9-bit DAC, is also provided.

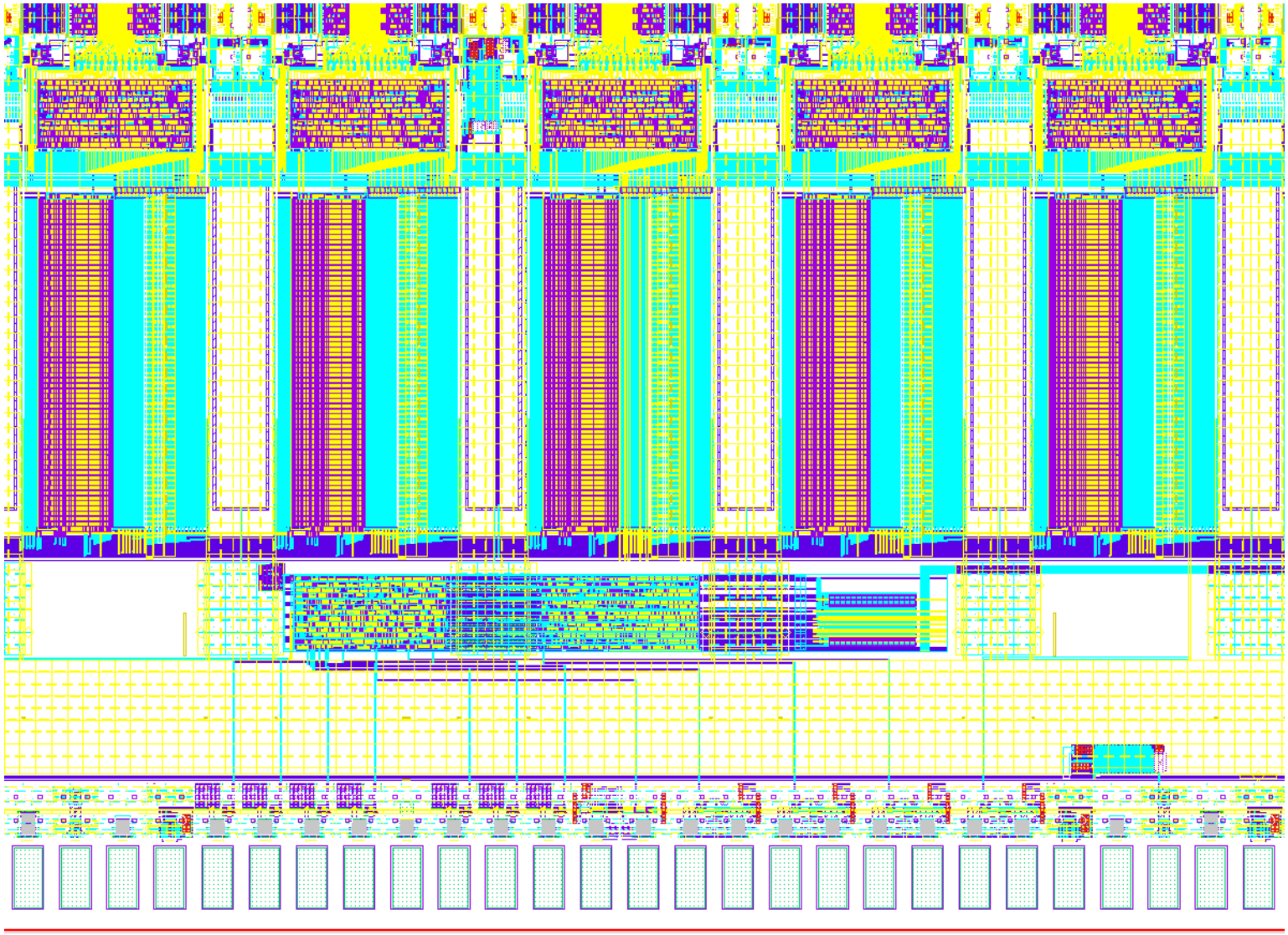
- MonRef allows direct monitoring of the current reference used for the LVDS I/O pads, without requiring any other circuitry to operate on the FE chip.
- MonVCal allows direct monitoring of the VCal voltage generated internally on the FE chip for charge injection calibrations. VCal is generated by a 9-bit current DAC and a resistor. The resistor is matched to the one used in the current reference, providing first-order cancellation of process variations.
- MonAmp includes an analog MUX which allows us to see the preamp waveform, the two sides of the second amplifier, and the chopper input. This is followed by a 100Ω buffer amplifier, which can drive a daisy-chained bus of test amplifiers, provided only one is enabled at any time.
- CapMeasure pin is attached to new capacitor measurement circuitry, which uses a charge pump circuit to measure accurate values for the critical capacitors used in the front-end (C(feedback), C(inj-low), C(inj-high)) by measuring a single DC current. This circuit has been used in the DMILL CapTest chip, and can provide accurate measurements of capacitor arrays at the sub-fF level. The circuit used here allows selection of 0, 1, 2, or 4 copies of each capacitor to be measured, as well as selection of 4 different clock frequencies for the measurement. This gives good control of systematics in the measurements.

Brief tour of the layout:

- Top level view of the chip (all 5 metals displayed):

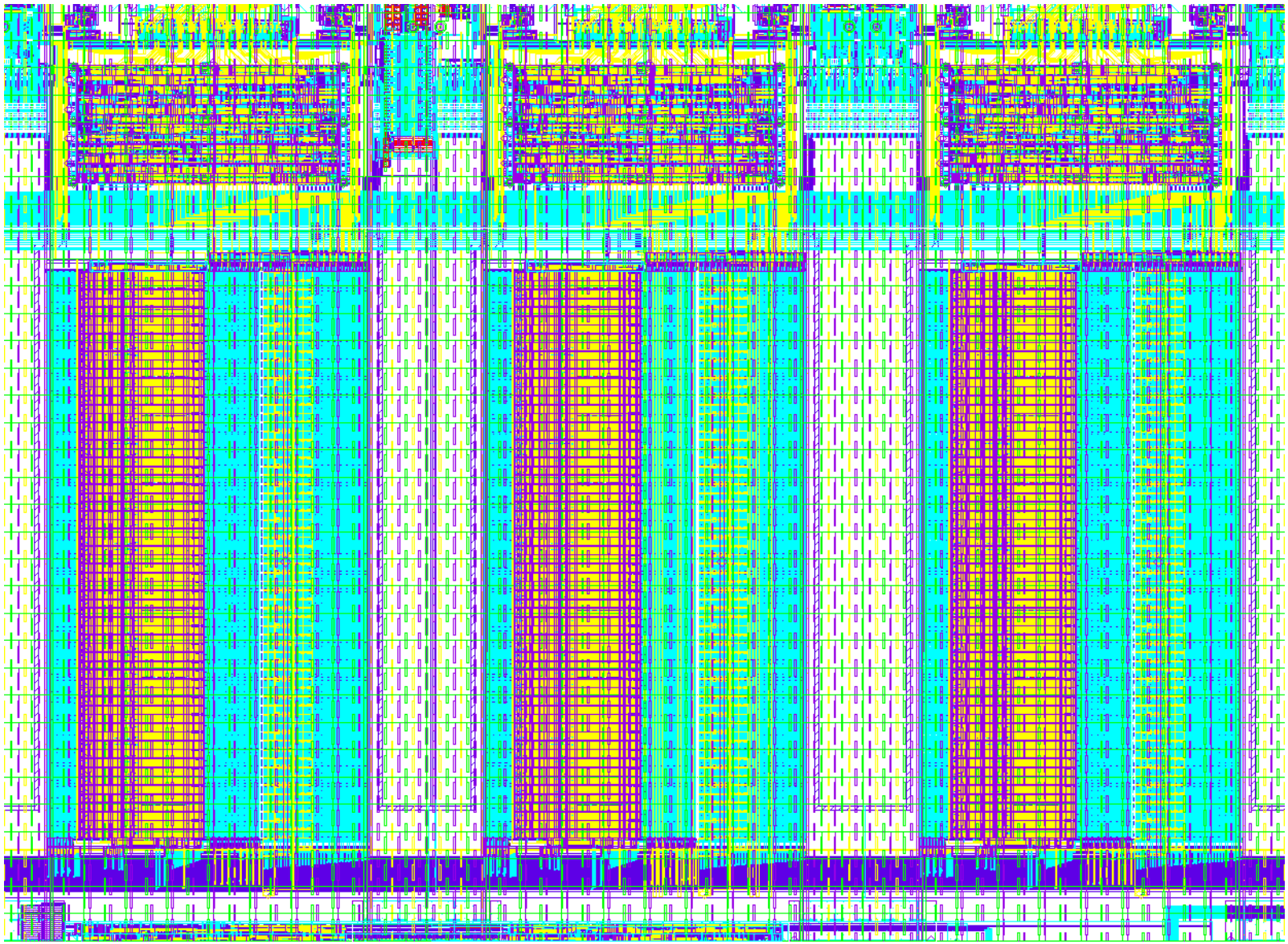


- Zoom showing EOC blocks and bottom of chip, including synthesized command decoder and readout controller blocks:



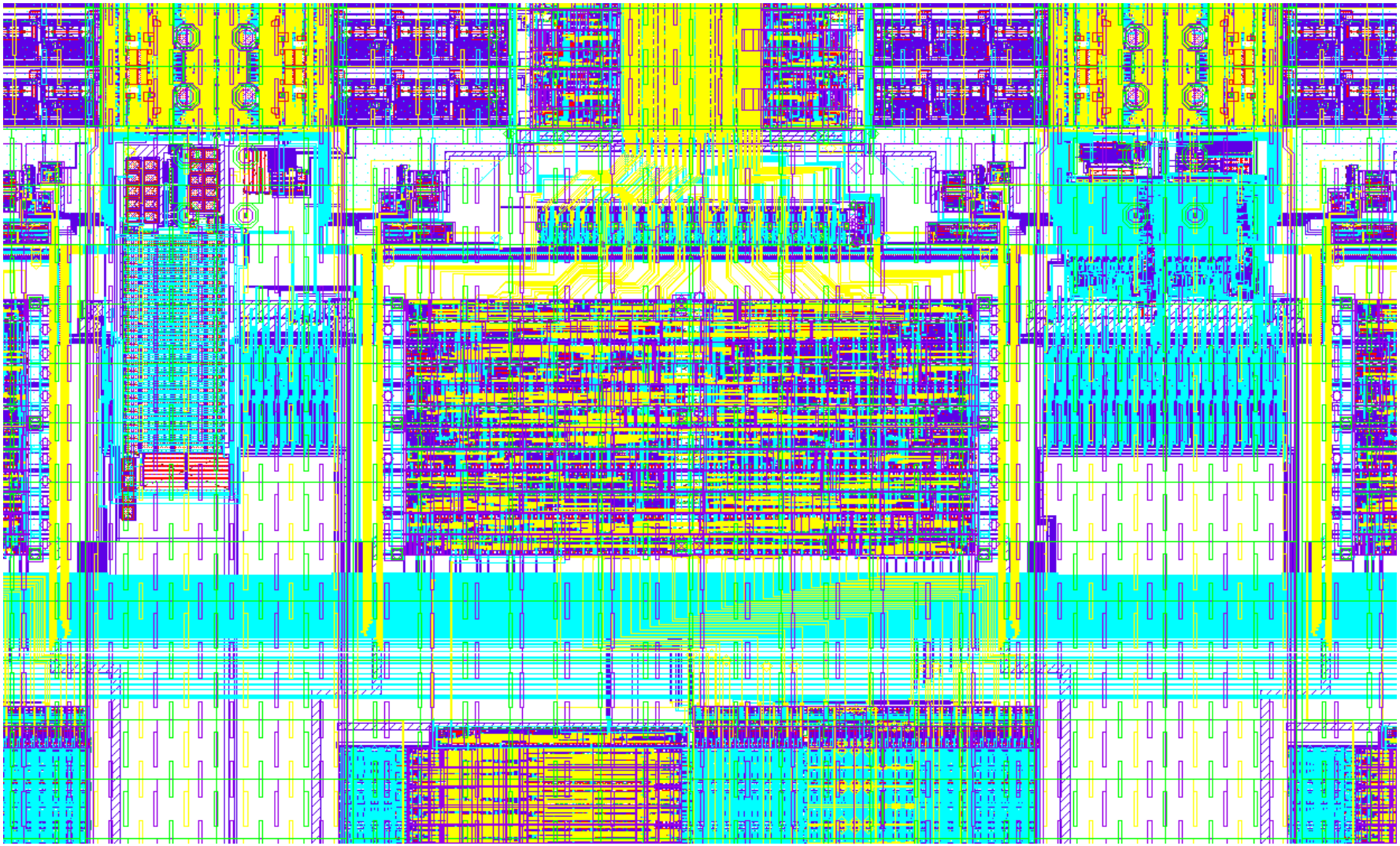
- Note that the bottom of the chip is still largely empty.

- Zoom into EOC buffer blocks, each containing 64 hit buffers for a column pair (requiring a total vertical height of about 1mm):



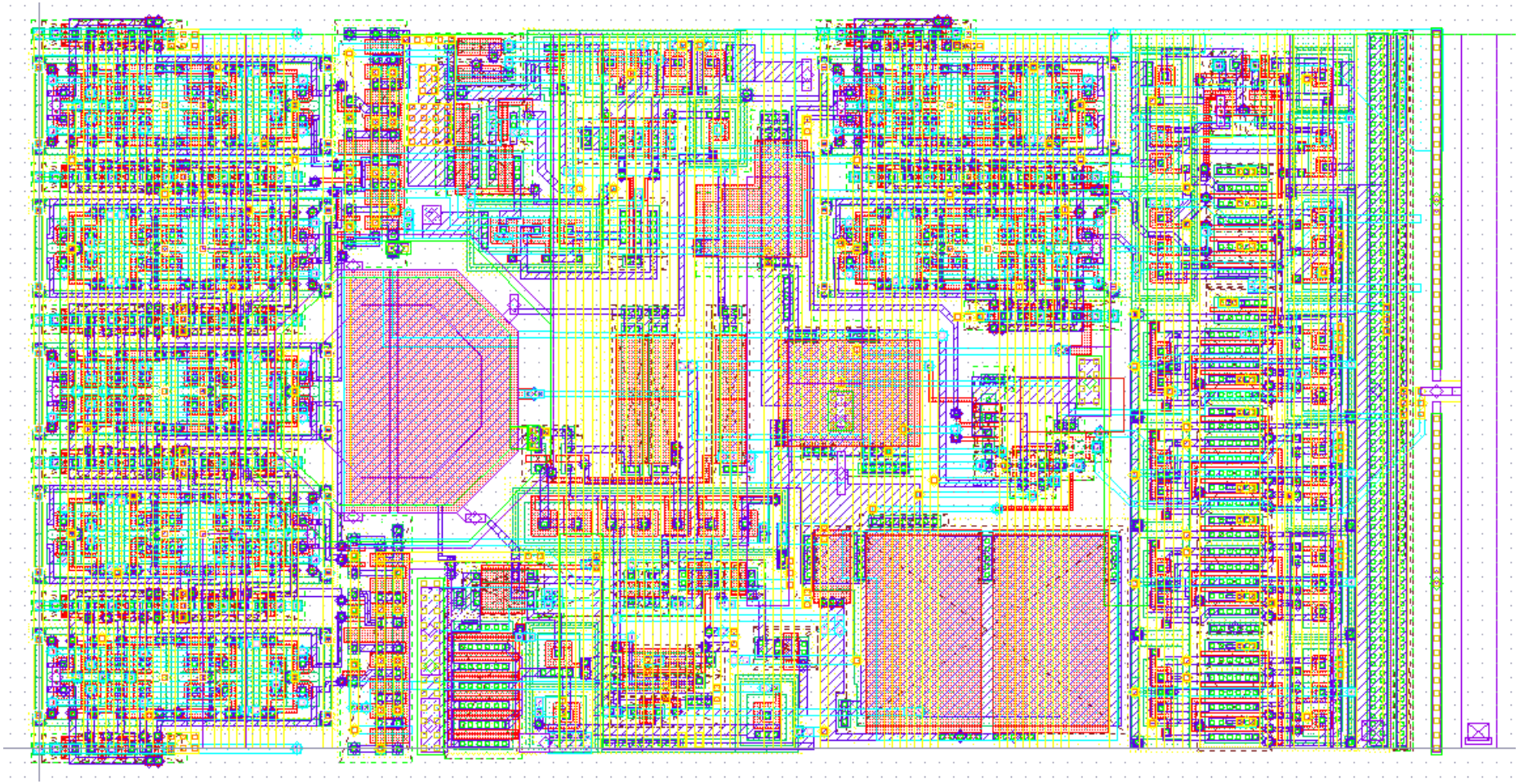
- TOT processor blocks feed into EOC blocks, horizontal bus is at bottom.

- Zoom into bottom of column region, showing integration of DACs and bias cells with analog columns, and CEU+TOT processor with digital columns:



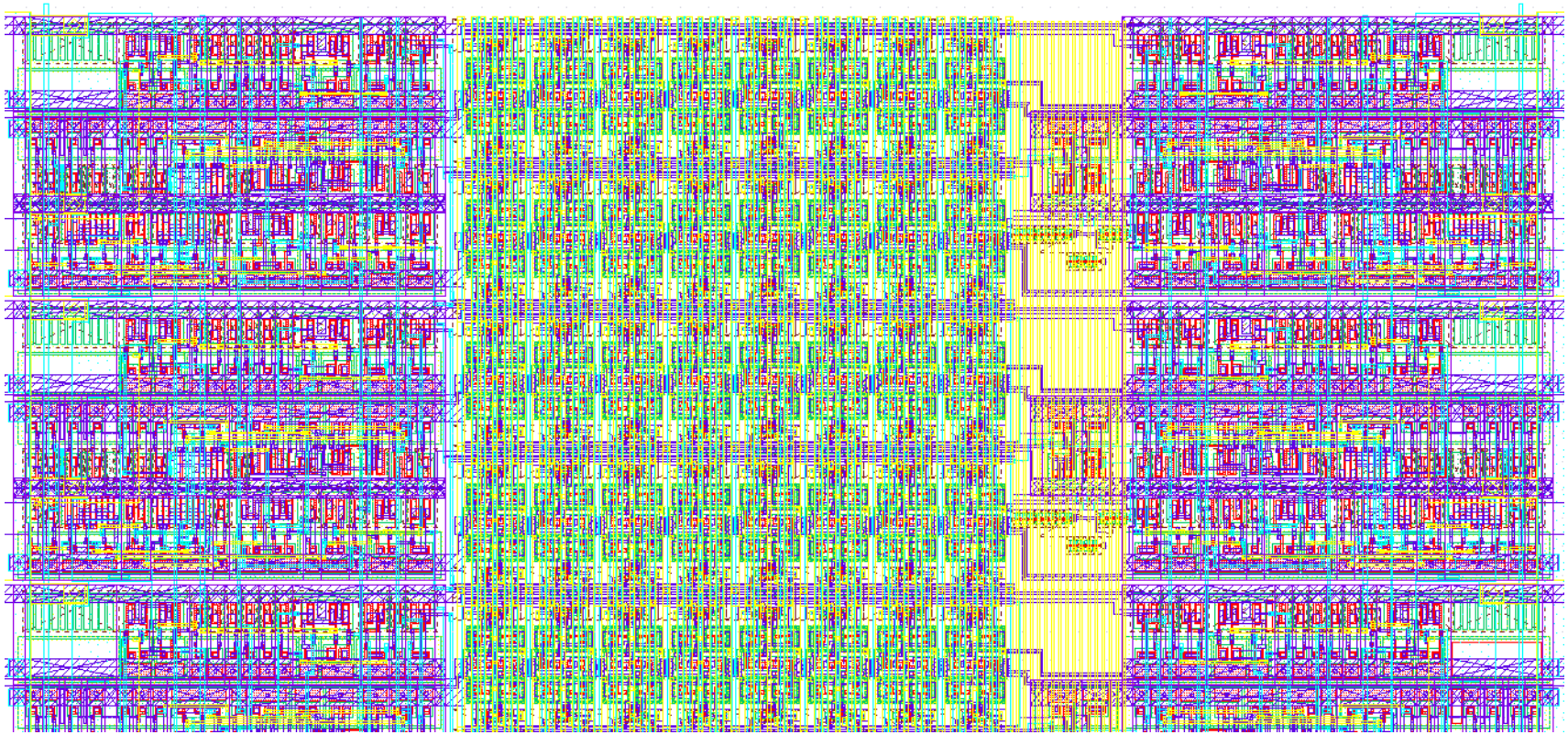
- Left analog column has current reference and register bits, right has pair of 8-bit DACs and register bits (all registers use SEU-tolerant latches).

- Zoom into Pixel FE block:



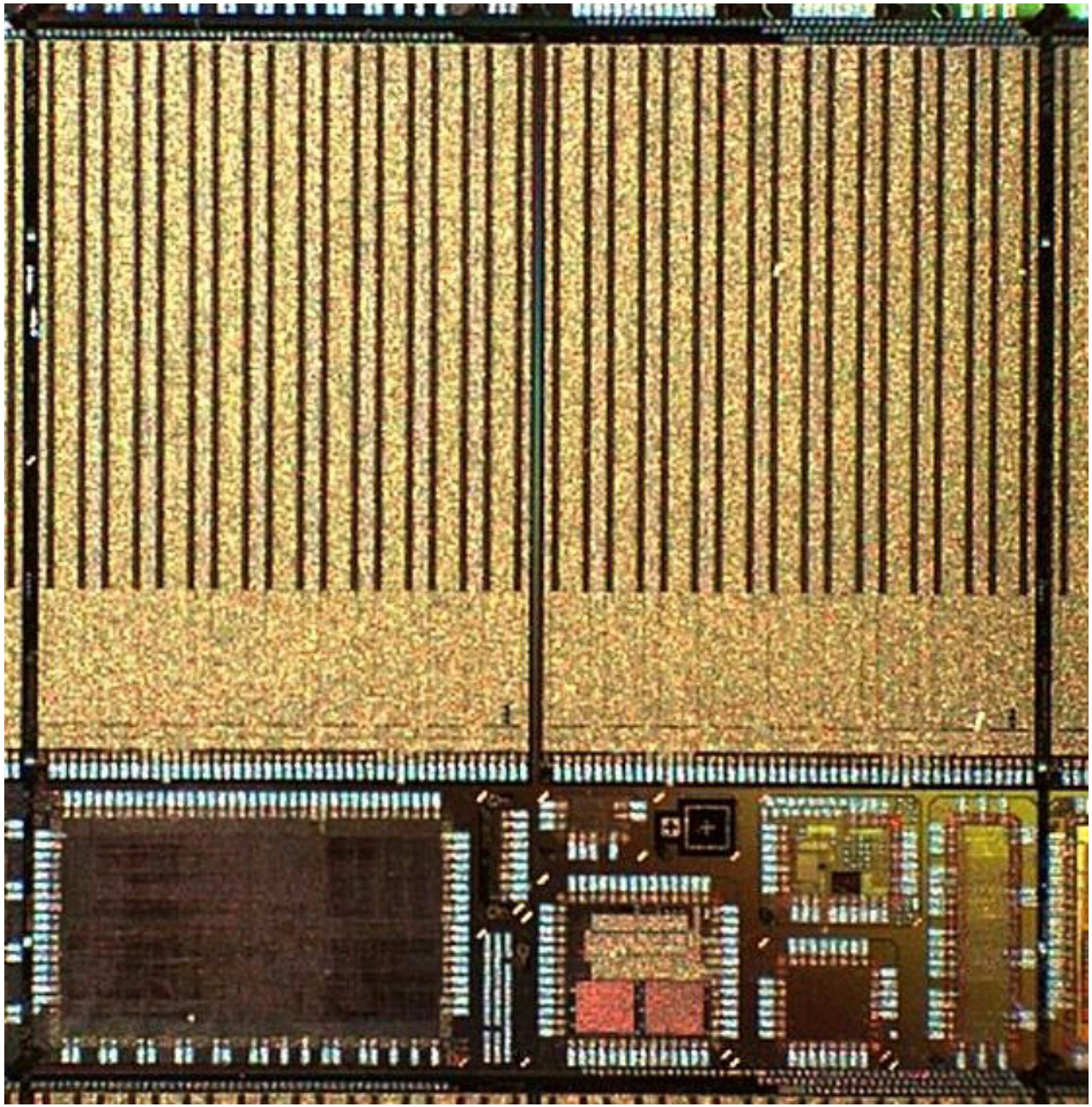
- Left of bump, can see 10 SEU-tolerant latches. Lower right below bump is preamp, center is feedback, top is second stage and discriminator. Right end includes leakage compensation capacitors and additional 4 latches and logic for control of hits, calibrations, and digital injection.

- Zoom into readout region of pixel (two back-to-back columns):



- Central region includes dual 8-bit differential SRAM for LE and TE information for each pixel plus address ROM. Everything is differential (timestamp input, plus RAM and ROM output).
- Left and right sides contain hit logic, sparse scan, and handshaking with CEU for data transfer.

Reticle for FE-I1 engineering run:



Reticle includes two different FE-I chips.

One is FE-I1A, with Cfb=10fF, the second is FE-I1B, with Cfb=5fF.

Reticle also includes MCC-I1, Analog Test Chip, LVDS Buffer chip, DORIC and VDC opto-chips, and several other small test chips.

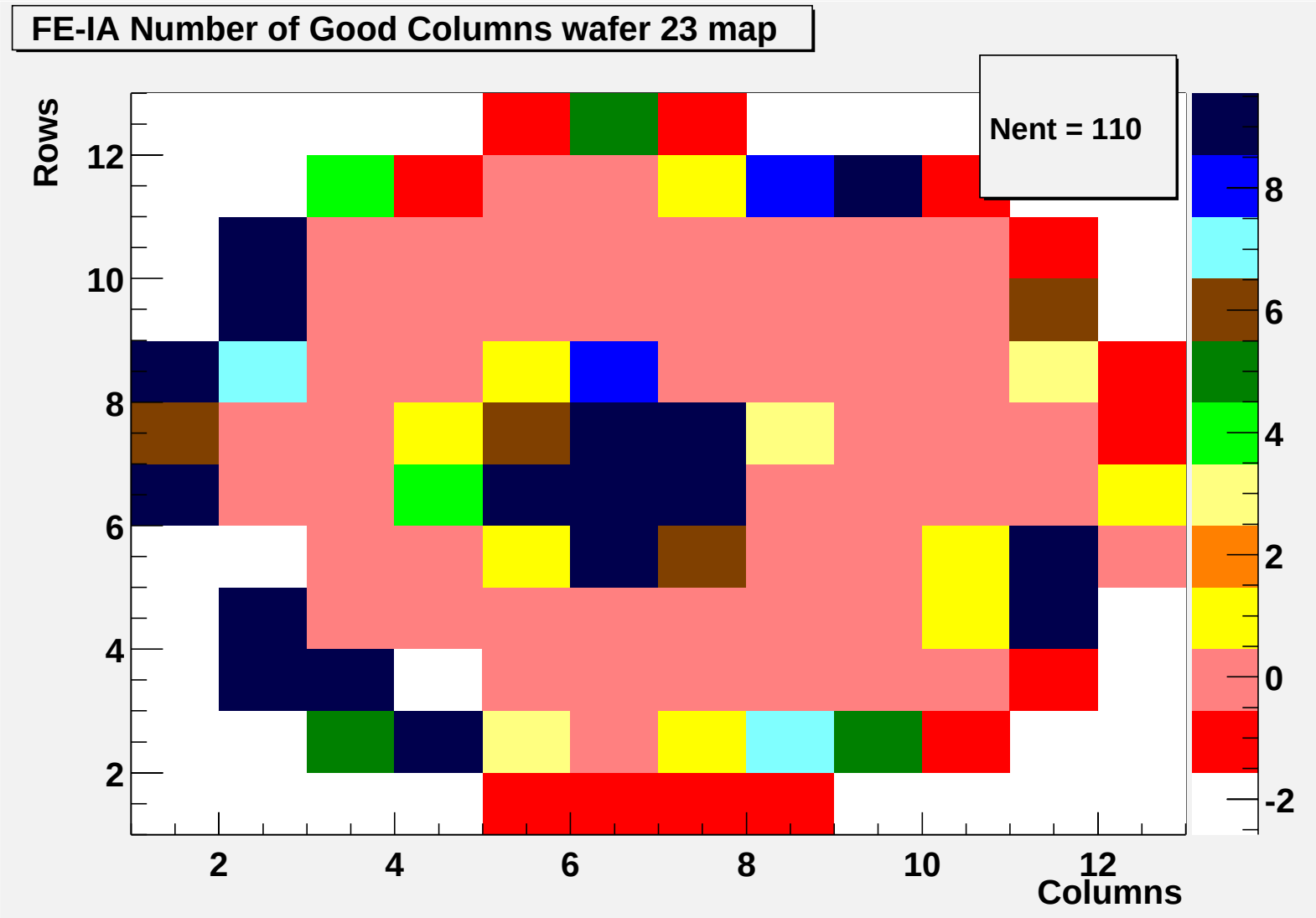
There are 112 potentially good reticles on a wafer.

FE-I1 Performance and Yield Issues

First wafers from 12-wafer Engineering Run arrived in Jan 02:

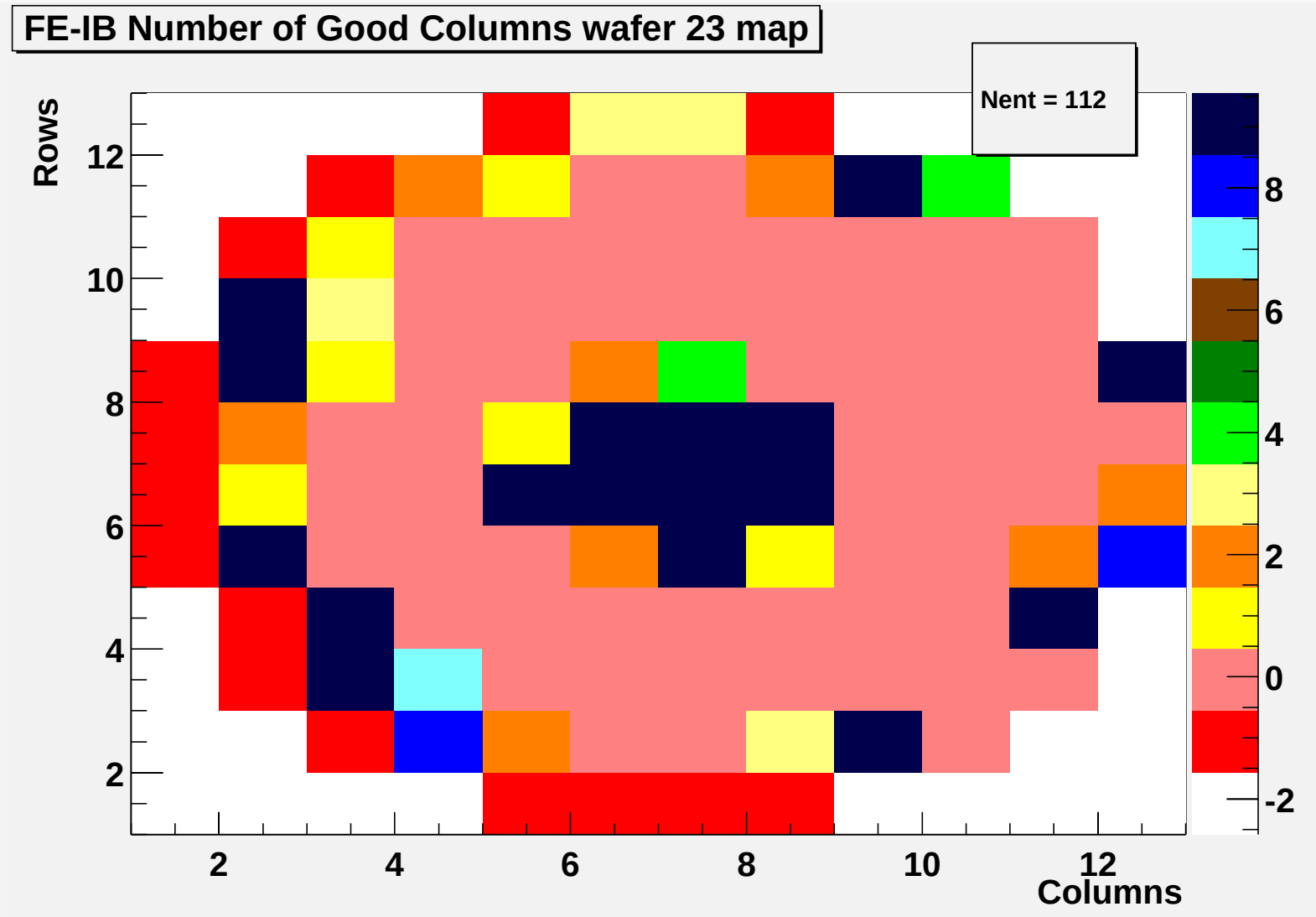
- All blocks worked roughly as expected. Remarkable success for 2.5M transistor chip submitted in new process !
- All performance features, even for new analog front-end, have performed close to expectations. Even threshold dispersion and timewalk, studied by large HSPICE simulations, agree reasonably well with the simulations.
- However, it was quickly realized that there was a serious yield problem. Yield to pass simple selection criteria (analog/digital currents OK, all registers working, and basic digital inject test working) was only about 15%. Even more striking, the good chips were all confined to a small area in the core of the wafer, or along the extreme edges.Finally, chips that passed basic register tests (few percent of transistors tested), would usually be perfect for full digital and analog tests, so defect density was not an issue.
- Extensive investigations of failure modes have been made, and fault analysis was performed by the foundry. Four additional wafers, initially held at back-end processing, were sent for evaluation. They showed very low yield (3%), and very basic failure modes consistent with metallization problems (mostly supply shorts).
- Example of wafer map of first wafer probed shows typical pattern seen in first 12 wafers. We find very little variation within a lot or wafer group, but very large variations in yield between groups !

Wafer Map for SESB23T (good column pairs for A chips):



- Chips with no data appear White, bad Global Registers are Red, and othe colors represent the Pixel Register test results. There are 18 (3) chips with working Global Registers and 9 (8) column pairs working in the Pixel Register.

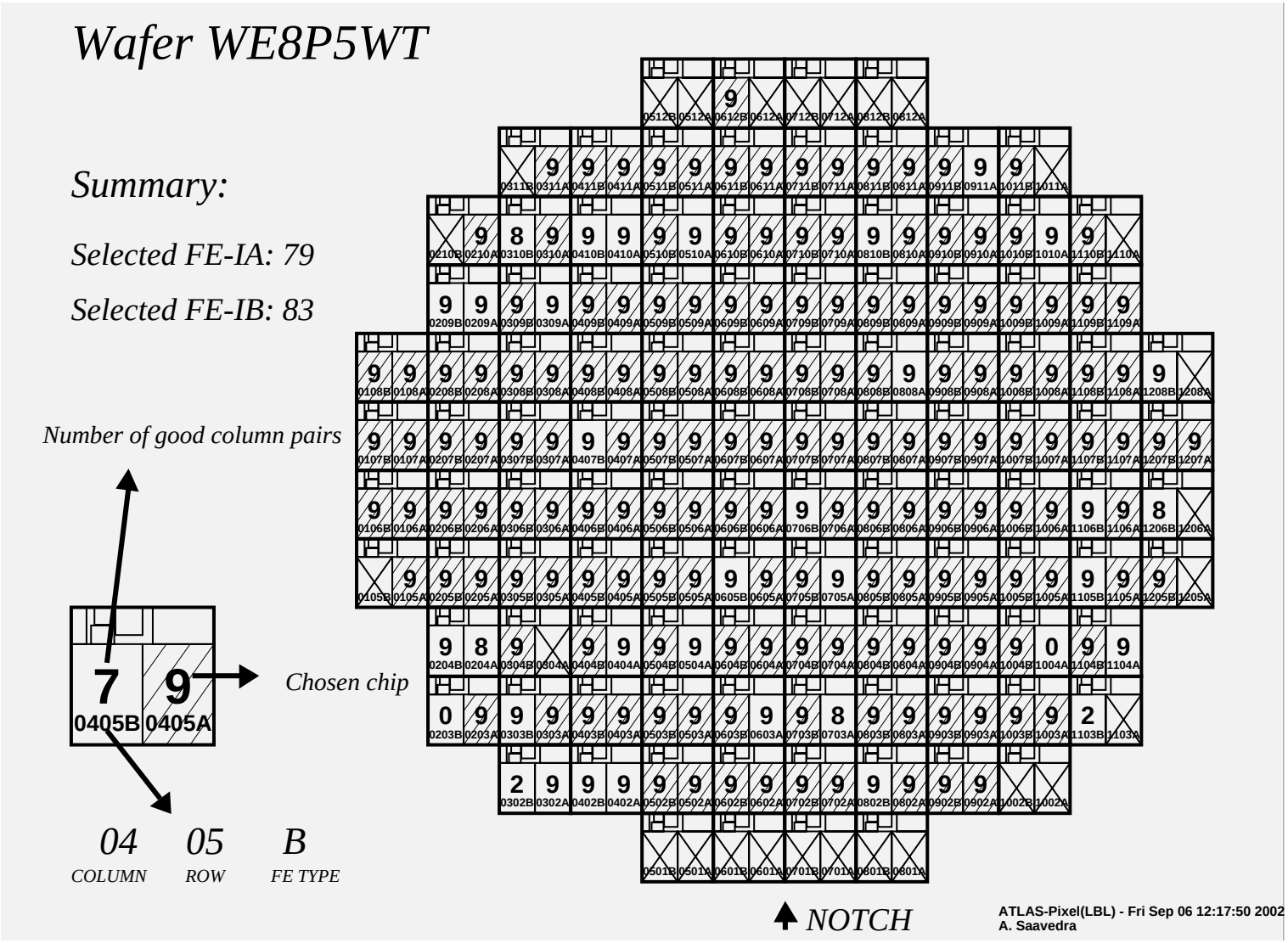
Wafer Map for SESB23T (good column pairs for B chips):



- Chips with no data appear White, bad Global Registers are Red, and othe colors represent the Pixel Register test results. There are 17 (2) chips with working Global Registers and 9 (8) column pairs working in the Pixel Register.

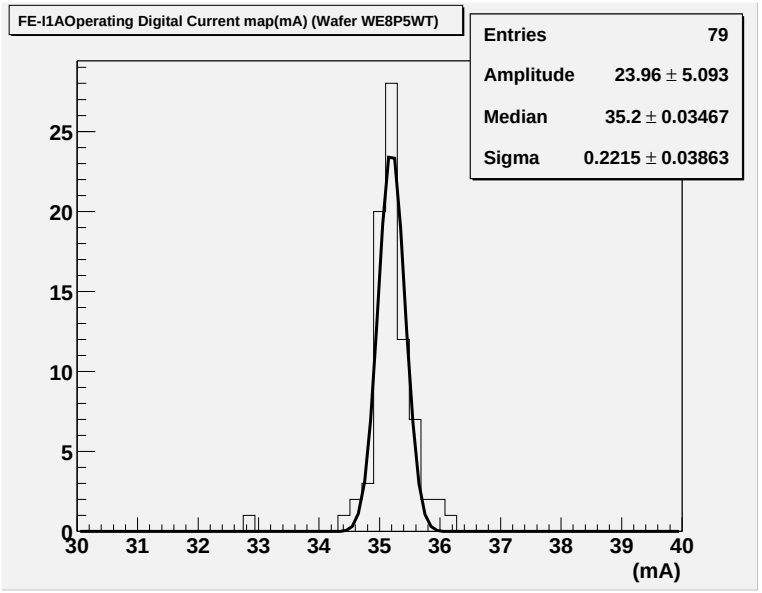
- Concurrent and similar problems seen with CMS APV25 chip, and some fault analysis clues were found. Two new lots were run for CMS and two new lots were run for ATLAS, with delivery in May/June.
- The vendor has not uncovered anything very useful, despite significant investment of resources. At this stage, I believe basic problems are in metallization stage. No clear evidence for real problems with our design. However, vendor claims no other customers, besides CERN, see such wild yield variations. Annular NMOS (unique CERN feature) extensively investigated by foundry, but no fabrication issues uncovered. Recent change of metallization fill rules by IBM may be relevant, and we will see what happens with our next run of two lots late this year.
- First of the two replacement ATLAS lots had “expected” yield behavior. For eight wafers recently sent for bump-bonding, “simple” cuts (supply currents and register tests) give an average yield was 79%. For more complete cuts, including perfect digital operation of every pixel and EOC buffer, the average yield was 64%.
- Second of the replacement lots had intermediate behavior. For a single wafer probed so far, the yield for complete cuts was 25%, with a large “donut of death” in the wafer where all chips failed even basic register tests, and usually showed supply shorts on one or both supplies.
- The basis for our production planning is an assumed 50% yield. This seems achievable, but there will apparently be large fluctuations between lots. Fortunately, within one lot, the yield behavior seems fairly consistent from wafer to wafer. In addition, the wafer cost is relatively low in large quantities (about 2K$).

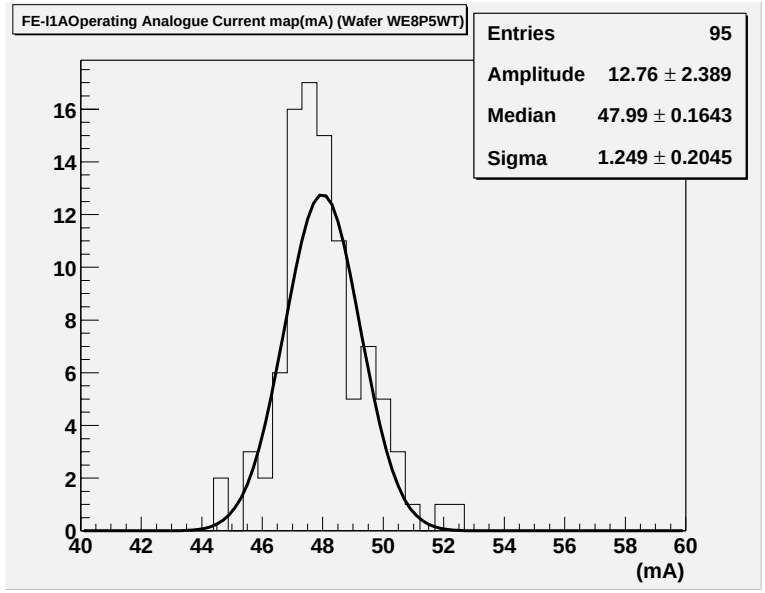
Typical wafer from good replacement lot:

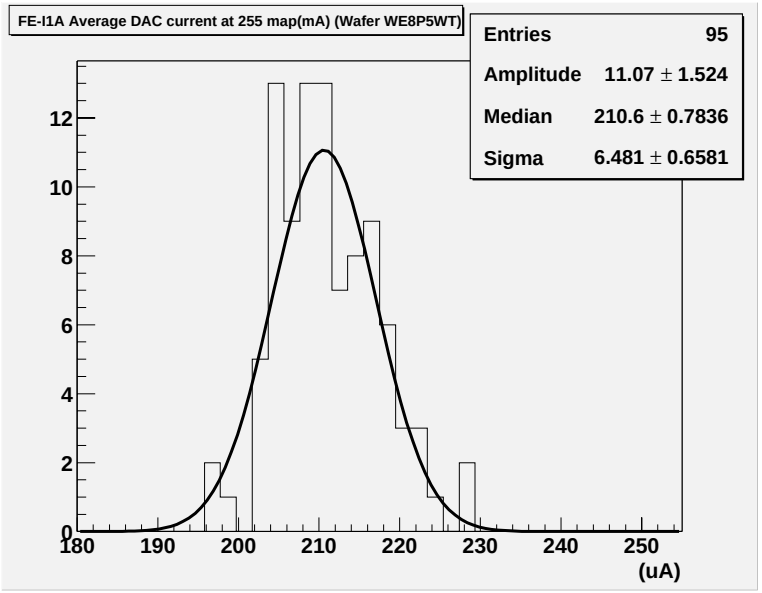


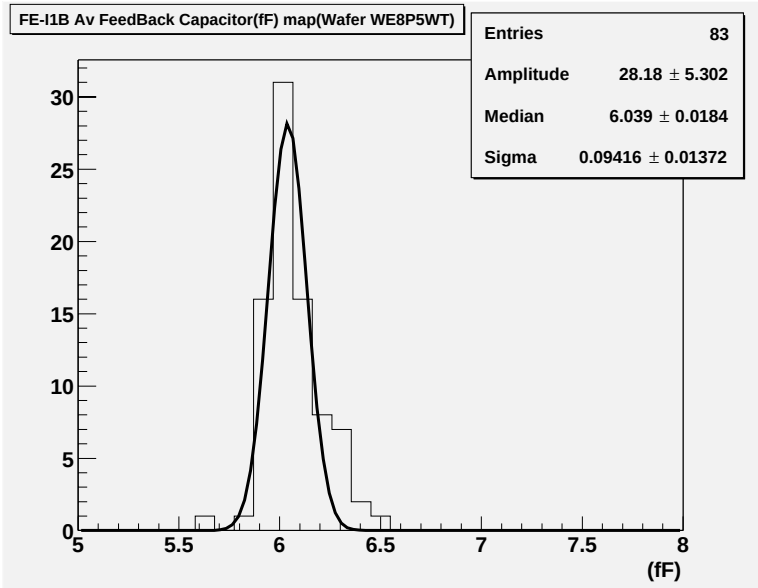
- Map shows chips passing complete cuts (supply surrents, registers, DACs, perfect digital pixels and EOC buffers). This has a yield of 73% for this wafer.

Examples of plots from wafer probing:









Improvements for next generation (FE-I2)

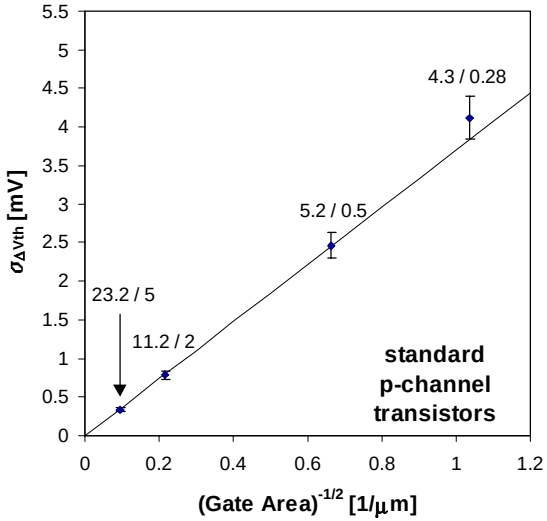
- Upgrade program is underway, and new submission expected by the end of this year. This should be a pre-production quality chip. This means that if all goes well, modules built with this chip can (and will ?) be used in ATLAS.
- A number of minor problems were uncovered, and are easily resolved.

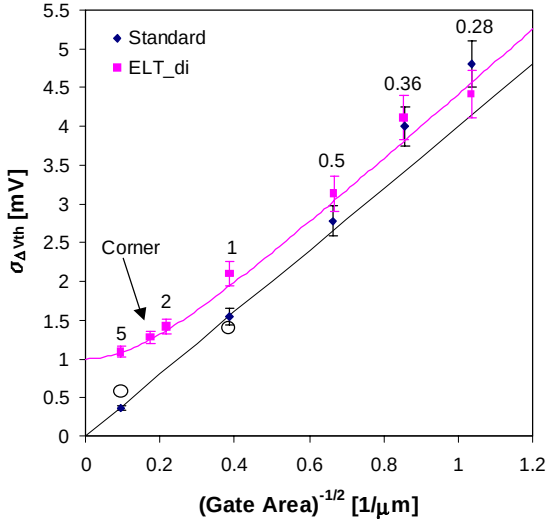
Serious issues for FE-I2:

- **Threshold dispersion and "re-dispersion":** Although the initial dispersion was roughly as expected, managing this dispersion through the lifetime of the chip has proven challenging. Can typically tune a given assembly to a sigma of better than 100e for a given set of conditions. However, changing the temperature from +20C to -10C re-disperses the thresholds to about 300e sigma. Similarly, a total dose of about 300KRad re-disperses the thresholds to about 300e sigma. Measurements at LBL Cyclotron show that this re-dispersion does not saturate at high total dose. Finally, small changes in the global threshold adjustment also cause re-dispersion.
- Threshold dispersion and re-dispersion in our design arise from matching errors between identical transistors. A careful Monte Carlo analysis has been performed, using the CERN matching data for IBM (Anelli thesis). This approach predicted the observed threshold dispersion, and has now been used to optimize the sizing of all transistors in the front-end. In some cases, there are real trade-offs between speed (timewalk) and threshold dispersion. Modifications to the present design should reduce the dispersion by a further 20-30% with little loss in performance.

- Each device had its VT modified using sigma taken from the thesis of G. Anelli:

Standard PMOS VT matching versus device size



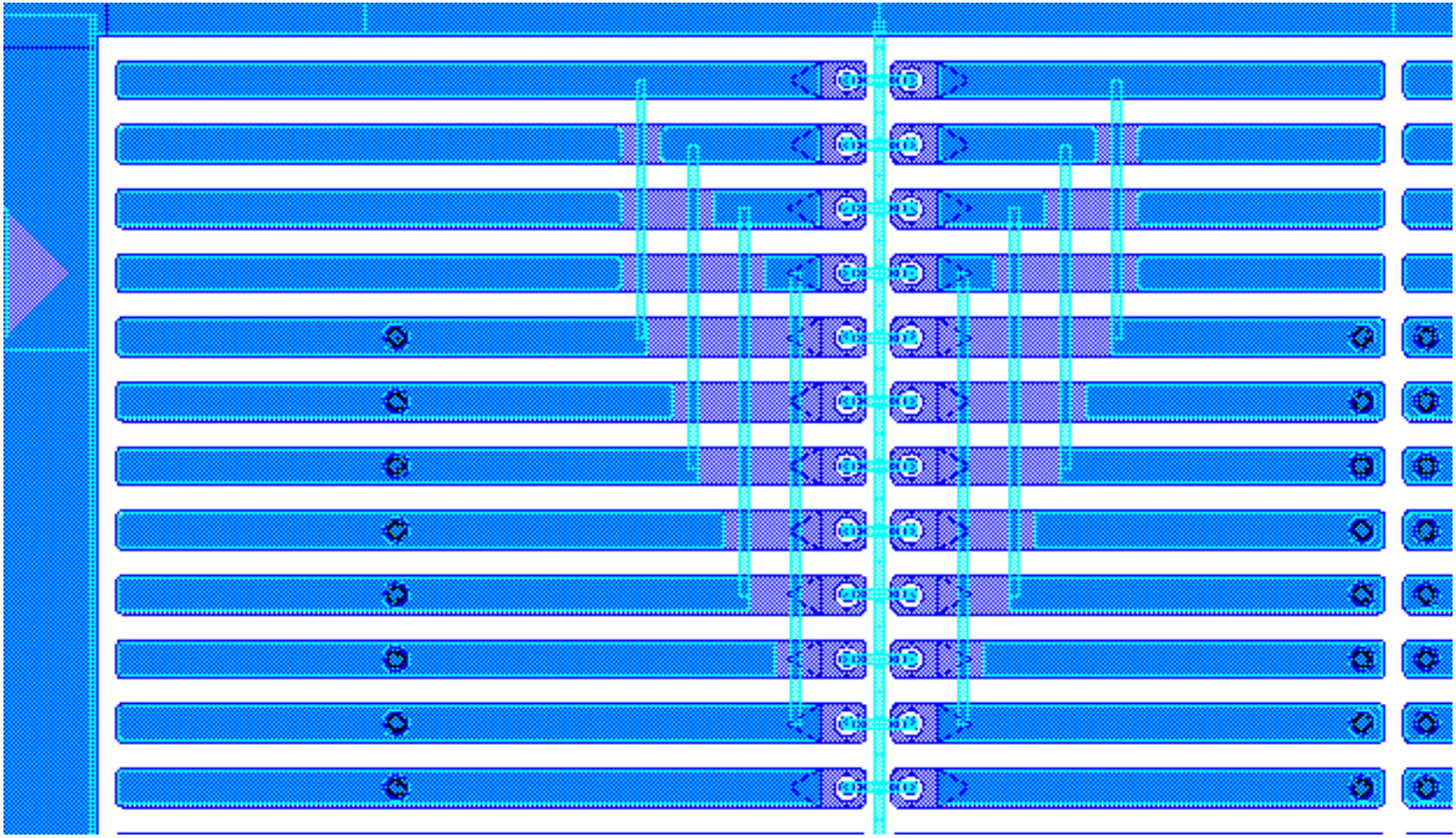


Enclosed NMOS VT matching versus device size

- Very little is known about additional differential threshold shifts in matched transistors that can occur after irradiation, etc. CERN data indicates some large irradiation effects that may be consistent with our results.
- **<u>Bias distribution:</u>** Significant top-bottom variations are seen in the timewalk performance of FE-I1. These arise from internal voltage drops on AGND, which in turn modify the Vgs used to distribute large bias currents like IP (the preamplifier input transistor bias, typically 8μA). All mirror transistors in our design are in weak inversion (sub-threshold), so the mirrored current is very sensitive to small changes in Vgs. In addition, the reference plane for our design is VDDA, which uses the LM top metal and is a low-resistivity plane. AGND is mainly used as a current return, and has much higher resistivity. A new active bias scheme, which distributes Vgs differentially has been designed, and should eliminate this problem.

- **Threshold control:** The very compact local DACs used for 5-bit threshold adjust have very poor linearity. In addition, the bias distribution issues mentioned above cause significant top/bottom variations. In order to optimize the predictability of the threshold tuning process, we need to improve the quality of the local DAC. New design will use 6-bit DAC based on identical unit cells (similar to the ones used at the bottom of column), and should result in much better linearity and control.
- **SEU-tolerance:** All configuration data (Command Register, Global Register, and Pixel Register) is stored in SEU-tolerant latches (40,547 per chip !) Initial measurements of upset rates of our SEU-tolerant latches at the 55MeV Cyclotron showed very low cross-sections. Measurements at the CERN PS (20GeV protons) showed more than an order of magnitude higher upset rates. This would lead to somewhat unreliable operation at the LHC, despite proposed global “periodic reset” every few hundred seconds in ATLAS TDAQ.
- We will improve the SEU-tolerance of FE-I2 by improving the layout of the latches in the Pixel latches, and by using a triple-redundancy scheme in the Command and Global latches. This should result in very stable configuration data even during operation of B-layer at design luminosity.
- SEU effects in the data paths (dynamic) are much harder to evaluate or measure. All state machines were designed to use no hidden states, and individual bit flips will normally have very localized effects (corruption/loss of individual hits), so effects are expected to be small. Further hardening of all this logic would be very challenging, and is not believed to be necessary.

- **Special pixels:** ATLAS pixel sensor contains four types of pixels, in order to provide 100% coverage in multi-chip module:



- Pixels at the two edges are 600μ instead of 400μ. Pixels at the top are ganged.
- Capacitive loads for the ganged pixels, as well as the normal pixels overlapped by ganging traces, are much higher. Plan to deal with this by using modified front-ends with 2*IP and 4*IP, in order to provide acceptable timewalk for all pixels.

Summary

- Lengthy design program has led to very sophisticated and high performance pixel arrays meeting all ATLAS pixel detector requirements.
- Present design contains 2.5M transistors for 2880 channels of readout. It is implemented in a commercial 0.25μ technology, using radiation-tolerant layout techniques to achieve 60MRad tolerance and good SEU hardness without latch-up or other fatal effects.
- First prototypes now extensively evaluated and meet all ATLAS requirements. Modest improvements being implemented in pre-production version of FE chip, to be submitted by the end of 2002.
- There remain uncertainties in the yield for the FE chip. Our experience with four groups of wafers, all made using the same mask set, has varied from 3% yield to 79% yield for simple selection criteria.