UNIVERSITY OF CALIFORNIA, IRVINE

Design and Scaling of a 16-Mega-Pixel CMOS Image Sensor for Electron Microscopy

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE in Electrical and Computer Engineering

by

Shiuh-hua Chiang

Thesis Committee: Professor Stuart Kleinfelder, Chair Professor Xuong Nguyen-Huu Professor Michael Green

© 2009 Shiuh-hua Chiang

The thesis of Shiuh-hua Chiang is approved:

.

Michael D Xubug Wgwycenhuu Stuart & Klefelder

Committee Chair

University of California, Irvine 2009

DEDICATION

То

My wife Camilla

and

my daughter Aeternia

TABLE OF CONTENTS

List of Figures	V
List of Tables	vii
Acknowledgments	'iii
Abstract of the Thesis	
Chapter 1 : Introduction	. 1
Chapter 2 : EM7 CMOS Image Sensor	4
Chapter 3 : Clock Distribution	. 8
3.1 Introduction	. 8
3.2 Interconnect Parasitics	8
3.3 MOS Device Loading	11
3.4 Clock Line Modeling	13
3.5 Column Shift-Register Clock Distribution	15
3.6 Row Shift-Register Clock Distribution	25
3.7 ADC Gray Code Distribution	31
3.8 Frame Rate	36
3.9 Summary	37
Chapter 4 : Power Distribution	39
4.1 Introduction	39
4.2 Pixel Current Source Power Distribution	41
4.3 Effect of IR Drop on Pixel Source Follower Performance	45
4.4 Opamp Power Distribution	51
4.5 Effect of IR Drop on Opamp Performance	56
4.6 Pixel Power Distribution	
4.7 Effect of IR Drop on Pixel Performance	63
4.8 Summary	63
Chapter 5 : Conclusion	64
References	65

LIST OF FIGURES

Figure 1: EM7 pixel and readout circuit block diagram.	4
Figure 2: EM7 block diagram at the global level	6
Figure 3: Side-by-side comparison of EM7 and EM5 layout.	7
Figure 4: Cross-sectional view of interconnect showing parasitic capacitance components	
Figure 5: Effective capacitance between two interconnects with a) both signals switching in the	
same direction, b) one signal switching and one at dc, and c) signals switching in the	-
opposite directions.	10
Figure 6: Cross-section of a MOS device with its parasitic capacitances shown.	
Figure 7: a) An inverter and b) a transmission gate with parasitic gate capacitances shown.	
Figure 8: Distributed <i>rc</i> model for the interconnect	
Figure 9: Dispersion of a signal in an interconnect.	
Figure 10: Shift-register showing master and slave stages.	
Figure 11: Column shift-register and counter buffer schematic.	
Figure 12: Column shift-register and counter buffer layout	
Figure 13: Two-phase clock schematic	
Figure 14: Two-phase clock layout	. 19
Figure 15: Block diagram of column clock distribution design with sub-sectioning	. 20
Figure 16: Layout of one readout section.	.21
Figure 17: Close-up view of the column shift-registers and the two-phase clock layout	22
Figure 18: Two-phase clock waveforms at the farthest end of the column shift-register from the	
clock, with sub-sectioning.	
Figure 19: Block diagram of column clock distribution design without sub-sectioning.	
Figure 20: Two-phase clock waveforms at the farthest end of the column shift-register from the	
clock, without sub-sectioning.	
Figure 21: Row shift-register and drivers schematic	
Figure 22: a) Row shift-register and b) driver layout.	
Figure 23: Block diagram of row clock distribution design with sub-sectioning.	
Figure 24: Layout of one row-addressing section.	
Figure 25: Close-up view of the row shift-registers, drivers, and the two-phase clock layout	. 28
Figure 26: Two-phase clock waveforms at the farthest end of the row shift-register from the	
clock, with sub-sectioning.	
Figure 27: Block diagram of row clock distribution design without sub-sectioning	
Figure 28: Two-phase clock waveforms at the farthest end of the column shift-register from the	e
clock, without sub-sectioning.	. 30
Figure 29: Gray counter schematic showing a) bit 0, b) bit 1-8, and c) bit 9.	32
Figure 30: Gray counter layout.	
Figure 31: Block diagram of Gray code distribution with sub-sectioning	
Figure 32: Gray code bit 0 waveform at the farthest end of the counter buffer from the driver,	
with sub-sectioning.	35
Figure 33: Block diagram of Gray code distribution design without sub-sectioning.	35
Figure 34: Gray code bit 0 waveform at the farthest end of the counter buffer from the driver,	. 55
	26
without sub-sectioning.	. 30
Figure 35: a) Linear array of N current sources b) modelled using N constant current sources	.40
Figure 36: Pixel current source layout	
Figure 37: Block diagram of current source power distribution design with sub-sectioning	
Figure 38: Ground IR drop for current sources with added ground supplies	
Figure 39: Block diagram of current source power distribution design without sub-sectioning	
Figure 40: Ground IR drop for current sources without added ground supplies	.44

Figure 41: a) Pixel and current source schematic and b) simplified circuit	45
Figure 42: Small-signal equivalent circuit of the pixel source follower	46
Figure 43: Pixel source follower gain versus supply IR drop.	47
Figure 44: High-frequency model of the pixel source follower	48
Figure 45: Pixel source follower -3dB bandwidth versus supply IR drop	50
Figure 46: Folded-cascode opamp schematic.	51
Figure 47: Folded-cascode opamp layout.	52
Figure 48: Block diagram of opamp power distribution design with sub-sectioning	52
Figure 49: Ground IR drop for opamp with sub-sectioning	53
Figure 50: Vdd IR drop for opamp with sub-sectioning.	54
Figure 51: Block diagram of opamp power distribution design without sub-sectioning	55
Figure 52: Ground IR drop for opamp without sub-sectioning	55
Figure 53: Vdd IR drop for opamp without sub-sectioning.	56
Figure 54: Opamp gain versus supply IR drop.	57
Figure 55: Opamp unity-gain bandwidth versus IR drop.	58
Figure 56: Pixel schematic	59
Figure 57: Pixel layout.	59
Figure 58: Block diagram of pixel power distribution design with added power supplies	60
Figure 59: Vdd IR drop in the pixel array with additional power supplies	61
Figure 60: Block diagram of pixel power distribution design without added power supplies	62
Figure 61: Vdd IR drop in the pixel array without additional power supplies	62

LIST OF TABLES

Table 1: Perf	ormance com	parison using	g the old and	d new cloc	k distribution	designs	
Table 2: Perf	formance com	parison using	g the old and	d new pow	er distribution	designs	63

ACKNOWLEDGMENTS

I would like to thank my advisor Professor Stuart Kleinfelder for his support, guidance, and technical insight during this research project. I would also like to thank my committee members, Professor Xuong Nguyen-Huu and Professor Michael Green for their time and effort in reviewing my dissertation.

I would like to thank my colleagues at UCI including Wei Huang, Di Wang, and Jeff Sloan for their work on the EM7 schematic and the pixel array layout verification. I thank Shengdong Li who designed EM5, upon which the design discussed herein was based. I also thank Liang Jin from UCSD for helpful discussions on different reset and read modes of the chip.

I thank my parents and my parents-in-law for their loving support. Above all, I thank my wife Camilla for her love and encouragement throughout my studies. This work would not have been possible without her. Last but not least, I thank my daughter Aeternia who arrived in the family during the course of my studies. Her adorable smiles and cuddly hugs always make my day after a day's hard work.

ABSTRACT OF THE THESIS

Design and Scaling of a 16-Mega-Pixel CMOS Image Sensor for Electron Microscopy

By

Shiuh-hua Chiang

Master of Science in Electrical and Computer Engineering University of California, Irvine, 2009 Professor Stuart Kleinfelder, Chair

The design and scaling of a large-scale $21 \times 21 \text{ mm}^2$ CMOS image sensor with digital readout for charged-particle imaging, "EM7," is presented. The sensor contains ~50 million transistors spanning its 16 million pixels, and includes over 4,100 parallel analogprocessing and A/D conversion circuits, over 4,100 double-buffered readout registers, and 12 parallel 10-bit readout busses for high data throughput. Scaling issues in moving from an earlier, smaller prototype, to the new large sensor are discussed. The clock distribution design in EM7 minimizes the clock delay by dividing the chip into multiple parallel sections driven locally by a tree-like clock structure. By this technique, simulations showed that the readout shift-register clock delay is reduced from 4.7 ns to 0.14 ns, and the row shift-register clock delay is reduced from 1.7 ns to 0.12 ns. With local buffering, the ADC Gray code counter delay is reduced from 35 ns to 0.9 ns. The improved architecture enables EM7 to sustain an image rate of 75 frames/s, for a continuous data throughput of over 10 Gb/s. The large chip dimensions and the increased power consumption in EM7 also require more robust power distribution. Utilizing higher metal layers and multiple supply points, a matrix-math simulation shows the pixel IR

drop is reduced from 20 mV to 8 mV. Similarly the pixel current source IR drop is reduced from 80.7 mV to 2.58 mV. The pixel source follower worst-case bandwidth is increased from 6.92 MHz to 14.4 MHz. The opamp ground IR drop is reduced from 236 mV to 20.2 mV, and VDD IR drop from 90.7 mV to 14.9 mV. The opamp gain variation is reduced from 525% to 28%. The worst-case opamp bandwidth is increased from 0.87 MHz to 764 MHz.

Chapter 1: Introduction

Electron microscopy images using electrons instead of photons, taking advantage of the fact that the equivalent wavelength of electrons is much smaller than that of visible light. Hence it provides higher resolution structural information for a wide range of research areas including material, physical, medical, and biological sciences. For example, cryoelectron microscopy can determine 3D structure of large protein complexes and viruses at 7-10 Å resolution [1]. In the past, film has been the medium of choice for recording images for electron microscopy due to its excellent modulation transfer function (MTF), typically 0.1 at 100 lines/mm [2]. However, drawbacks associated with film such as the tedious tasks of mechanical loading, development, and digitization prevents it from being used in applications where high frame rate and real-time data are required (for example, electron tomography) [3]. The emergence of charge-coupled devices (CCD) as image sensors provides an alternative method to film for electron microscopy imaging with more streamlined procedures [4]. However, due to radiation damage and signal saturation issues, a phosphorescent scintillation screen is required to first convert electron energy into photons before a CCD camera can perform imaging [5]. This indirect method of electron detection introduces deterioration in image resolution due to the scattering of electrons and light by the scintillation screen. Consequently, the resolution of the CCD method is limited by the scintillator, with MTF typically less than 0.1 at 20 lines/mm [6].

The disadvantages of CCD have prompted the development of CMOS image sensors for direct electron detection. CMOS image sensors utilize p-n junctions for the collection of charges generated from incident particles. The collected charges are then converted into

voltage signals before being processed and read out. Originally invented in the 1960's [7], CMOS image sensors have benefited from shrinking transistor dimensions, resulting in lower power consumption, lower cost, more integration of functionality on the same chip, miniaturization, higher resolution, and higher speed than CCD image sensors [8]. The trend of achieving higher pixel count, and therefore, higher resolution, has been ongoing in each new generation of CMOS image sensors [9, 10, 11, 12, 13], with the latest reported count at 52-mega-pixels by Iwane et al [14]. Several generations of CMOS image sensors customized for charged-particle imaging, including electron microscopy, have been designed, fabricated, and tested [15, 16, 17, 18, 19, 20, 21]. The most refined of these have achieved spatial resolution surpassing that of CCD-based electron microscopy sensors [22]. The first chip was EM1, an experimental chip for studying the influence of diode size on signal-to-noise ratio. EM1 was fabricated in standard 0.25 µm CMOS technology, and contains 50×50 pixels with analog readout. The next generation chip, EM2, studied the effect of pixel pitch value on electron lateral diffusion. EM2 was fabricated in 0.25 µm technology and contains four sensor array sectors that measure 360 \times 360, 180 \times 180, 45 \times 45, and 30 \times 30 pixels. Building on the results of its two predecessors, a full-scale sensor, EM3, was made in 0.25 µm technology. EM3 contains 512×550 pixels with analog readout circuitry, and demonstrated excellent spatial resolution and signal-to-noise ratio compared to CCD systems [23]. A more ambitious chip, EM4, was made in 0.25 μ m technology following the success of EM3. At 1024 \times 1024 pixels, EM4 contains nearly four times as many pixels as its predecessor, and uses 16 parallel analog readout channels. Major upgrade to the readout architecture was made in the next chip, EM5. Fabricated in 0.25 μ m technology, EM5 contains 460 \times 560 pixels

with on-chip noise-cancellation, programmable gain circuitry and analog-to-digital converters. In this thesis, the challenges of designing a next generation chip, EM7, by increasing EM5's resolution to 4140×3865 pixels (more than 62 times) are discussed, and appropriate solutions are presented.

Chapter 2: EM7 CMOS Image Sensor

To achieve high resolution images, EM7 uses building blocks from EM5 to form the basis of its pixel and readout design, and scales up the pixel count more than 62 folds to 4140×3865 , or just over 16-Mega-pixels. Figure 1 shows the design of a single pixel and its readout circuit:



Figure 1: EM7 pixel and readout circuit block diagram.

The pixel implements active pixel sensor (APS) architecture that consists of three NMOS transistors and a diode (shown in the dotted box). The transistors and the diode are placed on a continuous *p*-type epitaxial layer that extends throughout the entire pixel array area,

serving as the sensing region where incident electrons can liberate electrons from the lower energy bands. The excited electrons diffuse in the epitaxial layer and are eventually collected by the sensor diode. Because the epitaxial layer is continuous and the top circuitry appears transparent to the incident electrons, the proportion of the sensing region in each pixel, or "fill factor", is 100 %. The collected charges are converted to a voltage signal by the pixel's source follower transistor, and the output of the source follower appears at the column line through a row select transistor. The charges on the diode are cleared by the reset transistor, which connects the diode to the supply line. Each row of pixels in the chip shares a reset line and a select line, which are controlled by the row reset/select shift registers and drivers. The row reset has two modes – a global reset mode and a rolling reset mode. In the global reset mode, all rows are reset at the same time. The global reset mode is used with a mechanical shutter to begin the integration of signals by all the pixels at the same time. The rolling reset mode resets each row in succession, using a pointer inserted in the first cell of the row shift-register. Reading is similar to the rolling reset mode, except the select line is selected instead. Each column of pixels shares a column line, a current source, and per-column analog processing and analog-to-digital conversion circuitry. In the gain stage, a folded-cascode op amp with switched capacitors provides unity or $\times 10$ gain, offset cancellation, and correlated-double sampling (CDS). The gain stage is followed by a sample-and-hold and a single-slope Nyquist-rate ADC. The comparator in the ADC compares the pixel voltage against a voltage ramp as a 10-bit Gray counter sweeps across the equivalent set of values. Initially the ramp voltage is lower than the pixel voltage, and the comparator output is low, causing the counter buffer to load the Gray counter values. When the ramp voltage

exceeds the pixel voltage, the comparator output switches to high and freezes the Gray counter value in the counter buffer, thereby giving the digital representation of the pixel voltage. A 10-bit wide shifter register is clocked by a two-phase clock to sequentially shift out values stored in the counter latches. Figure 2 shows the block diagram of EM7 at the global level:



Figure 2: EM7 block diagram at the global level.

At the center is the pixel sensor array, with row reset/select control and clocking on the left. Two identical readout blocks are placed above and below the sensor array, with each

side processing either odd or even pixel columns. The readout block consists the aforementioned analog signal processing circuits, ADCs, shift registers, two-phase clocks, and a Gray counter. Figure 3 shows the layout of EM7, with EM5 placed next to it to demonstrate their relative sizes:



Figure 3: Side-by-side comparison of EM7 and EM5 layout.

With the chip dimensions measuring 21×21 mm, EM7 easily dwarfs EM5, which measures 3.1×4.8 mm. The increased size of EM7 creates issues in clock delay, skew, and IR drop that necessitate additional design considerations, all of which are discussed in detail in Chapter 3.

Chapter 3: Clock Distribution

3.1 Introduction

A clock synchronizes the flow of data in a chip by providing a temporal reference for a digital machine to determine the precise instance to change its state. For example, clock is used to trigger data from one set of sequential registers to the next through combinational logic that manipulates data in a functional manner, such as addition or multiplication as in the case of adders and multipliers. To distribute clock throughout the chip, global interconnects route one or more clocks from external pads to different chip sections, where local blocks such as buffers and local clock generators provide synchronization for the local elements. Optimized clock distribution reduces clock uncertainties such as delay and skew, thereby achieving high-performance with higher clock rates [24]. Different clock distribution topologies have been developed such as *grids* [25], *trees* [26], and *serpentines* [27], all aimed at minimizing clock uncertainties. As device size scales down and chip sizes increase, interconnect delay begins to dominate the total delay time [28]. The following sections discuss the factors that contribute to clock delay, and their modeling for simulation and analysis.

3.2 Interconnect Parasitics

Interconnect delay is caused by the parasitic resistance and capacitance of the physical wires. Resistance R_{int} is determined by using the following expression:

$$R_{\rm int} = \frac{\rho L}{WH} \tag{1}$$

Where ρ is the resistivity of the interconnect material, *L* is the length, *W* is the width, and *H* is the height of the interconnect. With ρ and *H* generally stay fixed for a given layer on a wafer, the ratio ρ/H , or sheet resistance, is typically given as a parameter in the foundary reports [29]. By multiplying L/W of the interconnect with the sheet resistance, the total resistance can be readily calculated.

The interconnect capacitance comes from the parasitic capacitance between the node of interest to its environment. The interconnect capacitance can be determined using the model shown in Figure 4:



Figure 4: Cross-sectional view of interconnect showing parasitic capacitance components.

The model shows the cross-section of three interconnect layers, with conductor 1 in the top layer, 2, 3, and 4 in the middle layer, and 5 in the bottom layer. The total capacitance of conductor 3 consists the following components: the area capacitances C_{a1} and C_{a2} between conductors in different planes, the fringe capacitance C_{f1} , C_{f2} , C_{f3} , and C_{f4} between conductors in different planes, and the lateral capacitance C_{11} and C_{12} between conductors in the same plane. Therefore, the total interconnect capacitance C_{int} of conductor 3 is [30]

$$C_{int} = C_{f1} + C_{f2} + C_{f3} + C_{f4} + C_{a1} + C_{a2} + C_{l1} + C_{l2}$$
(2)

If the geometry of the interconnects are known, the individual capacitance components can be calculated by using the per-unit area capacitance, per-unit length fringe capacitance, and per-unit length lateral capacitance that are tabulated in the foundary reports [29, 31]. Equation (2) can be further refined by scaling each capacitance component by a factor between 0 and 2. This factor is dependent on the switching characteristics of the signals that travel in the interconnects (Figure 5).



Figure 5: Effective capacitance between two interconnects with a) both signals switching in the same direction, b) one signal switching and one at dc, and c) signals switching in the opposite directions.

For example, when signals in two parallel interconnects switch in the same direction (Figure 5 a)), no charge is transferred and the effective capacitance between the two wires is zero. On the other hand, if one signal switches while the other is dc (Figure 5 b)), the effective capacitance is C. The worst case happens when the two signals switch in the opposite directions (Figure 5 c)), in which case the effective capacitance is 2C due to Miller effect. Using data sets that simulate the actual chip operations, the average scaling factor can be computed and applied to each capacitance component. Even more elaborate modeling methods have been proposed that take into account of 3D field effects [32, 33]. However such methods tend to be slow and computationally expensive. In this thesis Equation (2) is used for the simulations.

3.3 MOS Device Loading

The loading on a clock comes not only from the interconnect parasitics but also from the parasitic capacitance of the devices it drives. Shown in Figure 6 is a cross-section of a MOS device with its parasitic capacitances shown:



Figure 6: Cross-section of a MOS device with its parasitic capacitances shown.

 C_1 and C_2 are the gate-to-source/drain overlap capacitances defined by

$$C_1 = C_2 = W C_{ov} \tag{3}$$

Where W is the width of the transistor and C_{ov} is the per-unit length capacitance that takes into account of the overlapping and fringing effects. C_3 and C_4 are the source/drain-tobulk junction capacitances defined by

$$C_3 = C_4 = C_j L_s W + C_{jsw} (2L_s + W)$$
(4)

The first term on the right-hand side represents the bottom-plate capacitance of the diffusion area. C_j is the per-unit area junction capacitance, L_s is the length of the diffusion area, and W is the width of the transistor. The second term represents the side-wall capacitance of the diffusion area, with C_{jsw} being the per-unit length side-wall capacitance. The junction capacitance C_j in Equation (4) is given by

$$C_{j} = C_{j0} / (1 + V_{R} / \Phi_{B})^{m}$$
(5)

Where V_R is the reverse voltage across the junction, Φ_B is the built-in junction potential, and *m* is the grading coefficient. From Figure 6, C_5 is the gate-to-channel/bulk capacitance and has a value of $2/3C_{ox}WL$ in saturation, $C_{ox}WL$ in triode, and $C_{ox}WL||C_d$ in cutoff, where C_{ox} is the per-unit area capacitance of the oxide and C_d is the deletion region capacitance between the channel and the bulk. C_5 can be merged with C_1 if operating in the saturation region, or divided equally between C_1 and C_2 if operating in the triode region [31].

Typical load elements for a clock line are inverters and transmission gates, shown in Figure 7:



Figure 7: a) An inverter and b) a transmission gate with parasitic gate capacitances shown.

For an inverter, the gate capacitance consists of the gate-to-source and gate-to-drain capacitances, with the effective gate-to-drain capacitance multiplied by two due to Miller Effect. For a transmission gate, each clock drives the gate-to-drain and gate-to-source capacitances of a single transistor. Since during switching the transistors go through

different regions of operation, the capacitances exhibit a non-linear response. In this thesis, simulation is used predict the clock performance by inserting load transistors at the appropriate points in the interconnect model.

3.4 Clock Line Modeling

With R_{int} and C_{int} known, the interconnect can be modeled as shown in Figure 8:



Figure 8: Distributed rc model for the interconnect

The R_{int} and C_{int} are distributed over the interconnect length L, with per-unit length resistance r and capacitance c shown in the figure. Writing KCL, the voltage at node i can be expressed as

$$c\Delta L \frac{\partial V_i}{\partial t} = \frac{(V_{i+1} - V_i) + (V_{i-1} - V_i)}{r\Delta L}$$
(6)

If the number of segments is large, or equivalently $\Delta L \rightarrow 0$, the equation becomes

$$rc\frac{\partial V}{\partial t} = \frac{\partial^2 V}{\partial x^2} \tag{7}$$

Equation (7) is known as the *diffusion equation*, where *V* is the voltage at distance *x* from the input. Unfortunately, no closed-form solution exists for this equation. Various approximations of the solution for Equation (7) can be found in [34]. For computer-aided analysis, the interconnect model must be constructed with a finite number of segments to approximate the distributed delay line. Elmore delay formula can then be used to estimate the delay τ of an *N*-segment *rc* network [35]:

$$\tau = \left(\frac{L}{N}\right)^{2} \left(rc + 2rc + ... + Nrc\right) = L^{2}rc\frac{N+1}{2N} = R_{\text{int}}C_{\text{int}}\frac{N+1}{2N}$$
(8)

Where *L* is the total length of the interconnect, *N* is the number of segments, *r* and *c* are the per-unit length interconnect resistance and capacitance respectively. For an *N*segment model, the value of each the resistor and capacitor are *rL/N* and *cL/N* respectively. From Equation (8), if the number of segments is large, τ approaches $R_{int}C_{int}/2$. For the simulations in this thesis, the number of segments are generally >30. Equation (8) shows that the delay is proportional to the square of the length of the wire. Therefore, the farther away from the signal source the more dispersed the waveform becomes. Figure 9 shows the waveforms at different points in the interconnect model of Figure 8:



Figure 9: Dispersion of a signal in an interconnect.

 V_{in} is a sharp step input to the interconnect. Because of the *rc* parasitics, the waveform V_i at node *i* has a finite rise time with a delay of τ_i with respect to the input. As the signal travels to the end of the interconnect, the waveform becomes more dispersed, having a delay of τ_N at node *N*. To simulate the loading, transistors are inserted at the appropriate points in the interconnect model. The increased capacitance at these nodes will further slow down the signal. In a digital system where the clock signal is distributed over long

interconnects, the delay seriously degrades the performance of the system by placing a limit on the maximum frequency of the clock. Should the clock frequency exceed this limit the signal at the far end of the interconnect will fail to reach proper logic levels, causing catastrophic failure of the system. Further, the different arrival times of the clock to various location in the circuit, or *clock skew*, can potentially cause system malfunctioning by violating the setup and hold times. Therefore, to increase the performance of the system it is of prime concern to reduce the clock delay and skew by reducing (1) the parasitic rc of the interconnect, (2) transistor loading, and (3) the distance the clock signal has to travel. The following sections discuss the signal distribution designs in EM7.

3.5 Column Shift-Register Clock Distribution

EM7 uses shift-registers to sequentially output the digitized pixel values from the counter buffer, as explained in Chapter 2. A block diagram of a shift-register is shown in Figure 10:



Figure 10: Shift-register showing master and slave stages.

The shift register consists of two stages – a master stage and a slave stage. Each stage is implemented as a D-latch. A Two-phase clock is used to time the shift register. In the first phase, CKB is high, and the master is transparent to pass the input to the intermediate

node, while the slave is open to hold the previous output value. In the second phase, *CKB* is low, and the master holds while the slave passes the value in the intermediate node to *Out*. By linking multiple shift-registers in series, data can be propagated to an output port sequentially. *CK* and *CKB* must have non-overlapping phases, lest both the master and slave stages are transparent at the same time, destroying the value in the intermediate node. The schematic and layout of EM7's column shift register and counter buffer are shown in Figure 11 and Figure 12 respectively:



Figure 11: Column shift-register and counter buffer schematic.



Figure 12: Column shift-register and counter buffer layout

The shift-register uses clock signals p1, p1b, p2, and p2b generated from a two-phase clock. p1 and p2 have non-overlapping phases to ensure that the slave and master stages are never transparent at the same time. p1b and p2b are simply the compliment of p1 and p2, respectively. The master stage consists of a transmission gate switch followed by an *NMOS* transistor and a *PMOS* transistor whose drain nodes are connected to the switch of the slave stage. The slave stage is completed with an inverter. As explained in Chapter 2, the Gray counter value is loaded into the counter buffer during the ADC conversion. The Gray counter value dc is shown to the left in the schematic. It goes to the Gray counter buffer consists of an *NMOS* transistor and a *PMOS* transistor whose drains are connected to the buffer output switch, which is controlled by tr, trb. During Gray counter loading both ld and tr are low. ld goes high when the input ramp to the comparator exceeds the pixel value, freezing the Gray counter in the buffer. The value

stored in the buffer is transferred into the shift-register by pulsing *tr*. Thus the counter buffer allows simultaneous ADC conversion and readout by double-buffering. The combining of an inverter and a transmission gate switch (also called C^2MOS [36]) made it possible for a compact layout that fits within the pitch defined by the widths of two pixels (since each pixel column requires a shift-register, and there are top/bottom readout sides). The entire layout cell, consisting of three inverters, four switches, and eight control signal wires, occupies an area less than $10 \times 16 \mu m$. The schematic and the layout of the two-phase clock are shown in Figure 13 and Figure 14 respectively.



Figure 13: Two-phase clock schematic.



Figure 14: Two-phase clock layout.

EM7 contains 4,140 pixel columns spanning a total length of over 20 mm. Reading out the pixel columns from the top and bottom sides of the chip (Figure 2), each readout side contains 2,070 10-bits wide shift-registers for a total of 20,700 shift-register cells. From Equation (8), the time constant of the clock waveform is proportional to the square of the length of the interconnect. Therefore, reducing the distance the clock signal has to travel will have a strong impact in keeping the clock waveforms ideal. For example, if the distance is reduced by half, the time constant of the clock waveform will be reduced by three-fourth. The sheer number and dimension of the readout logic elements in EM7 necessitate the sub-sectioning of the column shift-registers and the introducing of local buffers and clock generators for more efficient clock distribution. Figure 15 shows the block diagram of EM7's readout design:



Figure 15: Block diagram of column clock distribution design with sub-sectioning.

The readout shift-registers in EM7 are divided into six identical sections. A global clock *Ext Ck* is supplied externally and routed to the six local two-phase clocks via global interconnects buffered by inverters. Each two-phase clock drives 345 10-bit wide shift-registers. The 2,070 ADCs provide digitized 10-bit pixel values. Each of the six sections has its own set of 10 output pads, giving a total of 12 parallel 10-bit readout channels for the chip. Figure 16 shows the layout of one complete readout section:



Figure 16: Layout of one readout section.

The gray area across the top half of Figure 16 shows the analog processing circuitry and ADCs, and the darker area just below shows the column shift-registers. The data propagates from right to left, with the readout bus and the two-phase clock placed at the extreme left of the section. By having the data and clock travel in opposite directions, the hold-time of the registers is unconditionally met, ensuring no race conditions [37]. The digital output pads are near the bottom of Figure 16, showing the interleaving power and ground pads to provide shielding to reduce capacitive and inductive coupling between the digital output bond wires, and to decrease the equivalent inductance for the bond wires. Figure 17 provides a close-up view of the area at the extreme left of the readout section:



Figure 17: Close-up view of the column shift-registers and the two-phase clock layout.

The vertical repeating pattern shows the 10-bit width of the shift-registers, and the horizontal repeating pattern shows the length of the shift-registers. The wires running vertically to the extreme left of the figure form a 10-bit wide output bus leading to the output pads. The thick vertical wire near the center of the figure is the power supply line. The two-phase clock can be seen just below the shift-registers. To its right are the Gray counter input inverters and bus. The layout of the column shift-registers is simulated and the clock waveforms at the farthest end (worst case) from the two-phase clock are shown in Figure 18:



Figure 18: Two-phase clock waveforms at the farthest end of the column shift-register from the clock, with sub-sectioning.

The clock signals show non-overlapping phases between p1 (solid line) and p2 (dotted line) with an input clock frequency of 100 MHz. The clock delay between the closest shift-register and the farthest shift-register is 0.14 ns. For comparison, the clock distribution design from EM5 is directly used in the scaled-up EM7 chip without modification (no sub-sectioning). The block diagram of this design is shown in Figure 19:



Figure 19: Block diagram of column clock distribution design without sub-sectioning. Without sub-sectioning, a two-phase clock would drive 2,070 column shift-registers. Using the same 100 MHz clock, the simulation of the clock waveforms at the farthest end of the column shift-registers is shown in Figure 20:



Figure 20: Two-phase clock waveforms at the farthest end of the column shift-register from the clock, without sub-sectioning.

With the increased distance and loading, the two-phase clock signals fail to reach the proper logic levels due to their large time constant. The result is the catastrophic malfunctioning of the readout circuitry. The clock delay between the closest and farthest shift-register is 4.7 ns.

3.6 Row Shift-Register Clock Distribution

EM7 uses shift-registers to sequentially address each pixel row for reset and read, as explained in Chapter 2. The schematic and layout of a row shift-register are shown in Figure 21 and Figure 22 respectively:



Figure 21: Row shift-register and drivers schematic


Figure 22: a) Row shift-register and b) driver layout.

Similar to the column shift-registers, the row shift-register uses clock signals *p1*, *p1b*, *p2*, and p2b generated from a two-phase clock. p1 and p2 have non-overlapping phases and *p1b* and *p2b* are the compliment of *p1* and *p2*, respectively. A *NAND* gate in the master stage and a NOR gate in the slave stage are used to reset the row-shift register. The drivers consist of combinational logic for choosing rolling/global/read mode, and two large inverters to drive the pixel reset and select lines. Since each of the 3,865 pixel rows has its own shift-register and drivers, there are a total of 3,865 row shift-registers in series. A special shift-register is placed at the head of the chain with the locations of its NAND and NOR gates swapped. When *rrst* is high (and *rrstb* low), a pointer is inserted into the head shift-register while all the other shift-registers are reset. To reset the pixels in the rolling reset mode, *rollingb* is set low, *global* is low, and *readb* is high. By clocking the shift-registers the pointer propagates through the chain and resets the pixel rows sequentially. Pixel read uses similar operation by setting *readb* low, *rollingb* high, and global low. To reset all the pixels at once, global is set high. Due to the large number of row shift-registers and the distance the clock signals have to travel, four two-phase clocks are used. Figure 23 shows the row addressing design:



Figure 23: Block diagram of row clock distribution design with sub-sectioning.

The row shift-registers in EM7 are divided into four sections. A global clock *Ext Ck* is supplied externally and routed to the four local two-phase clocks via global interconnects buffered by inverters. The routing uses a tree structure to ensure the clock skew to the four two-phase clocks is minimized. Each two-phase clock drives 966 row shift-registers. In addition, the four clocks' outputs are connected with each other to synchronously drive the row shift-registers. Figure 16 shows the layout of one complete row-addressing section:



Figure 24: Layout of one row-addressing section.

The gray areas across the top are the pixels. Below the pixels are the row shift-registers and drivers, shown as darker horizontal stripes. The power and ground pads are at the bottom. The two-phase clock is the small gray rectangle in the center of the section. By placing clock in the center, clock skew at the boundary between sections is minimized. Figure 25 provides a close-up view of the area at the center of the row-addressing section (rotated by 90°):



Figure 25: Close-up view of the row shift-registers, drivers, and the two-phase clock layout.

The two-phase clock is seen to the left. The row shift-registers and drivers are in the middle. To the extreme right is the first column of pixels. Because each row of pixels requires a cell of shift-register and drivers, the chain of cells make repeating patterns every 5 μ m. The row address pointer propagates from bottom to top in the figure. The layout of the row shift-registers is simulated and the clock waveforms at the farthest end (worst case) from the two-phase clock are shown in Figure 26:



Figure 26: Two-phase clock waveforms at the farthest end of the row shift-register from the clock, with sub-sectioning.

The clock signals shows non-overlapping phases between p1 (solid line) and p2 (dotted line) with an input clock frequency of 100 MHz. The clock delay between the closest shift-register and the farthest shift-register is 0.12 ns. For comparison, the clock distribution design from EM5 is directly used in the scaled-up EM7 chip without modification (no sub-sectioning). The block diagram of this design is shown in Figure 27:



Figure 27: Block diagram of row clock distribution design without sub-sectioning.

Without sub-sectioning, a two-phase clock would drive 3,865 row shift-registers. Using the same 100 MHz clock, the simulation of the clock waveforms at the farthest end of the column shift-registers is shown in Figure 28:



Figure 28: Two-phase clock waveforms at the farthest end of the column shift-register from the clock, without sub-sectioning.

With the increased distance and loading, the two-phase clock signals show much longer rise and fall times. The reduced noise margin makes the circuit less robust and more prone to malfunctioning. The clock delay between the closest and farthest shift-register is 1.7 ns.

3.7 ADC Gray Code Distribution

EM7 uses 10-bit ADCs to convert the pixel analog voltage into digital values. As explained in Chapter 2, an ADC compares the pixel voltage against a voltage ramp as a Gray counter simultaneously counts up. When the ramp exceeds the pixel voltage the comparator freezes the Gray code in the counter buffer. Thus the Gray code stored is the digital representation of the pixel voltage. The Gray counter counts from 0 to 2^{10} - 1 with only one bit making a transition between any two successive values, preventing large quantization error from the asynchronous comparator latch. The counter buffer is placed next to the column shift-register for readout. The schematic of a counter buffer, combined with a column shift-register is shown in Figure 11, and the layout is shown in Figure 30 respectively:



Figure 29: Gray counter schematic showing a) bit 0, b) bit 1-8, and c) bit 9.



Figure 30: Gray counter layout.

The 10-bit counter logic can be seen as the repeating cells in Figure 30, with the 10-bit output bus at the bottom. The Gray counter uses timing signals generated from a two-phase clock (not shown) to increment the code every half clock cycle. Therefore, to count from 0 to 2^{10} - 1, the counter uses $2^{10} / 2 = 512$ clock cycles. There are two Gray code counters in EM7, one for the top readout circuit and one for the bottom (Figure 2). Each counter is physically located on the right side of the chip, and a 10-bit bus distributes the Gray code to 2,070 counter buffers. Figure 31 shows the block diagram of the Gray code distribution design:



Figure 31: Block diagram of Gray code distribution with sub-sectioning.

Dividing the readout into six identical sections, each section contains 345 10-bit counter buffers for a total of 3,450 counter buffer cells. The Gray code is first distributed to the local tapered drivers for each section. The layout of the local drivers and the counter buffers is shown in Figure 17. The tapered drivers are just to the right of the two-phase clock. The layout is simulated and the Gray code waveform for bit 0 at the farthest end from the driver is shown in Figure 32:



Figure 32: Gray code bit 0 waveform at the farthest end of the counter buffer from the driver, with sub-sectioning.

The Gray code reaches proper logic levels with a 100 MHz external clock. The counter delay between the closest buffer and the farthest buffer is 0.9 ns. For comparison, the Gray code distribution design from EM5 is directly used in the scaled-up EM7 chip without modification (no sub-sectioning and local buffers). The block diagram of this design is shown in Figure 33:



Figure 33: Block diagram of Gray code distribution design without sub-sectioning.

Without sub-sectioning, the Gray counter would drive 2,070 column shift-registers. Using the same 100 MHz clock, the simulation of the bit 0 waveforms at the farthest end from the counter is shown in Figure 34:



Figure 34: Gray code bit 0 waveform at the farthest end of the counter buffer from the driver, without sub-sectioning.

With the increased distance, loading and reduced driving, the Gray code signal fails to reach proper logic levels, causing the catastrophic malfunctioning of the data conversion circuitry. The clock delay between the closest and farthest buffer is 35 ns.

3.8 Frame Rate

Whereas EM5 uses single-buffering for the Gray code counter, EM7 uses doublebuffering to enable simultaneous A/D conversion and readout operations. The frame rate calculation for EM7 is as follows: Running at 10-bit resolution for the A/D conversion and clocking the Gray code counter with a 100-MHz clock, the time it takes to count from 0 to 2^{10} -1 is $2^{10} / (2 \times 100M) = 5.12$ µs. The factor of two in the denominator is because the Gray code counter increments at both the rising and falling edges of the clock. Since there are 12 parallel readout channels and each channel contains 345 readout shift-register columns that operate at 100-MHz, the time it takes for the digital readout is 345 / 100M = 3.45 µs. Taking the maximum of 5.12 µs and 3.45 µs (since theses two operations are done in parallel), the total readout time per row is 5.12 µs. Reading 3,865 rows, the total time per frame is $5.12\mu \times 3,865 =$ 19.8 ms or 50.5 frames/s. If the resolution reduced to 9-bits, the A/D conversion time is reduced to $2^9 / (2 \times 100M) = 2.56$ µs, and the total time per frame is max(2.56µ, 3.45µ) × 3865 = 13.3 ms or 75 frames/s. The average throughput at 75 frames/s is 75 × 3,865 × 4,140 × 9 = 10.8 Gb/s. Further reducing the ADC resolution will not increase the frame rate, as the speed bottleneck then becomes the digital readout time.

For comparison, if only single-buffering is used the total time per frame with 10-bit A/D conversion would be $(5.12\mu + 3.45\mu) \times 3,865 = 33.1$ ms or 30.2 frames/s. With 9-bit A/D conversion the total time is $(2.56\mu + 3.45\mu) \times 3,865 = 23.2$ ms or 43.1 frames/s.

3.9 Summary

Comparison of the performances for the old and new clock distribution designs are shown in Table 1:

	Old design	New design
Column shift-register clock delay (ns)	4.7	0.14
Row shift-register clock delay (ns)	1.7	0.12
Gray code delay (ns)	35	0.9
Frame rate at 9-bit A/D conversion (frames/s)	43.1	75.0

Table 1: Performance comparison using the old and new clock distribution designs.

Chapter 4: Power Distribution

4.1 Introduction

Robust power distribution is vital in integrated circuit design to ensure reliable operation at the guaranteed performance. Due to parasitic resistance in the power grid, current flowing in the conductive layers will introduce reduction in voltage, or IR drop, between the voltage sources and the power/ground rails of the load elements. The reduced rail voltages will decrease the speed and noise margin of the circuit, degrading performance and potentially causing system failures [38]. IR drop can be divided into two types: dynamic and static IR drops. Dynamic IR drop concerns with the transient behavior of the circuit and is modeled using a linear RCL network and independent time-variant current sources as the switching elements [39]. Static IR drop deals with constant voltage reductions due to time-invariant current sources [40]. For example, a linear array of Ncurrent sources in a resistive power distribution network can be modeled using Nindependent current sources (Figure 35):



Figure 35: a) Linear array of *N* current sources b) modelled using *N* constant current sources.

Assuming $I_1 = I_2... = I_N = I$, the voltage V_i at the *i*th node can be written using superposition as

$$V_{i} = \frac{IR(N+1-i)}{N+1} + \frac{2IR(N+1-i)}{N+1} + \dots + \frac{iIR(N+1-i)}{N+1} + \frac{iIR(N+1-(i+1))}{N+1} + \frac{iIR(N+1-(i+2))}{N+1} + \dots + \frac{iIR(N+1-N)}{N+1}$$
(9)
$$= \frac{IR(N+1)i - IRi^{2}}{2}$$

The node where maximum IR drop occurs can be determined by differentiating Equation (9) with respect to *i* and solving for the root:

$$\frac{\partial V_i}{\partial i} = \frac{IR(N+1)}{2} - IRi = 0$$

$$i\Big|_{\max V} = \frac{N+1}{2}$$
(10)

This result indicates that the maximum IR drop occurs in the center of the array, which makes intuitive sense. Inserting this value of i back to Equation (9), we get

$$V_{\rm max} = \frac{IR(N+1)^2}{8}$$
(11)

Equation (11) shows that the maximum IR drop V_{max} is proportional to the current and resistance and to the square of the number of branches. Therefore, to minimize IR drop the current, resistance, and especially, the number of branches must be minimized. In arriving at this result, the current in each branch was assumed to be equal. In reality, a drop in rail voltage will cause the current to decrease because of the reduced V_{GS} ($I \propto (V_{GS}-V_{th})^2$). Therefore, the maximum IR drop will be smaller than predicated by Equation (11), and can be solved numerically or through circuit simulation by taking into account of the current source's non-linear voltage-dependence. The following sections present the power distribution designs in EM7.

4.2 Pixel Current Source Power Distribution

Each pixel column in EM7 uses a current source to provide the bias current for the pixel's source follower (Figure 1). The current source is an NMOS transistor with the gate voltage defined by a current mirror. The transistor has a W/L ratio of 13.5/2.7 and carries a nominal current of 10 μ A. The layout of the current source is shown in Figure 36:



Figure 36: Pixel current source layout

The pixel column metal interconnect is to the far right of the figure. EM7 has two readout sides with each side having 2,070 current sources. As explained in Section 4.1, to minimize the IR drop the current, resistance, and/or number of branches must be reduced. With the large number of current sources in EM7, multiple ground supply points are used to reduce IR drop, as shown in Figure 37:



Figure 37: Block diagram of current source power distribution design with subsectioning.

Using a total of seven ground supply pads, the current sources are divided into six sections, each having 345 current sources. In addition, Metal4 that was not utilized in the previous generation sensor was used to increase the total width of the ground interconnect, reducing the resistance. The IR drop is simulated including the interconnect resistance and the voltage-dependence of the current sources. The results are shown in Figure 38:



Figure 38: Ground IR drop for current sources with added ground supplies.

The x-axis represents the current sources numbered from 1 to 2,070 along the chip's xaxis. The y-axis represents the IR drop at each of the current source's ground node. The IR drop is zero near the ground supplies, and increases to a maximum of 2.58 mV at halfway point between two ground supplies. The current varies from 10 μ A (near ground supplies) to 9.63 μ A (halfway between ground supplies). For comparison, the power distribution design from EM5 is directly used in the scaled-up EM7 chip without modification (only two ground pads, no extra Metal4 layer). The block diagram of this design is shown in Figure 39:



Figure 39: Block diagram of current source power distribution design without subsectioning.

Without sub-sectioning, there are 2,070 current sources between the two ground supplies. The IR drop for this design is simulated and the results are shown in Figure 40:



Figure 40: Ground IR drop for current sources without added ground supplies.

The increased resistance and number of branches cause the maximum IR drop to reach 80.7 mV. The current varies from 10 μ A near supply points to 2.21 μ A at halfway points.

4.3 Effect of IR Drop on Pixel Source Follower Performance

To study the effect of IR drop on the pixel source follower performance, the circuit is analyzed (Figure 41):



Figure 41: a) Pixel and current source schematic and b) simplified circuit.

Figure 41 a) is the schematic of the pixel and the current source. Reducing to just the input and the load transistors, b) shows the simplified circuit where R is the sum of the resistance of the column interconnect and the equivalent resistance of the select transistor when operated in the triode region. The small-signal equivalent circuit of b) is shown in Figure 42:



Figure 42: Small-signal equivalent circuit of the pixel source follower.

where g_m and gmb are M_l 's transconductance and body transconductance respectively. The output voltage can be expressed as

$$V_{out} = g_{m1} V_{gs1} \left[\left(R + r_{O2} \right) \| r_{O1} \| \frac{1}{g_{mb1}} \right] \frac{r_{O2}}{R + r_{O2}}$$
(12)

and V_{gs1} can be written as

$$V_{gs1} = V_{in} - g_{m1} V_{gs1} \left[\left(R + r_{O2} \right) \| r_{O1} \| \frac{1}{g_{mb1}} \right]$$
(13)

Substituting (13) into (12) and solving for the gain $A_v = V_{out} / V_{in}$, the result is

$$A_{\nu} = \frac{\left(R + r_{O2}\right) \|r_{O1}\| \frac{1}{g_{mb1}}}{\frac{1}{g_{m1}} + \left(R + r_{O2}\right) \|r_{O1}\| \frac{1}{g_{mb1}}} \cdot \frac{r_{O2}}{R + r_{O2}}$$
(14)

To gain more insight, Equation (14) is simplified by ignoring channel-length modulation (r_O is assumed to be large). The expression for A_v thus reduces to

$$A_{v} = \frac{1}{1+\eta} \tag{15}$$

where

$$\eta = \frac{g_{mb}}{g_m} = \frac{\gamma}{2\sqrt{2\Phi_F + V_{SB}}}$$
(16)

 γ is the body effect coefficient, Φ_F is the work function of polysilicon gate and silicon substrate, and

$$V_{SB} = V_{IN} - V_{th} - \sqrt{2I_D / k_n (W/L)_1}$$

= $V_{IN} - V_{th} - \sqrt{\frac{(W/L)_2}{(W/L)_1}} (V_{ov2})$ (17)

Equations (15), (16), and (17) show that as the IR drop increases, the overdrive voltage V_{ov2} decreases, V_{SB} increases, η decreases, and A_v increases. The gain variation among pixels increases fixed-pattern noise (FPN) for the readout and is an undesirable effect. Simulation of the source follower gain versus IR drop is shown in Figure 43:



Figure 43: Pixel source follower gain versus supply IR drop.

Source followers using the old power distribution design have gains that vary from 0.826 V/V at supply points to 0.843 V/V at the maximum IR drop point (2% gain variation). The source followers with the new power distribution design show more consistent performance, with gains that vary between 0.826 and 0.827 V/V (0.1% variation). The gain variation can be corrected in software after readout by using a calibration reference, at the expense of having more complicated post-processing steps.

The frequency response of the source follower also has a dependency on the IR drop. Figure 44 show the high-frequency model of the pixel source follower:



Figure 44: High-frequency model of the pixel source follower.

 C_L is the interconnect capacitance plus the output load capacitance. Channel length modulation and body effect have been neglected to simplify the analysis. Summing the current at V_{out} :

$$\frac{V_X - V_{out}}{R} = V_{out} C_L s \tag{18}$$

Summing the current at V_x :

$$(V_{gs} - V_{x})C_{GS}s + g_{m}V_{gs} = V_{X}C_{SB}s + \frac{V_{X} - V_{out}}{R}$$
(19)

Also, V_{in} can be written as

$$V_{in} = V_X + V_{gs} \tag{20}$$

Solving for V_x from Equation (18) and substituting it into Equation (19), then solving for V_{gs} and substituting all into Equation (20), the transfer function is obtained as

$$\frac{V_{out}}{V_{in}}(s) = \frac{g_m + C_{GS}s}{(2C_{GS} + C_{SB})RC_Ls^2 + (g_m RC_L + 2C_{GS} + C_{SB} + C_L)s + g_m}$$
(21)

Equation (21) shows that there is a zero and two poles in the left-half plane. The denominator of Equation (21) can be written in the following form:

$$D = (1 + s/\omega_1) \cdot (1 + s/\omega_2)$$

= $\frac{s^2}{\omega_1 \omega_2} + \frac{\omega_1 + \omega_2}{\omega_1 \omega_2} s + 1$ (22)

where ω_1 and ω_2 are the magnitudes of the pole frequencies. Assuming that the second pole is much faster than the first pole $|\omega_2| >> |\omega_1|$ [41], Equation (22) reduces to

$$D \approx \frac{s^2}{\omega_1 \omega_2} + \frac{1}{\omega_1} s + 1 \tag{23}$$

Comparing the second term in Equation (23) with the second term in (21), the expression for the magnitude of the first pole is arrived:

$$\omega_{1} = \frac{g_{m}}{g_{m}RC_{L} + 2C_{GS} + C_{SB} + C_{L}}$$
(24)

Since the interconnect and load capacitance C_L typically dominates over the other parasitic capacitances, Equation (24) can be reduced to

$$\omega_1 \approx \frac{1}{\left(R + 1/g_m\right)C_L} \tag{25}$$

The transconductance g_m is given by

$$g_{m} = \sqrt{2k_{n}(W/L)_{1}I_{D}} = k_{n}\sqrt{(W/L)_{1}(W/L)_{2}}V_{ov2}$$
(26)

Therefore, as IR drop increases, V_{ov2} decreases, g_m decreases, and ω_1 decreases. The reduced pole frequency increases the response time of the source follower. Simulation of the source follower's -3dB bandwidth versus IR drop is shown in Figure 45:



Figure 45: Pixel source follower -3dB bandwidth versus supply IR drop.

Figure 45 shows that for the old power distribution design, the -3dB bandwidths of the source followers vary between 14.6 MHz near supply points and 6.92 MHz at the maximum IR drop point (53% bandwidth variation). The source followers with the new power distribution design has more consistent -3dB bandwidths, varying between 14.6

MHz and 14.4 MHz (1.4% variation). To account for the bandwidth variation, the subsequent stages must wait for the slowest source follower to settle to within a required precision, slowing down the overall readout speed.

4.4 **Opamp Power Distribution**

The gain stage in EM7 uses an opamp in the closed-loop configuration with switched capacitors to programmably achieve a unity or $\times 10$ gain. The comparator uses another opamp in the open-loop configuration for analog-to-digital conversion (Chapter 2). The opamps are implemented as a folded-cascode amplifier whose schematic and layout are shown in Figure 46 and Figure 47 respectively:







Figure 47: Folded-cascode opamp layout.

The opamp uses PMOS transistors as the input pair and a wide-swing current mirror for single-ended output. The bias voltages are generated from current mirrors. The opamp layout fits within the 10- μ m pitch defined by two pixels. The opamp consumes a nominal static current of 69 μ A. To reduce IR drop, the power and ground wires in EM7 have been widened by using the Metal4 layer. Seven power and ground supply points are added to divide the linear array of opamps into six sections, each section having 345 opamps (Figure 48):



Figure 48: Block diagram of opamp power distribution design with sub-sectioning.

The IR drop for Vdd and ground are simulated, taking into account of the resistance and voltage-dependent static current consumption. The Vdd IR drop and ground IR drop simulation results are shown in Figure 49 and Figure 50 respectively:



Figure 49: Ground IR drop for opamp with sub-sectioning.



Figure 50: Vdd IR drop for opamp with sub-sectioning.

The IR drop reaches a maximum 20.2 mV and 14.9 mV for ground and Vdd rails respectively. The asymmetry is due to the slightly wider Vdd wire than the ground wire due to layout space, resulting in smaller resistance in the Vdd wire. The static current varies between 69 μ A and 57 μ A. For comparison, the power distribution design from EM5 is directly used in the scaled-up EM7 chip without modification (only two supply points, no extra Metal4 layer). The block diagram of this design is shown in Figure 51:



Figure 51: Block diagram of opamp power distribution design without sub-sectioning.

Without sub-sectioning, there are 2,070 opamps between the two supply points. The IR drop for this design is simulated and the ground and Vdd IR drops are shown in Figure 52 and Figure 53 respectively:



Figure 52: Ground IR drop for opamp without sub-sectioning.



Figure 53: Vdd IR drop for opamp without sub-sectioning.

The increased resistance and number of branches cause the maximum ground and Vdd IR drops to reach 236 mV and 90.7 mV respectively. The static current varies between 69 μ A and 1.7 μ A. The large IR drop causes the opamp's current sources enter the cutoff region, causing circuit failure.

4.5 Effect of IR Drop on Opamp Performance

The opamp shown in Figure 46 has a gain of

$$A_{v} \approx g_{m1,2} \left\{ \left[\left(g_{m7,8} + g_{mb7,8} \right) r_{O7,8} r_{O9,10} \right] \| \left[\left(g_{m3,4} + g_{mb3,4} \right) r_{O3,4} \left(r_{O1,2} \| r_{O5,6} \right) \right] \right\}$$
(27)

If assuming all g_m 's are equal, all r_O 's are equal, and neglecting g_{mb} , Equation 27 reduces to $A_v \approx (2/3)g_m^2 r_O^2$. Since the intrinsic gain $g_m r_O = \sqrt{2k(W/L)I_D}(1/\lambda I_D)$ increases with decreasing I_D ($\lambda \propto I_D$), the gain of the opamp increases as the IR drop increases. Simulation of the opamp gain versus IR drop is shown in Figure 54:



Figure 54: Opamp gain versus supply IR drop.

In the old design the large IR drop causes the opamp gain to increase, peaking at around 220 mV with a gain over 8,500 V/V, then dropping rapidly as the current sources enter into the cutoff region, in which case the opamp fails to function. With the new power distribution design, the gain varies between 1,360 V/V near supply points and 1,740 V/V at the maximum IR drop point (28% gain variation).

The IR drop also affects the bandwidth of the opamp. Considering only the dominant pole, which is located at the output node, the magnitude of its frequency is

$$\omega_{1} \approx \frac{1}{\left[g_{m8}r_{O8}r_{O10} \parallel g_{m4}r_{O4}(r_{O6} \parallel r_{O2})\right]C_{L}}$$
(28)

where C_L is the total load capacitance of the output node. If g_m 's are assumed to be equal and r_O 's are assumed to be equal, the equation reduces to $\omega_1^{-1} \approx (1/3)g_m r_O^2 C_L =$ $(1/3)\sqrt{2k(W/L)I_D}(1/\lambda I_D)^2 C_L$. Therefore, increasing the IR drop decreases I_D and decreases the pole frequency. Simulation of the opamp's unity-gain bandwidth versus IR drop is shown in Figure 55:



Figure 55: Opamp unity-gain bandwidth versus IR drop.

In the old power distribution design the large IR drop causes the bandwidth to decreases from a maximum of 856 MHz, bottoming out at 868 kHz when the IR drop reaches 230 mV. The new power distribution sees less bandwidth variation, with a maximum bandwidth of 856 MHz near supply points and a minimum of 764 MHz between two supply points (11% bandwidth variation). Due to the variation in bandwidth, the gain stage must wait for the slowest opamp to settle to within a required precision. For the opamp used in the ADC, the bandwidth and gain variation introduces FPN, which can only be corrected by software in post-processing.

4.6 Pixel Power Distribution

The pixel contains a reset transistor, a source follower input transistor, and a row select transistor in Active Pixel Sensor (APS) configuration. The schematic and layout of a pixel are shown in Figure 56 and Figure 57 respectively:







Figure 57: Pixel layout.

The 4,140 \times 3,865 pixel array uses grid-like power distribution network for the pixels. Each source follower draws a static current of 10 μ A when selected. Since only one row of pixels is selected at any time, the pixel array consumes a total of 4,140 \times 10 μ = 41.4 mA of static current. The block diagram of the power-distribution grid is shown in Figure 58:



Figure 58: Block diagram of pixel power distribution design with added power supplies.

The pixel array uses eight power supply points on the left and right sides (the top and bottom sides are used for readout circuitry). The power grid is shown in the expanded view in Figure 58. The resisters represent the parasitic resistance of the interconnect, and the row of constant current sources represent the selected pixel row. To simulate the IR drop, nodal analysis is used [42]. The problem is formulated into a linear system as:

$$\mathbf{G}\mathbf{v} = \mathbf{i}$$
 (29)

Where **G** is the conductance matrix for the resister network, **v** is a vector of node voltages, and **i** is a vector of current sources. Using MatlabTM, the node voltages are solved, and the results are shown in Figure 59:



Figure 59: Vdd IR drop in the pixel array with additional power supplies.

Figure 59 shows a 2D view of the pixel array with the column numbers labeled on the xaxis and the row numbers labeled on the y-axis. The shades of gray represent varying levels of IR drop, with the voltage index on the right side. The row being read is arbitrarily selected to be row #1,932 (middle row). From the figure, the greatest IR drop for this case occurs at the center of the array, with a value of 8 mV. For comparison, the power distribution design from EM5 is directly used (only four supply points). The block diagram of this design is shown in Figure 39, and simulation results shown in Figure 61:



Figure 60: Block diagram of pixel power distribution design without added power supplies.



Figure 61: Vdd IR drop in the pixel array without additional power supplies.

Without the added supply points, the maximum IR drop is 20 mV at the center of the array.

4.7 Effect of IR Drop on Pixel Performance

The drop in supply voltage has a direct impact on the pixel reset level. With an NMOS reset transistor, the reset level of the diode is $V_{reset} = V_{DD} - V_{th}$. Therefore, a reduction by V_x in the rail voltage will reduce V_{reset} by the same amount. The inconsistent V_{reset} across the same row increases the FPN. Also, the lower reset level reduces the total number of electrons that can be collected in the diode. The FPN can be corrected by CDS, while the reduced electron collection capacity remains. A solution is to de-select the pixel before resetting, eliminating the static current consumption and therefore the IR drop during reset.

4.8 Summary

Comparison of the performances using the old and new power distribution designs are shown in Table 2:

	Old design	New design
Pixel current source ground IR drop (mV)	< 80.7	< 2.58
Pixel source follower gain (V/V)	0.826 - 0.843	0.826 - 0.827
Pixel source follower -3dB bandwidth (MHz)	6.92 - 14.6	14.4 - 14.6
Opamp ground IR drop (mV)	< 236	< 20.2
Opamp Vdd IR drop (mV)	< 90.7	< 14.9
Opamp gain (V/V)	1,360 - 8,500	1,360 - 1,740
Opamp unity-gain bandwidth (MHz)	0.868 - 856	764 - 856
Pixel V _{DD} IR drop (mV)	< 20	< 8

Table 2: Performance comparison using the old and new power distribution designs.

Chapter 5: Conclusion

This thesis discussed the clock distribution and power distribution issues and solutions for building a next-generation electron microscopy CMOS image sensor. The new sensor, named EM7, is designed in standard 0.25 μ m technology and measures 21 \times 21 mm². It contains 16 million pixels, over 4,100 parallel analog-processing circuits, over 4,100 ADCs, over 4,100 readout registers, and over 3,800 row address registers. Simulations of the chip demonstrate improved performance for clock distribution and power distribution against using the designs from the previous generation image sensor: With local twophase clock generation, the column shift-register clock delay is reduced from 4.7 ns to 0.14 ns. A clock tree is used to minimize the row shift-register clock skew and reduces the delay from 1.7 ns to 0.12 ns. With local buffering, the ADC Gray code delay is reduced from 35 ns to 0.9 ns. The frame rate is improved from 43 frames/s to 75 frames/s. Robust power/ground pad placement and routing reduce the pixel current source IR drop from 80.7 mV to 2.58 mV. The worst-case pixel source follower bandwidth is increased from 6.92 MHz to 14.4 MHz. The gain-stage opamp and comparator ground IR drop is reduced from 236 mV to 20.2 mV, V_{DD} IR drop from 90.7 mV to 14.9 mV, and the gain variation is reduced from 525% to 28%. The worst-case opamp bandwidth is increased from 868 kHz to 764 MHz. The pixel IR drop is reduced from 20 mV to 8 mV.

REFERENCES

- A. Milazzo, P. Leblanc, F. Duttweiler, L. Jin, J. Bouwer, S. Peltier, M. Ellisman, F. Bieser, H. Matis, H. Wieman, P. Denes, S. Kleinfelder, N. Xuong, "Active Pixel Sensor Array As a Detector for Electron Microscopy," *Ultramicroscopy*, Sep. 2005, vol. 104, pp. 152-159.
- [2] G. Deptuch, A. Besson, P. Rehak, M. Szelezniak, J. Wall, M. Winter, Y. Zhu, "Direct Electron Imaging in Electron Microscopy with Monolithic Active Pixel Sensors," *Ultramicroscopy*, Sep. 2007, vol. 107, pp. 674-684.
- [3] W. Baumeister, "Electron Tomography: Towards Visualizing the Molecular Organization of the Cytoplasm," *Biophysical Methods*, Oct 2002, vol. 12, pp. 679-684.
- [4] A. Fauqi, S. Subramaniam, "CCD Detectors in High-Resolution Biological Electron Microscopy," *Quarterly Reviews of Biophysics*, 2000, vol. 33, pp. 1-27.
- [5] P. Roberts, J. Chapman, A. MacLeod, "A CCD-based Image Recording System for the CTEM," *Ultramicroscopy*, 1982, vol. 8, pp. 385-396.
- [6] R. Meyer, A. Kirland, R. Dunin-Borkowski, J. Hutchison, "Experimental Characterisation of CCD Cameras for HREM at 300 kV," *Ultramicroscopy*, Aug. 2000, vol. 85, pp. 9-13.
- [7] E. Fossom, "CMOS Image Sensors: Electronic Camera on a Chip," *IEEE Trans. Electron Devices*, Oct. 1997, vol. 44, pp. 1689-1698.
- [8] M. Bigas, E. Cabruja, J. Forest, J. Salvi, "Review of CMOS Image Sensors," *Microelectronics J.*, 2006, vol. 37, pp. 433-451.
- [9] A. Krymski, N. Bock, N. Tu, D. Van Blerkom, and E. Fossum, "A High-Speed, 240-frames/s, 4.1-Mpixel CMOS Sensor," *IEEE Trans. Electron Devices*, Jan. 2003, vol. 50, pp. 130-135.
- [10] S. Yoshihara, Y. Nitta, M. Kikuchi, K. Koseki, Y. Ito, Y. Inada, S. Kuramochi, H. Wakabayashi, M. Okano, H. Kuriyama, J. Inutsuka, A. Tajima, T. Nakajima, Y. Kudoh, F. Koga, Y. Kasagi, S. Watanabe, and T. Nomoto, "A 1/1.8-inch 6.4 Mpixel 60 frames/s CMOS Image Sensor with Seamless Mode Change," *IEEE J. Solid-State Circuits*, Dec. 2006, vol. 41, no. 12, pp. 2998-3006.
- [11] I. Takayanagi, M. Shirakawa, K. Mitani, M. Sugawara, S. Iversen, J. Moholt, J. Nakamura, and E. Fossum, "A 1.25-inch 60-frames/s 8.3-M-pixel Digital-Output CMOS Image Sensor," *IEEE J. Solid-State Circuits*, Nov. 2005, vol. 40, no. 11, pp. 2305-2314.
- [12] G. Meynants, D. Scheffer, B. Dierickx, and A. Alaerts, "A 14-Megapixel 36 x 24mm² Image Sensor," *Proc. SPIE*, Jul. 2004, vol. 5301, pp. 168-174.
- [13] S. Ay, E. Fossum, "A 76 x 77 mm², 16.85 Million Pixel CMOS APS Image Sensor," 2006 Sym. VLSI Circuits Digest Technical Papers, Nov. 2006, pp. 19-20.
- [14] M. Iwane, T. Matsuda, T. Sugai, K. Tazoe, T. Okagawa, T. Ono, T. Watanabe, K. Ogawa, H. Takahashi and S. Inoe, "52 Mega-pixel APS-H-size CMOS Image

Sensor for Super High Resolution Image Capturing," Proc. 2007 In'l Image Sensor Work Shop, Jun. 2007, pp.295-298.

- [15] N. Xuong, A. Milazzo, M. Ellisman, S. Peltier, J. Bouwer, F. Duttweiler, P. Leblanc, J. Matteson, H. Wieman, H. Matis, F. Bieser, P. Denes, S. Kleinfelder, "First Use of a High-Sensitivity Active Pixel Sensor Array as a Detector for Electron Microscopy," *Proc. SPIE, Electronic Imaging Science and Technology*, Vol. 5301, Jan. 2004, pp 242-249.
- [16] S. Li, J. Bouwer, F. Duttweiler, M. Ellisman, L. Jin, P. Leblanc, A. Milazzo, S. Peltier, N. Xuong, S. Kleinfelder, "A New Direct Detection Camera System for Electron Microscopy," *Proc. SPIE, Electronic Imaging Science and Technology*, 2006, vol. 6068, 60680O.
- [17] S. Li, S. Kleinfelder, L. Jin, N. H. Xuong, "A CMOS Sensor for Nano-Imaging," Proc. IEEE Conf. on Nanotechnology, Jul. 2006, vol. 2, pp. 544- 547.
- [18] L. Jin, A. Milazzo, S. Kleinfelder, S. Li, P. Leblanc, F. Duttweiler, J. C. Bouwer, S. T. Peltier, M. Ellisman, N. Xuong, "The Intermediate Size Direct Detection Detector for Electron Microscopy," 2007, *Proc. SPIE*, Vol. 6501, 65010A.
- [19] S. Li, S. Kleinfelder, "Direct Charged-Particle Imaging Sensors," *Nuclear Instruments and Methods in Physics Research A*, Aug. 2007, vol. 579, no. 1, pp. 227-230.
- [20] S. Kleinfelder, S. Li, Y. Chen, "Optimization of Monolithic Charged-Particle Sensor Arrays," *Nuclear Instruments and Methods in Physics Research A*, Sept. 2007, vol. 579, no. 2, pp. 695-700.
- [21] S. Li, "Modeling, Design, and Analysis of Monolithic Charged-particle Image Sensors," doctoral dissertation, Dept. Electrical and Computer Engineering, Univ. California, Irvine, 2007.
- [22] L. Jin, A. Milazzo, S. Kleinfelder, S. Li, P. Leblanc, F. Duttwiler, J. Bouwer, S. Peltier, M. Ellisman, N. Xuong, "Applications of Direct Detection Device in Transmission Electron Microscopy," *J. Structural Biology*, 2008, vol. 161, pp. 352-358.
- [23] N. Xuong, L. Jin, S. Kleinfelder, S. Li, P. Leblanc, F. Duttwiler, J. Bouwer, S. Peltier, A. Milazzo, M. Ellisman, "Future Directions for Camera Systems in Electron Microscopy," *Methods in Cell Biology*, 2007, vol. 79, pp. 721-739.
- [24] M. Jackson, A. Srinivasan, E. Kuh, "Clock Routing for High-Performance ICs," *Proc.* 27th ACM/IEEE Design Automation Conf., IEEE, 1990, pp. 573-579.
- [25] B. Gieseke *et al.*, "A 600 MHz Superscalar RISC Microprocessor with Out-of-Order Execution," *ISSCC Tech. Dig.*, Feb. 1997, pp. 176-177.
- [26] C. Webb et al., "A 400-MHz S/390 Microprocessor," *IEEE J. Solid-State Circuits*, vol. 32, Nov. 1997, pp. 1665-1657.
- [27] G. Geannopoulos, X. Dai, "An Adaptive Digital Deskewing Circuit for Clock Distribution Networks," *IEEE SSCC Tech. Dig.*, Feb. 1998, pp. 400-401.
- [28] P. Restle, A. Deutsch, "Designing the Best Clock Distribution Network," *Symp. VLSI Circuits Dig. Tech. Papers*, 1998, pp. 2-5.
- [29] MOSIS, http://www.mosis.com/cgi-bin/cgiwrap/umosis/swp/params/tsmc-025/t4cs_lo_epi-params.txt

- [30] N. Arora, K. Raol, R. Schumann, L. Richardson, "Modeling and Extraction of Interconnect Capacitances for Multilayer VLSI Circuits," *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, vol. 15, Jan. 1996, pp. 58-67.
- [31] J. Rabaey, A. Chandrakasan, B. Nikolic, *Digital Integrated Circuits*, Prentice Hall, 2003.
- [32] J. Gu, Z. Wang, X. Hong, "Hierarchical Computation of 3D Interconnect Capacitance Using Direct Boundary Element Method," *Proc. ASP-DAC Design Automation Conf.*, 2000, pp. 447-452.
- [33] J. Xu, H. Li, W. Yin, J. Mao, L. Li, "Capacitance Extraction of Three-Dimensional Interconnects Using Element-by-Element Finite Element Method (EBE-FEM) and Preconditioned Conjugate Gradient (PCG) Technique," *IEICE Trans. Electronics*, 2007, pp. 179-188.
- [34] A. Kahng, S. Muddu, "Delay Analysis of VLSI Interconnections Using the Diffusion Equation Model," 32st Conf. Design Automation, Jun. 1994, pp. 563-569.
- [35] W. Elmore, "The Transient Response of Damped Linear Networks with Particular Regard to Wideband Amplifiers," J. Applied Physics, vol. 19, Jan. 1948, pp. 55-63.
- [36] Y. Suzuki, K. Odagawa, T. Abe, "Clocked CMOS Calculator Circuitry," *IEEE J. Solid-State Circuits*, vol. 8, Dec. 1973, pp. 462-469.
- [37] J. Neves, E. Friedman, "Optimal Clock Skew Scheduling Tolerant to Process Variations," *Proc.* 33rd Design Automation Conf., 1996, pp. 623-628.
- [38] S. Lin, N. Chang, "Challenges in Power-Ground Integrity," Proc. 2001 IEEE/ACM Int'l Conf. Computer-Aided Design, 2001, pp. 651-654.
- [39] M. Zhao, R. Panda, S. Sapatnekar, D. Blaauw, "Hierarchical Analysis of Power Distribution Networks," *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, 2002, pp. 159-168.
- [40] Y. Zhong, M. Wong, "Fast Algorithms for IR Drop Analysis in Large Power Grid," Proc. 2005 IEEE/ACM Int'l Conf. Computer-Aided Design, 2005, pp. 351-357.
- [41] B. Razavi, Design of Analog CMOS Integrated Circuits, McGraw Hill, 2001.
- [42] C. Ho, A. Ruehli, P. Brennan, "The Modified Nodal Approach to Network Analysis," *IEEE Trans. Circuits and Systems*, vol. 22, Jun. 1975, pp. 504-509.