Abstract—Visible light from silicon junctions under avalanche breakdown can be used to create microdisplay systems with integrated light sources. Junctions available in standard CMOS usually break down at much larger voltages than the typical operating voltage for integrated circuitry. It is possible to reduce the operating voltage of by making use of techniques which changes the electric field profile in light sources based on hot carrier electroluminescence such as electric field reach through between two highly doped implant regions. This work successfully demonstrates the possibility of tailoring the operating voltage and quantifying the optical performance in an integrated microdisplay consisting of a 128 by 96 pixel array based on light sources in standard CMOS. Based on the approach followed it becomes possible to integrate light sources in such a manner that it can coexist and interact with other on-chip analog and digital circuitry. The requirements for architectural features of a microdisplay in standard CMOS is discussed and it is shown to be possible to create large scale integrated circuits containing integrated light sources in standard CMOS without the need for postprocessing or additional back end modifications.

Index Terms—Avalanche, CMOS, electroluminescence, microdisplay, silicon.

I. INTRODUCTION

Silicon p-n-junctions under avalanche breakdown have been known to emit visible light since the first published observation by Newman [1] in 1955. It is possible to make use of this phenomenon in CMOS integrated circuits, thereby providing opportunity for creating integrated light sources [2]. Emission occurs due to hot carrier interactions under high electric fields [3] and provides a broad spectrum [4] which sufficiently covers the photopic and scotopic spectral ranges resulting in light observable by the human eye. This feature of the emission spectrum provides for the opportunity to utilize hot carrier light sources to create microdisplays in standard CMOS. Since avalanche breakdown occurs at much higher voltages than the typical operating voltage of conventional CMOS circuitry, integration of light sources based on avalanching junctions presents a challenge, especially when required to coexist and interact with other on-chip circuitry in an integrated system.

Microdisplays find application in near-to-eye (NTE) systems and projection displays [5] and are distinguished from other miniature displays by the fact that magnification is required for observing text or images formed on readout. An increase in recent commercial products in the form of wearable glasses hints at an active and developing wearable display market of which microdisplays form a fundamental part of the system with information and content conveyed to the user through an NTE interface. In these systems the size and form factor, resolution, luminance, weight, power consumption and functionality become key performance criteria for the microdisplay. Microdisplays are generally based on either organic light emitting diodes (OLEDs) [6], which are emissive in nature, or on liquid crystal on silicon (LCoS) technology which relies on transmissive principles.

The principle of using silicon light sources in microdisplay applications has been demonstrated by using porous silicon [7]–[9] or avalanche electroluminescence in addition to image intensification [10]. We have previously presented another approach where hot carrier electroluminescence is used to create a microdisplay with devices directly available in CMOS [11] without reverting to techniques incompatible with CMOS, image intensification or additional post processing. By constraining the light sources to be completely manufacturable in a standard CMOS and utilizing techniques to interact with on-chip digital circuitry and drivers, it was proven possible to create a CMOS microdisplay which is visible to the naked eye and fabricated as if it was an ordinary ASIC.

The problem with our prior work was the high operating voltage of above 12 V. This work addresses the voltage restrictions by utilizing a newly developed hot carrier light source in standard CMOS. The improved light source operates at a lower voltage and produces substantially more optical power in the photopic region. These devices were implemented in a microdisplay boasting higher resolution as well as integrated intensity control. The previous results [11] for a 64 × 8 pixel microdisplay are compared to the performance of the 128 × 96 pixel microdisplay presented in this work and the results are discussed.

This work was supported by INSiAVA (Pty) Ltd. at the Carl and Emily Fuchs Institute for Microelectronics, University of Pretoria, Pretoria, South Africa.

The authors are with the Carl and Emily Fuchs Institute for Microelectronics, Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria, South Africa (e-mail: pjventer@ieee.org; jannes.venter@up.ac.za; monuko@up.ac.za).
II. LIGHT SOURCE

A. Principle of Operation

It is known that a p-n-junction operating under avalanche breakdown emits light. Our prior work [11] makes use of this phenomenon in a standard CMOS process to create light sources using silicon as the active material. Fig. 1 shows a plan view of a junction created in a CMOS design. With a p-type background region doped in the order of $10^{17}$ cm$^{-3}$ a junction formed by the addition of an n$^+$-implant region will break down between 9 V and 12 V. A p$^+$-implant is used to provide a low impedance substrate tap. In order to form a single localized light spot, the structure terminating the n$^+$-substrate tap acts as a avalanche junction such that the electric field magnitude due to an increase in an n$^+$-implant region will break down between 9 V and 12 V. A p$^+$-implant is used to provide a low impedance substrate tap. In order to form a single localized light spot, the structure terminating the n$^+$-substrate tap acts as a avalanche junction such that the electric field magnitude resulting from edge effects [12] due to the approximate spherical profile of the depletion region around the point. This localization assists in improving the light spot intensity [13] and makes it possible to improve uniformity by creating matrices of light spots. Both n$^+$-p- and p$^+$-n-junctions can be formed by using an n-$^+$-well in the latter case. Optical power is a linear function of device current as shown in Fig. 3 for the light source design used in the microdisplay. No appreciable difference is observed when comparing the emission from an n$^+$-p-junction to that of a p$^+$-n-junction.

B. Operating Voltage and Reach Through Conditions

A high operating voltage means that any driving circuitry will be subjected to large voltage variations. For example, where a transistor is used to switch a light source on or off, the drain would experience voltages from $V_{on}$, the driving NMOS transistor’s “on” voltage in the linear operating region, to $V_{supply}$ which is the voltage level of the supplying rail powering the light sources. $V_{supply}$ is of sufficient magnitude to support both avalanche breakdown and the transistor overheads.

The technique used to lower the operating voltage is based on spacing the highly doped substrate tap sufficiently close to the avalanching junction such that the electric field reaches through and the substrate tap acts as a field stop. It is possible to lower the breakdown voltage of the light sources by increasing the background doping concentration. In a CMOS process, this approach lies beyond a circuit designer’s influence.

We propose an alternative method. Assume $W_D$ is the depletion region width which, according to an abrupt junction approximation, extends mainly in the lightly doped background region. As the reverse bias voltage increases, so does $W_D$, save for the compressive variation close to the pointed edge of the light source due to the edge effects mentioned in Section II-A.

At breakdown, an incremental increase in voltage results in a large increase in device current, effectively pinning the voltage to just above the onset of breakdown. The key to lowering the voltage lies in ramping up the electrical field to the critical value $E_{crit}$ required for avalanche multiplication, while reducing the potential integral. This can be done using a highly doped field stop region of same doping kind as the background. Position this region at a separation distance closer than $W_D$ laterally near the avalanching junction, the potential integral can be reduced without reducing the maximum electric field.

This effect can be understood by examining Fig. 3 where $\Delta x$ is the separation distance between the two highly doped regions of opposite kinds and $W_D$ is the depletion region width extending laterally into the background doped region in the absence of a field stop region. Although there is room for further depletion elsewhere around the junction, it is the maximum electric field which pins the voltage at the onset of avalanche breakdown near the pointed source and closely spaced field stop region as shown in Fig. 2(a), while the voltage is determined by the potential integral in the direction of the depletion region as shown in Fig. 2(b).

This “reaching through” of the electric field to the field stop region results in a lower operating voltage for the device for the same series current flow in the device, under the assumption that only the electric field region reaching the critical field will be responsible for generating avalanche current. Using the reach through technique, it is possible to tailor the operating voltage for the light sources used in this work as a function of separation distance $\Delta x$. Another advantageous effect when the junction is operating in reach through is that the breakdown region is further localized. Since the potential across the junction...
needs to remain constant, the depletion region terminated by a field stop will reach a higher peak electrical field magnitude, the limits thereof being $E_{\text{crit}}$, where avalanche breakdown occurs. The light spot will manifest in the lightly doped region where the $n^+$- and $p^+$-implants are in closest proximity to each other.

**C. Polarity Considerations**

A single well CMOS process was used to fabricate the 128×96 pixel microdisplay, where the p-type substrate and p-wells share a conductive body and an n-well implant is used for creating PMOS devices. Since n$^+$- and p$^+$-implants are available, there are generally three available junctions, n$^+$-p, p$^+$-n and n-well to p-well. However, all junctions of the n$^+$-p-type share a common anode in the substrate, while it is possible, thanks to the n-well, to create p$^+$n junctions as a floating circuit component. The latter is the junction of choice for light sources on integration with other circuit components as the polarity allows for direct connection between the p-n junction’s anode and an NMOS transistor’s drain. It is necessary to consider whether the characteristics of both polarities are substantially similar. Fig. 3 shows a comparison between the light sources used in the microdisplay of both polarity kinds. The average difference between light sources across multiple samples is less than 1% over a large operating range. This feature allows for interchangeable use between devices of different polarities, depending on the application, without affecting the electro-optical conversion efficiency.

**D. Emission Spectra in Reach Through**

Microdisplays provide images discernable to the human eye. It is therefore necessary that a light source used in a microdisplay exhibits sufficient spectral content in the visible region. The luminosity function provides a photopic response curve [14], [15] which can be used to quantify the visible light by weighting the spectrum of the light source. Fig. 4 shows a power corrected spectrum of the light source when operated in an arrayed fashion at a device current of $I_D = 100$ mA. The ratio of the separation distance between the two highly doped regions, $\Delta x$ in Fig. 1(a), and the typical depletion region width $W_D$ for a line junction as shown in Fig. 2 is varied as $\Delta x / W_D$ in Fig. 5. Based on this ratio, the value for $\Delta x$ is selected where the peak photopic power emission occurs for maximum brightness. The total integrated emission power is also shown in Fig. 5 as detected by a silicon radiometer detector.

**E. Radiation Pattern**

Light emitted by the chip surface takes shape in the form of a radiation pattern which is strongly influenced by the front end of line surface features of the light source and CMOS process in which it is fabricated. In order to quantify this, a three dimensional radiation pattern was determined using a goniometric measurement setup. This allowed for the characterization of the emission pattern and the determination of half-power angles which are useful to determine luminance values. The display described in this work makes use of a reach through light source without light directing structures [16] in the BEOL-stack.
in an attempt to improve the array density individual light source points and to reduce the pixel pitch by reducing the area of the emissive part of the pixel. Fig. 6(a) shows the three dimensional radiation pattern for the modified light sources and Fig. 6(b) shows a cross sectional cut through the 0°–180° and 90°–270° axes since the beam is not symmetrical. The peak power point for each cut is measured after which the average half power angles can be derived. An average half power apex angle is subsequently calculated for the entire emission beam and approximated to $2\theta = 10^\circ$. An assumption is made that the spectral nature of emission does not vary significantly over the beam of emission.

### III. INTEGRATED PIXEL

Each pixel consists of two NMOS devices in series driving an array of 63 individual junctions and connects to row and column address lines, denoted as $R_{\text{row}}$ and $C_{\text{col}}$ respectively in Fig. 7(a), while the cathode of the light source connects to a high voltage supply line. Both NMOS transistors share a common substrate (bulk terminal) since the implementation technology is of single well type. The body effect has been taken into consideration upon sizing the devices. On applying a logic “1” to both address lines, the light source anode is drawn to a low potential and avalanche breakdown occurs establishing the current necessary for electroluminescence. Fig. 7(a) shows an example of a simplified circuit schematic of a pixel, while Fig. 7(b) shows the layout of a pixel with the emission area and driver identified. Since CMOS is planar in nature, the fill factor suffers from having the driver integrated into the pixel and because it is necessary to reserve sufficient real estate for providing the necessary metal layers for the address lines and power supply. The pixel pitch is 25 $\mu$m. This pixel configuration allows having both the column and row addressing lines integrated in a pixel and supplying a continuous high voltage, necessary for breakdown, to all of the pixels. This approach enables interfacing circuitry to directly interact with the pixels at standard voltage levels without the need to manipulate high voltage lines feeding pixel lines or rows individually. It then becomes possible to create a digital interface completely compatible with the rest of the on-chip electronics.

### IV. ARCHITECTURE

#### A. Overview

The emissive area dimensions for the 128 $\times$ 96 display are 3.2 $\times$ 2.4 mm$^2$ which necessitate the use of additional optics in order to perceive an image. As with OLEDs the contrast ratio in the absence of ambient light is extremely high since the display is emissive in nature and pixels can be completely turned off. The pixels are arranged as an array of 128 columns and 96 rows with each row and column sharing address lines $R_{\text{row}}$ and $C_{\text{col}}$ respectively. Each address line shared by a row is connected to the output of a pulse width modulator (PWM) while the column address lines are directly connected to a multiplexer. The whole array is connected with a mesh of the high voltage and ground supplies provided by metal interconnect layers. An integrated controller handles the interface between I/O and the row and column multiplexers. An overview of the system architecture is shown in Fig. 8.

#### B. Addressing Scheme

Hot carrier luminescence seen from avalanching junctions is known to be extremely fast [17], [18] which means that the limit to the rate at which the pixels can be toggled is limited by other factors such as the electronics and architecture itself. Since the planarity of CMOS burdens the pixel designer to aim for the simplest and smallest driver circuit in order to maximize the fill
factor, no memory element is integrated into the pixel. Instead, a persistent image is created by a passive matrix addressing scheme by scanning the array sufficiently fast at a refresh rate of above 60 Hz in order to avoid the effects of flicker [19]. Unlike passive display matrices, each pixel enjoys its own integrated driver. In order to relax the instantaneous local current density requirements on the power supply lines the display is energized one column at a time, while all rows are energized in parallel. This approach results in 96 pixels simultaneously scanning the display as a column “line” sweeping the display area with each row modulating its associated pixel based on the content of an external frame buffer while being synchronized to the column positioning counter in order to form an image. A trade-off exists in that each of the rows, when driven in parallel, requires one PWM per row which tends to complicate the circuitry.

C. Intensity Control

The row-parallel addressing scheme makes use of the effects of Bloch’s law on vision [20] in order to establish gray scale capability. Bloch’s law states that the perceived luminance is a function of energy and brightness and is dependent on the product of time and luminance [21]. The theory only holds up to a critical duration where, if the duration of the stimulus is longer, no increase in luminance is perceived. If temporal modulation of the light sources is well below 100 ms, Bloch’s law should hold [22] and the perceived brightness can be modulated by temporal modulation of the pulse energizing the pixel. With a refresh rate of 60 Hz covering 128 columns during each refresh cycle, the “on” duration of any pixel will be limited to 130.2 μs which is well below the critical duration and within the valid region of Bloch’s law.

A single complete scan of the emissive array takes \( T_{\text{refresh}} = 1/60 \) for a 60 Hz refresh rate. Since the 128 columns are energized sequentially, each column enjoys an addressed time window of \( T_{\text{col}} = T_{\text{refresh}}/128 \) wherein the light source can be energized. Each pixel in a column is on for a maximum of \( T_{\text{pixel max}} = T_{\text{col}}/64 \). One PWM exists for each row, thereby modulating the duty cycle per row for each pixel in a column. These 96 parallel PWMs uses pulse width modulation to establish gray scale operation where each period is further subdivided into 256 time quantized divisions. Based on a loaded 8-bit gray scale value the duty cycle can be varied from 0 to 255 depending on the input to the PWM for a specific row. This approach is preferred over an analogue approach in order to shift as much function to the digital controller for repeatable behavior and the high intrinsic bandwidth of the light sources presents no challenge to support this scheme for even higher resolution PWMs.

V. RESULTS AND DISCUSSION

A. Electrical Performance

Digital circuitry in the 0.35 μm CMOS process used for implementing the 128 x 96 pixel microdisplay operate at typical voltages of 3.3 V. Avalanche breakdown typically occurs in junctions in the region of 9 V and above. As described in Section II-B, electric field reach through can be used to lower the breakdown voltage of junctions as a function of separation distance. Fig. 9 shows the average measured breakdown voltage \( V_{BD} \) of the light sources as a function of separation distance \( \Delta x \) in units of the ratio \( \Delta x/W_D \) for both single junctions and junctions placed and driven in arrays. Over a certain domain of \( \Delta x/W_D \) the breakdown voltage \( V_{BD} \) exhibits a linear relationship to the separation distance \( \Delta x \). This is to be expected under the reach through conditions as the value of the potential integral is directly proportional to the depletion region width. The width in turn is determined by the separation distance \( \Delta x \). The upper limit of the linear region is reached when the separation distance falls beyond the reach through condition, while the lower limit is determined by the onset of Zener-mode tunnelling. The measurement results were obtained from 8 samples in two different manufactured batches. The error bars shows the spread across all samples, indicating a reasonable accurate predictive result even across different lots. Although the ratio \( \Delta x/W_D \) over which the devices exhibit a linear dependency on \( \Delta x \) can be smaller that the critical process dimensions, we have developed a method to allow for a reasonably repeatable reduction in \( \Delta x \) which is not directly limited to the lithographic limits of a standard CMOS process.

The choice of separation distance \( \Delta x \) is determined by the point where the maximum photopic power emission is observed which should result in the brightest point for display purposes. The resulting light source configuration resulted in a breakdown voltage of \( V_{BD} = 6.63 \) V at an individual junction current of
\[ I_D = 100 \, \mu A \] Electrical simulations of the driving circuitry and the characterized light sources are shown in Fig. 10 along with the measured results, which correlates reasonably well within the design boundaries where a limit was set for the maximum electrical power consumption of \( P_e \leq 1.4 \, W \). It should be restated that the results shown are for the emissive matrix including the part of the pixel consisting of driving circuitry, but excludes the power consumption for the rest of the digital circuitry on-chip such as the PWMs and multiplexers.

### B. Optical Performance

In order to observe the optical power emitted by the light sources under test, a silicon radiometer detector was placed in close proximity to the surface while the radiometer detector current was recorded. This was combined with the spectral characteristics of the light source in order to quantify the optical emission in terms of radiant power. It is assumed that the proximity of the detector captures practically all of the emitted light from the chip surface. Fig. 11 shows the integrated radiant power emitted by the 128×96 pixel microdisplay as a function of input electrical power. For the measurement, all rows were fully energized in order to establish a “flashlight” mode of operation. Since the cathodes of the light sources share a common node through the high voltage supply line, this line was simply driven by a voltage sweep while monitoring the current consumed by the pixel array. Luminance is a measure of the perceived brightness often used to characterize displays and microdisplays. Based on the emitted power weighted by the luminosity function, the resulting measure of luminous flux subtending a solid angle can be used to estimate a luminance value for the microdisplay. The radiation pattern half power angle is used for the purpose of estimating a solid angle of emission, while the photopic power is converted to luminous flux by proper scaling. Fig. 12 shows the calculated luminance as a function of input electrical power for the 128×96 pixel microdisplay.

### C. Comparison of Results With Previous Work

Our prior work [11] resulted in CMOS microdisplay with a 64×8 pixel array and it is worth comparing the current microdisplay with the previous in order to highlight which aspects are important in standard CMOS design. Both light source designs used in the 128×96 and 64×8 pixel display are based on hot carrier electroluminescent light sources operating in avalanche, which allows for a reasonable comparison in performance and operating voltage. The 128×96 pixel is modified slightly to allow for an increase in extraction and power efficiency as shown in Fig. 11. The optical power leaving the chip surface is almost an order of magnitude larger for the 128×96 pixel display. Furthermore, the voltage of the reach through light sources is substantially lower as shown in Fig. 13 where the operating voltage of the emissive array is compared. The combined result
is that the 128×96 pixel display in this work is an order of magnitude more efficient than the 64×8 pixel display from our prior work. Improvement factors for the total and photopic output power of 14 and 21, respectively are observed at an input power of 1 W for the 128×96 pixel over the 64×8 pixel display. There are however some important differences which make a direct comparison between the two works difficult: 1) the 64×8 pixel display utilizes BEOL light directing structures [16] while the 128×96 pixel display contains no light directing structures; 2) the reach through light sources differ fundamentally in design and 3) the 64×8 pixel microdisplay required direct high voltage signals for addressing columns, necessitating the use of external driver circuitry. Table I summarizes the different optical parameters at an input electrical power of 1 W with \( \Phi_e \) the radiant flux emitted from the chip surface, \( 2\theta \) the apex angle, \( \Omega \) the solid angle in steradians, \( \Phi_v \) the luminous flux as determined by the spectrum weighted by the luminosity function, \( A \) the emissive area of the microdisplay and \( L_v \) the luminance.

From Fig. 12 and Table I it is clear that the new light source design outperforms the previous design by more than an order of magnitude. The two key parameters which result in lower luminance for the 128×96 display than anticipated are the increased beam angle and the larger display area, both of which reduce the luminance value. We anticipate that the luminance can be improved through the use of BEOL light directing structures in the current design as well.

### VI. CONCLUSION

Hot carrier electroluminescence in avalanching CMOS pn-junctions affords CMOS technologies the opportunity to directly fabricate microdisplays. This work described the first completely integrated microdisplay where fully on-chip digital addressing, intensity control and what can be considered large scale integration along with integrated light sources are implemented on the same die. Furthermore, it was shown that the operating voltage of the light sources can be adjusted within certain limits as required by the specific application. Although our prior work exhibits higher luminance, the 128×96 pixel microdisplay emits more power and the reason for the drop in luminance is ascribable to the increased emissive area and substantially larger solid angle of the light sources used. An image of a logo is displayed in Fig. 14 on the 128×96 pixel microdisplay presented in this paper. Still images and video become clearly visible once the ambient light levels drop and the appropriate optical stack is introduced in front of the microdisplay.

### ACKNOWLEDGMENT

The authors would like to thank T. Joubert for her input in the digital design, and A. Alberts, M. Goosen, N. Fauré, and C. Janse van Rensburg for measurements and testing. The authors would also like to thank T. Joubert, Carl and Emily Fuchs Institute for Microelectronics, University of Pretoria, for her input in the digital design.

### REFERENCES


Monuko du Plessis (M’76–SM’81) was born in King Williamstown, South Africa, in 1949. He received the B.Eng., M.Eng., and D.Eng. degrees in electrical engineering from the University of Pretoria, Pretoria, South Africa, in 1972, 1978 and 1984, respectively. He also received the BA (Psychology) and BCom (Hons) (Economics) degrees from the University of South Africa in 1989 and 1998 respectively.

He joined the Civil Aviation Division in the Department of Transport, South Africa, in January 1973, as Telecommunications Engineer. In November 1973, he joined the Department of Electrical Engineering, University of Pretoria, as Research Officer in the field of microelectronics and optoelectronics. He was the Director of the Carl and Emily Fuchs Institute for Microelectronics (CEFIM) at the University of Pretoria from 1991 to 2012, and is currently a Full Professor in the Department of Electrical, Electronic and Computer Engineering at the University of Pretoria. His current research interests are silicon photonics and analog CMOS integrated circuit design.

Prof. du Plessis is a member of SPIE, and a Registered Professional Engineer with the Engineering Council of South Africa (ECSA).

Petrus J. Venter (M’04–SM’13) received the B.Eng. degree in electronic engineering, and the M.Eng. and Ph.D. degrees from the University of Pretoria, South Africa, in 2004, 2009, and 2013, respectively.

From 2005 to 2006, he was an assistant lecturer at the University of Pretoria. In 2007, he was appointed lecturer and started to practice engineering in a start-up company called INSiAV, founded in the same year. He currently holds an appointment both as senior lecturer and research engineer at the University of Pretoria. His fields of interests are CMOS integrated circuits, light emission from indirect band gap semiconductors and microdisplay systems.

Dr. Venter received the SAMES award for best microelectronic student in 2004 and received a first prize in the Design Automation Conference’s student paper contest in 2005. In 2006 he was awarded the Tanner Award for Outstanding Achievement in Educational Design. He is a Registered Professional Engineer at the Engineering Council of South Africa (ECSA).