Dynamic Analysis of Tester Operated Integrated Circuits Stimulated by Infra-Red Lasers

vorgelegt von
M. Sc.
Tuba KIYAN
aus Istanbul

Von der Fakultät IV -Elektrotechnik und Informatik
Institut für Hochfrequenz- und Halbleitersystemtechnologie
der Technischen Universität Berlin
zur Erlangung des akademischen Grades
Doktorin der Ingenieurwissenschaften
Dr.-Ing.

genehmigte Dissertation

Promotionsausschuss:
Vorsitzender: Prof. Dr.-Ing. R. Thewes
Berichter: Prof. Dr.-Ing. C. Boit
Berichter: Prof. Dr.-Ing. A. Ataman

Tag der wissenschaftlichen Aussprache: 23.03.2010

Berlin 2010

D 83

Erklärung

Ich versichere an Eides statt, dass ich die Dissertation selbstständig verfasst habe. Die benutzten Hilfsmittel und Quellen sind in der Arbeit vollständig angegeben.

Zusammenfassung

Laser Stimulationstechniken benutzen Laserstrahlung, um die elektrischen Eigenschaften von integrierten Schaltungen zu verändern. Die Bauelemente werden von der Rückseite oder Vorderseite Pixel für Pixel bestrahlt. Die Defekte können großere elektrische Veränderungen auslösen. Zur Lokalisierung werden die elektrischen Veränderungen pro Pixel gespeichert.

Das stetige Erhöhen der Taktfrequenz von integrierten Schaltkreisen kann Geschwindigkeitsfehler auslösen. Deswegen sind zeitliche Analysen von integrierten Schaltkreisen äußerst wichtig.

Die meisten Defekte können durch statische Laser-Stimulationsmethoden gefunden werden. Soft Defekt Lokalisierung und Performanz Charakterisierung jedoch können lediglich durch Dynamische Laser Stimulations- (DLS-) Methoden durchgeführt werden. DLS-Techniken ermöglichen eine zeitliche Charakterisierung von Bauelementen nach Signallaufzeit, Frequenz, Jitter und weiteren Eigenschaften. DLS benötigt eine volle Inbetriebnahme der Bauelemente. Zudem kann die Auswertung von Laser-induzierten Ausgangsignalen sehr kompliziert sein.

Das Hauptaugenmerk dieser Arbeit ist auf die Untersuchungen durch DLS von CMOS Bauelementen, sowie Verzögerungsketten und Rasterketten gerichtet.

In dieser Arbeit werden zwei verschiedene Laser Dioden, mit 1300 nm und 1064 nm Wellenlänge, in Dauerstich- und Pulsbetrieb verwendet. Je nach Wellenlänge ist die dominierende Auswirkung entweder thermisch oder photoelektrisch. Die verwendeten Bauelemente sind Teststrukturen von Infineon Technologies AG in 90nm Prozesstechnologie.

Um die sehr schwachen Laufzeitänderungen von Verzögerungsketten mit hoher Auflösung zu detektieren, werden verschiedene Messaufbauten vorbereitet und verglichen bezüglich Erfassungszeit und Signalqualität. Außerdem wird demonstriert, dass beim Laser-Pulsbetrieb die im Dauerstrich-Betrieb unerwünschten Nebenwirkungen vermieden werden können.

Für die zeitliche Analyse von integrierten Schaltkreisen wurde eine neue systematische Methode entwickelt. Diese benutzt das “soft fault injection” Phänomen in Scan Chain Strukturen. Der Signalzustand kann mit Laser Stimulation durch Anwendung von gepulstem Laserlicht mit einer Wellenlänge von 1064 nm auf das Empfindlichste identifiziert werden.

Der photoelektrischer Effekt ist auf Schaltkreiselevel simuliert und die Einwirkungen auf Signallaufzeit sind analysiert worden. Zudem wird der “soft fault injection“ Mechanismus in die Software- Ebene reproduziert und die zeitliche Empfindlichkeit von Signalen bei Laser Stimulation geprüft.

Abstract

Laser stimulation techniques use laser beam to change the electrical characteristics of active devices. The devices are irradiated from the backside or the frontside pixel by pixel. The defects present in the device might cause larger electrical changes. For the defect localization, the electrical changes are correlated to the laser beam position and saved for image acquisition.

High clock rates may result in speed faults in Integrated Circuits (ICs). Due to the continually increasing clock rates, timing analysis of ICs is crucial.

Most of the defects can be found by static laser stimulation techniques. On the other hand, soft defect localization and performance characterizations can only be obtained by dynamic laser stimulation (DLS) methods. DLS techniques allow timing analysis of devices, such as propagation delay, frequency, jitter etc. DLS requires full activation of the device under test (DUT). Moreover, the evaluation of the laser-induced output signals can be very hard.

The scope of this work is to investigate the CMOS devices, like delay chains and scan chains by DLS techniques.

In this work, two different laser diodes, with 1064 nm and 1300 nm laser wavelength, are utilized in continuous wave (CW) and pulsed mode. Depending on the wavelength, the dominant effect is either thermal or photoelectric. The investigated devices are produced in 90 nm process technology and provided by Infineon AG.

In order to detect very small laser-induced propagation delay variations with high resolution on the delay chains, different experimental setups are built and they are compared considering the acquisition time and signal quality. Moreover, it is demonstrated that the pulsed laser can suppress the laser induced secondary effects in opposition to the CW laser.

For the timing analysis of ICs, a new and systematic method is developed, which uses “soft fault injection” phenomenon on the scan chains. It can identify the most sensitive signal condition to the laser stimulation by using 1064 nm pulsed laser beam.

The photoelectric effect is simulated in circuit level and the effects on the propagation delay of a delay chain are analyzed. In addition to that, soft fault injection mechanism in software level reproduced and the timing sensitivity of the signals on the laser stimulation is proved.

Contents

List of Selected Abbreviations

Symbol	Description
ATE	Automatic Test Equipment
ATPG	Automatic Test Pattern Generation
BIST	Built-in Self Test
CCD	Charge Coupled Device
CW	Continuous -wave
DALS	Dynamic Analysis of Laser Stimulation
DLS	Dynamic Laser Stimulation
DUT	Device Under Test
DVM	Delay Variation Mapping
FET	Field Effect Transistor
IC	Integrated Circuit
JTAG	Joint Test Action Group
LADA	Laser Assisted Device Alteration
LOC	Launch-on-Capture
LOS	Launch-on-Shift
LSM	Laser Scanning Microscope
MOS	Metal Oxide Semiconductor
NIR	Near infra-red
PLS	Photoelectric Laser Stimulation
PMU	Parametric Measurement Unit
PSD	Phase Sensitive Detector
PVM	Phase Variation Mapping
RIL	Resistive Interconnect Localization
ROM	Read Only Memory
SBGA	Super Ball Grid Array
SCR	Space Charge Region
SDL	Soft Defect Localization
SEI	Seebeck Effect Imaging
SEU	Single Event Upset
Si	Silicon
SOM	Scanning Optical Microscope
SRAM	Static Random Access Memory
TAC	Time-to-Amplitude Converter
TLS	Thermal Laser Stimulation
VLSI	Very Large Scale Integrated
V_{in}	Input Voltage
V_{out}	Output Voltage

List of Selected Symbols

Symbol	Description	Unit
α	Absorption coefficient	1/cm
δ	Penetration depth	cm
C_{ox}	Gate oxide capacitance	F/cm^2
D_n, D_p	Diffusion coefficients for electron and holes	cm^2/s
ε_o	Permittivity of free space	F/cm
ε_{ox}	Dielectric constant of the gate oxide	-
Ec	Conduction band energy	eV
Ev	Valance band energy	eV
ϕ_0	Surface potential	V
γ	Body effect parameter	$V^{1/2}$
I_{DS}	Drain current	amps
J	Current density	$amps/cm^2$
L	Length of MOSFET channel	m
L_n, L_p	Diffusion length for electron and holes	m
μ	Mobility of the charge carriers	$cm^2/V.s$
n, p	Electron/hole concentration	cm^{-3}
N_A	Acceptor concentration	cm^{-3}
q	Magnitude of electronic charge	C
T	Temperature	K
t_{ox}	Gate oxide thickness	m
V_{BS}	Bulk-source voltage	V
VDD	CMOS supply voltage	V
V_{DS}	Drain-source voltage	V
V_{FB}	Flat-band voltage	V
V_{GS}	Gate-source voltage	V
VSS	CMOS ground potential	V
V_{th}	Threshold Voltage	V
W	Width of MOSFET channel	m

1. Introduction

Scanning an IC with a 1064 nm wavelength laser beam stimulates the active devices because of heating and photocurrent induction due to electron-hole pair generation in silicon (Si). Similarly, exposing the active devices with 1300 nm laser beam creates thermal gradients on the interconnects and the active devices. These techniques are based on scanning optical microscope (SOM) and are called photoelectric laser stimulation (PLS) and thermal laser stimulation (TLS), respectively.

Static PLS and TLS techniques were introduced in early 1990's to localize metal shorts/opens, bad contacts/vias, latch-up and junction defects. However, these techniques were not enough to investigate ICs with internal activity. Moreover, timing is becoming an important issue considering the modern day technology ICs. Shrinking geometry and higher clock rate, which result in speed faults, push failure analysis laboratories more and more towards delay testing. Therefore, DLS techniques were developed in order to locate dynamic defects, to analyse the IC performance or to perform timing characterization.

Soft Defect Localization (SDL) [BBE02] and Resistive Interconnect Localization (RIL) [CTH01] are DLS techniques that localize soft defects. The soft defects may be caused by metallization and interconnect defects such as resistive vias, oxide defects, inter-level dielectric defects and process variations (threshold shift, shifts in transistor sizes). All these soft defects cause timing failures and can be only detected while the device is fully running. Induced localized heating in IC interconnections changes the pass/fail status of the functional test identifying the failing site. The IC is activated in a loop while a laser beam is scanned across it pixel by pixel. The pass/fail statuses are recorded for each laser position to fill an image. Any change from pass to fail or from fail to pass pinpoints the soft defect. This technique was successfully applied for localization of the resistive vias.

Laser Assisted Device Alteration (LADA) Technique [RE03] further extends the application of DLS techniques to localize speedpaths. It creates timing shifts by laser scanning. Since thermally induced delay shifts in a golden device are too weak to observe, they use PLS technique. Considering the set-up, the only difference from SDL techniques is the use of PLS instead of TLS. Both SDL and LADA techniques use tester to perform functional test.

Some difficulties of the DLS applications are the synchronization between the laser scanning microscope (LSM) and the tester and the long loop lengths at tester, which makes the experiments impractical. Thus, other approaches are used to speed up the acquisition time. With the Delay Variation Mapping (DVM) technique [SDB05], the laser beam induced delay shift is obtained by measuring the delay time per laser scan position. A high-resolution time measurement unit is employed to measure the picosecond delays. Unfortunately, picosecond delay shifts can be very difficult to measure. Therefore, phase variation mapping (PVM) technique [SPB07] is introduced which uses a phase sensitive detector to detect laser-induced phase variations caused by timing shifts.

The scope of this work is to investigate the timing characteristics of a modern day fully functional "real" IC through laser stimulation. The feature size of the device under test (DUT) is 90 nm. The device runs in dynamic mode, which is driven by automatic test equipment (ATE). The device is scanned by laser beam. The laser illumination induces some effects such as delay variation and soft fault injection. In this work, delay variation and soft fault injection phenomenon are studied. We used two different kinds of laser diodes, emitting laser light at wavelength of 1064 and 1300 nm.

There are two different modes of laser operation: CW and pulsed. We employed both CW

and pulsed PLS as well as CW TLS technique to perform timing analysis of a 90 nm technology test device. All DLS techniques in the literature use CW laser. Because, CW laser is present in most of the laboratories whereas pulsed laser is not very common. On the other hand, pulsed laser has some advantages. For example, the laser can be pulsed on a desired frame of the test sequence to avoid side effects. We used pulsed laser to make the use of its advantages.

First of all, we built various experimental setups for the timing investigation of the devices and they are compared in the means of timing accuracy and acquisition time. Timing analysis of a buffer in a delay chain is carried out in detail. It is shown that localized photocurrent injection or thermal gradients on nMOS or pMOS devices perturb the switching time.

PLS creates a photocurrent on all p-n junctions including bulk-substrate junctions. Therefore, some undesired dummy structures may be stimulated and the well charge up may mask single transistor signatures. The best way to avoid these side effects is to pulse laser at the switching edges of the investigated signals. It limits the substrate current and therefore the stimulation on the secondary structures is avoided and the laser perturbations are more localized as introduced in experimental parts of the thesis.

Moreover, we introduced a new timing analysis methodology for clock driven scan design Integrated Circuits, based on externally triggered pulsed laser stimulation. The 1064 nm laser diode is used in pulsed and CW mode and the scan flip-flops are exposed from the backside of the chip causing soft fault injection. Soft fault injection technique can be used for validation of fault tolerant designs [MFS98], to simulate Single Event Upset (SEU) in memory circuits [BWK87], and for signal propagation imaging [PLF04]. In this work, soft fault injection technique is used to identify the most sensitive signal condition for fault injection with a time resolution correlated to the signal switching time with a better time resolution than the pulse width of the laser, offering new opportunities to failure analysis. The stimulation is done with a laser pulse that is tunable in intensity and delay with respect to clock and scan pattern in order to get maximum precision of the time information.

It is also presented that a current source connected to the drain of the stimulated transistor electrically models the PLS effect. The delay variations on a buffer and soft fault injection in a scan flip-flop are simulated to reproduce the PLS mechanism in software level.

The content of this work is structured as follows: First, the investigated devices and the design for test techniques will be introduced in Chapter 2 and 3, respectively. The tools that are employed for the experiments are described in Chapter 4. This is followed by a description of the laser light-silicon interaction (Chapter 5). Laser stimulation and photon emission techniques are discussed in Chapter 6. The experimental results which are the main part of this work are presented in Chapter 7, 8 and 9. Finally, simulation results will be given in Chapter 10.

2. Device Under Test

The structures that we use for our experiments are delay chains and scan chains. Because, delay chains and scan chains are the most suitable and basic structures for time related analysis. They are located on the same device and this device is provided by Infineon Technology AG, Munich. It is produced with 90nm triple well Complementary Metal Oxide Semiconductor (CMOS) technology and it is in a flip-chip Super Ball Grid Array (SBGA) package. The design information is given in Table 2.1.

The operation principal of the DUT can be seen in Figure 2.1. It consists of n modules that can be connected to the common IO bus, DIO by putting the corresponding address on the block select bus, BSEL. Each module can select which of the DIO pins are used as inputs and which are used as outputs individually. For structural logic testing, scan chains are implemented in the chip. The TESTSE signal switches the scan flip-flops between normal functionality or scan mode. In addition to switching the test enable pin to logic '1', a JTAG programming is needed to enable the clock signal to reach all modules in parallel.

Table 2.1. General design information of the DUT

Chip width (excl. sealring) [µm]	8053
Chip height (excl. sealring) [µm]	8053
Chip area (excl. sealring) [mm^2]	64.851
Sealring width [µm]	43
Number of pad power domains	1
Nominal pad voltage [V]	2.5
Number of power/well power domains	16
Nominal core voltage [V]	1.2
Number of power pads	94
Number of signal pads	128
Total number of pads	232
Total SRAM size [M bit]	9.0
Total ROM size [M bit]	1.0

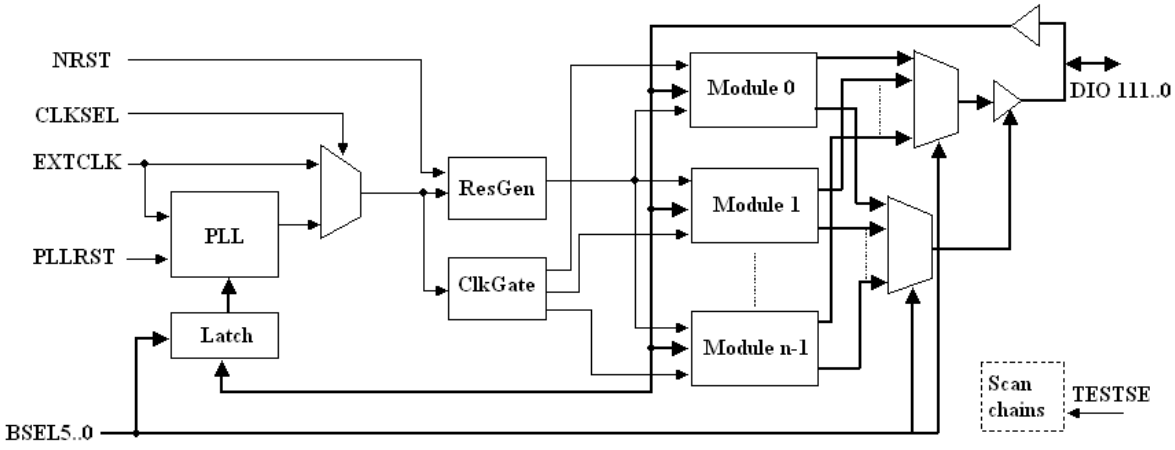


Figure 2.1. Operation principal of the DUT

This complex state of the art IC is a test device and there are many test structures to do some investigations, such as leakage and performance analysis (timing and correlation with the simulations). Some structures that are on the DUT are delay chains, ring oscillators, shift registers, scan chains, SRAMs and ROMs.

The reflected light image of the DUT from the backside is shown in Figure 2.2. There is no block assigned for the delay chains, but the inverters and buffers that are between input and output pads are spread all around the chip area.

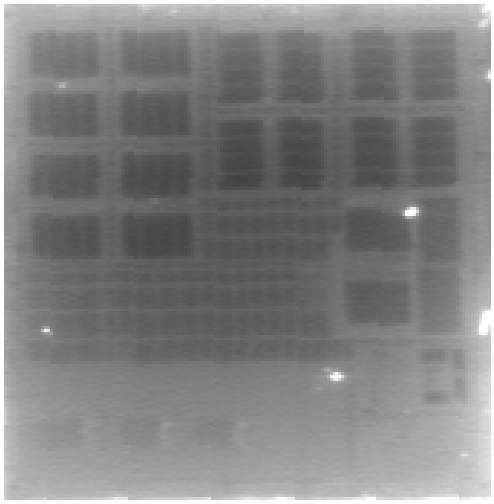


Figure 2.2. The reflected light image of the DUT obtained by 0.5x objective lens

The DUT is produced with a triple-well CMOS process. In the triple-well process, a p-well is placed inside an n-well, resulting in three types of well structures, as shown in the illustration in Figure 2.3. The triple n-well covers all the devices including the nMOS transistors forming a reversed biased junction.

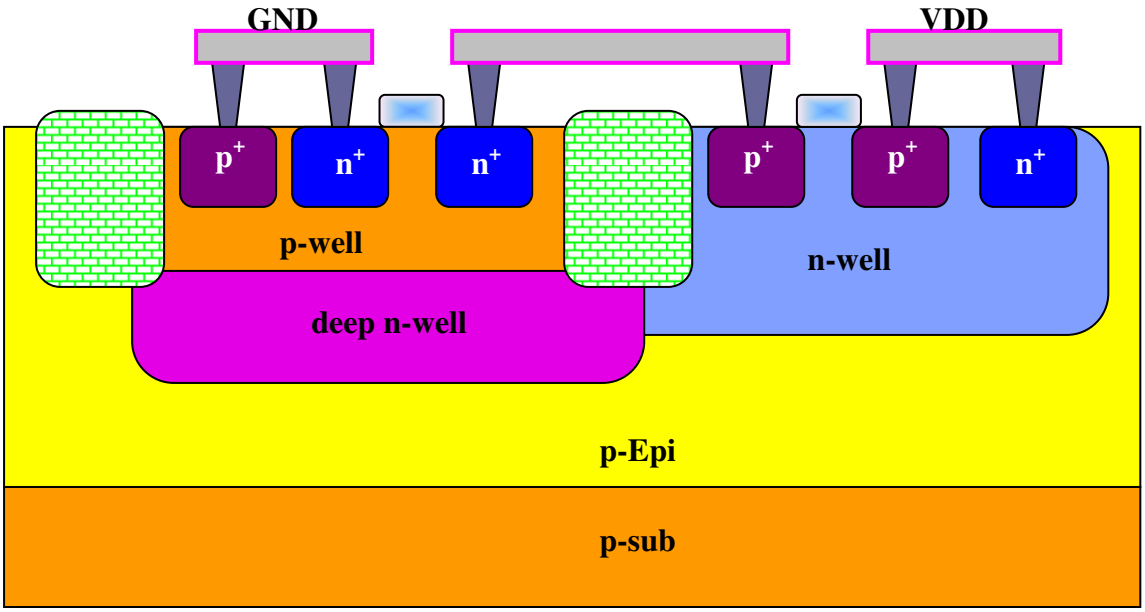


Figure 2.3. The cross-section of an nMOSFET and a pMOSFET

Triple-well technology is mostly used in mixed signal devices to isolate the analog circuitry from the digital logic where a noise injection can be a problem. The isolation is achieved by the reverse biased junction between the n-well and the p-substrate. The p-well connected to the analog VSS is isolated from the digital VSS/ground by the reversed biased junction, which avoids a resistive path between analog and digital circuitry.

2.1. Characterization of MOS Transistors

The metal-oxide-semiconductor field-effect transistor (MOSFET) is the most important device for state of the art complex very large-scale integrated (VLSI) circuits.

Considering the long channel MOSFET (the channel length L is much longer than the sum of the source and drain depletion layer widths), drain current, I_{DS} can be expressed as

$$I_{DS} = \frac{\mu\varepsilon_{ox}}{2t_{ox}} \cdot \frac{W}{L} \cdot \left[2(V_{GS} - V_{th})V_{DS} - V_{DS}{}^2\right] \quad \text{(linear region, } (V_{GS}\text{-}V_{th}) > V_{DS}\text{)} \tag{2.1}$$

$$I_{DS} = \frac{\mu\varepsilon_{ox}}{2t_{ox}} \cdot \frac{W}{L} \cdot (V_{GS} - V_{th})^2 \quad \text{(saturation region, } (V_{GS}\text{-}V_{th}) \leq V_{DS}\text{)} \tag{2.2}$$

where V_{th} is the threshold voltage of the transistor, V_{GS} is the gate-source voltage, V_{DS} is the drain-source voltage, μ is the electron mobility, ε_{ox} is the dielectric constant of the thin oxide (SiO_2), t_{ox} is the transistor oxide thickness, and W and L are the effective gate width and length of the transistor.

The threshold voltage can be written as

$$V_{th0} = V_{FB} + \phi_0 + \gamma\sqrt{\phi_0} \tag{2.3}$$

where V_{FB} is the flat band voltage, ϕ_0 is the surface potential in strong inversion, γ is the body effect parameter [Tsi99]. When a substrate bias is applied, the threshold voltage changes. This is called body effect. Since the body influences the threshold voltage (when it is not tied to the source), it can be thought of as a second gate, and is sometimes referred to as the "back gate". For an nMOS transistor body effect upon threshold voltage can be expressed as

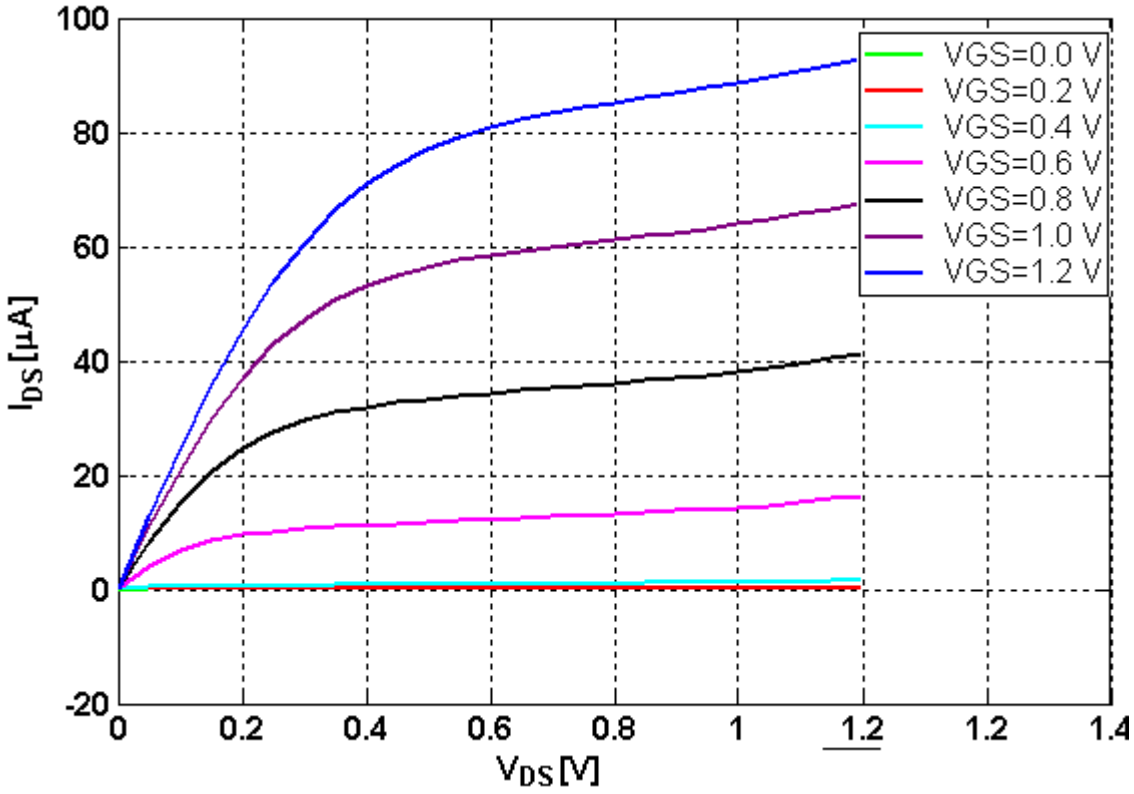
$$V_{th} = V_{FB} + \phi_0 + \gamma\left(\sqrt{V_{SB} + \phi_0}\right)$$

γ can be written as

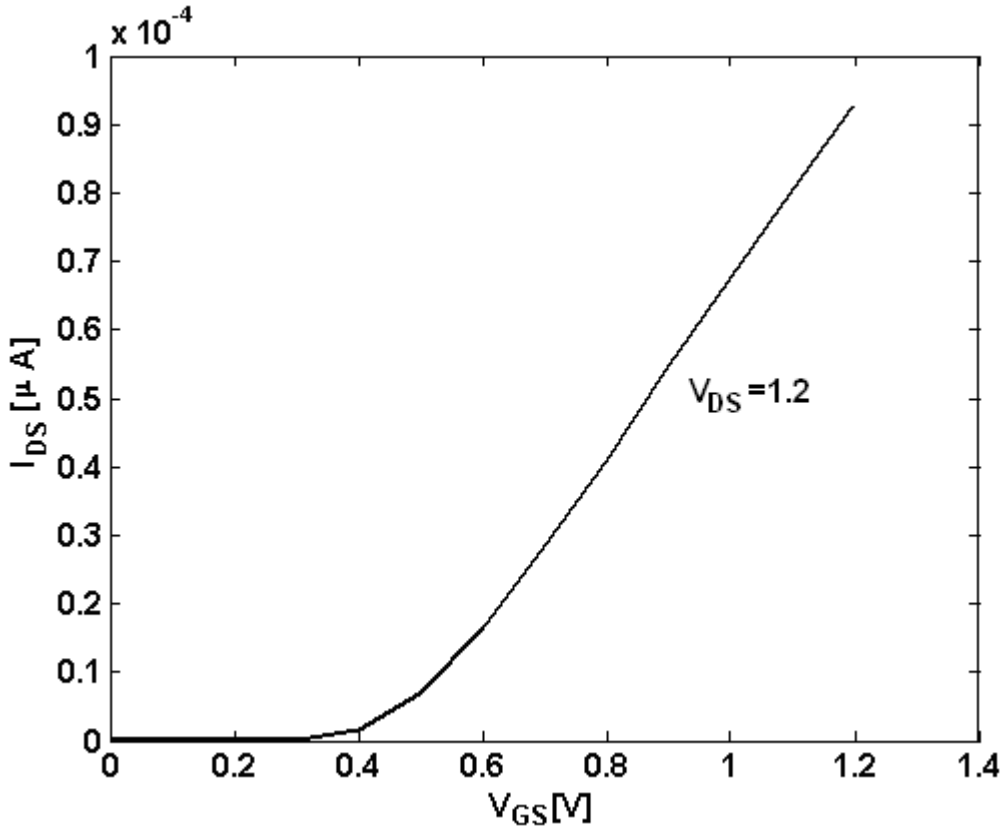
$$\gamma = \frac{\sqrt{2\varepsilon_s q N_A}}{C_{ox}} \tag{2.5}$$

where ε_s is the permittivity of silicon, q is the magnitude of electronic charge, N_A is the acceptor concentration and C_{ox} is the oxide capacitance.

The characteristics of 90 nm single transistor test structures are measured by a parameter analyser. The output characteristics of sub-micron (W/L=0.12μm/0.08μm) nMOS and pMOS transistors are shown in Figure 2.4 and Figure 2.5, respectively.

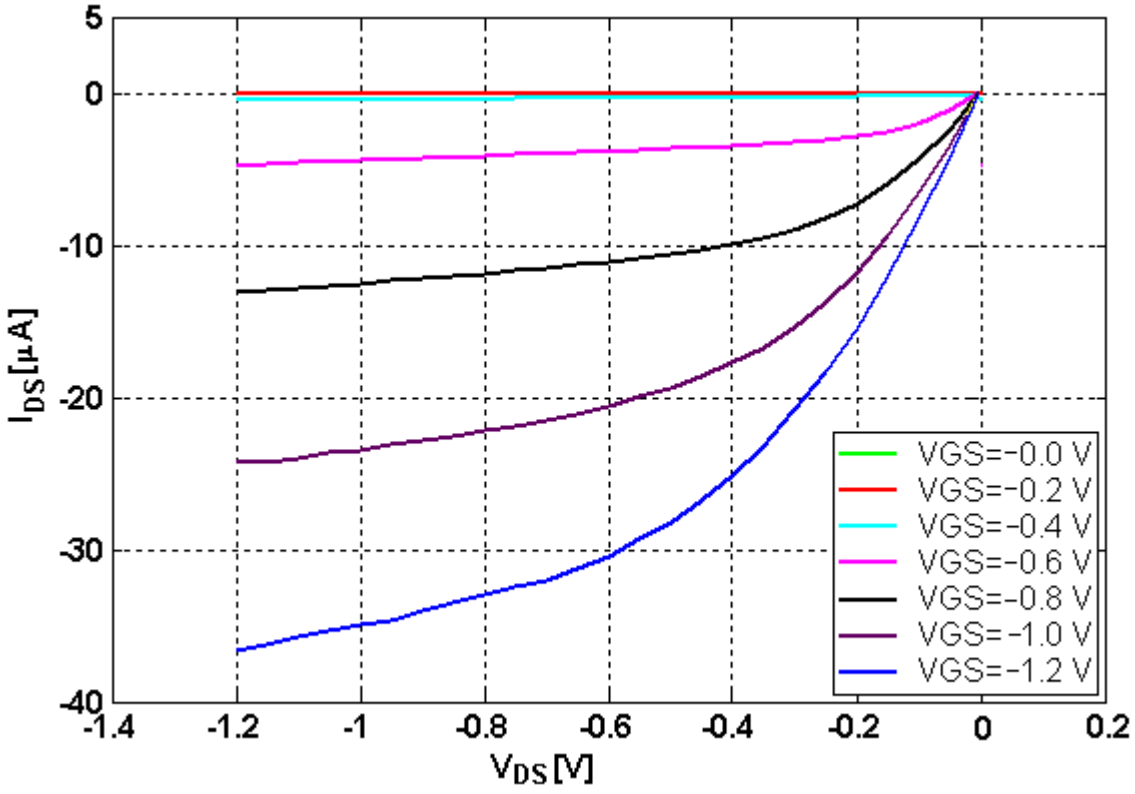


(a)

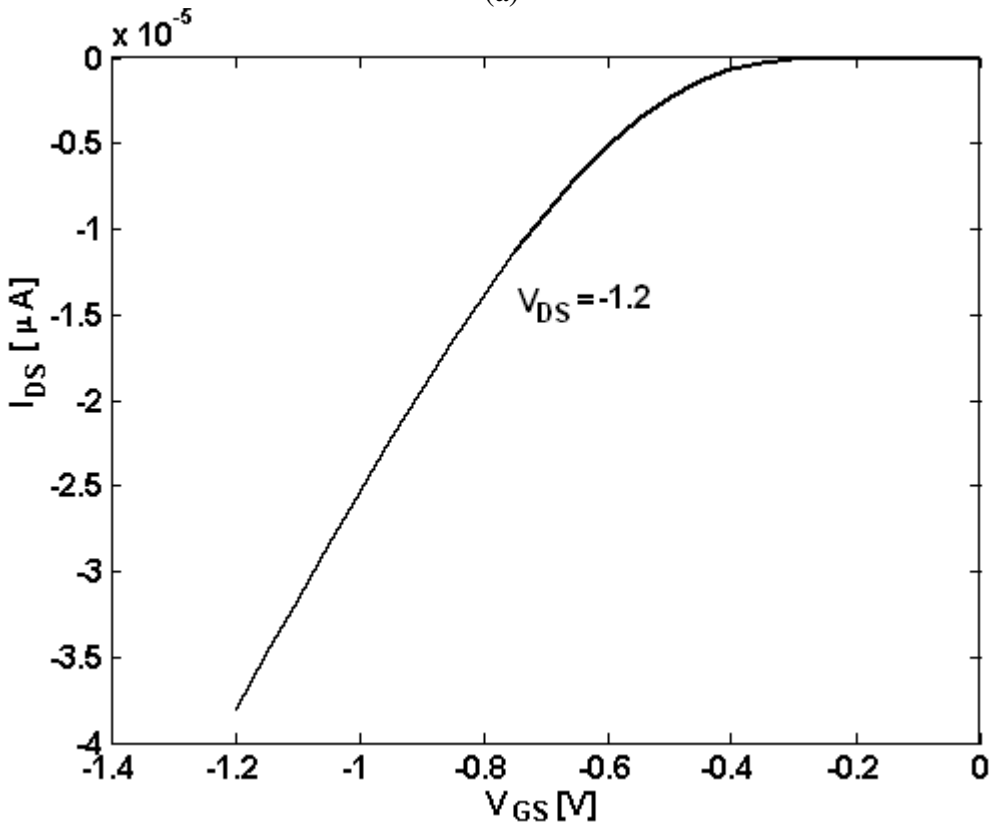


(b)

Figure 2.4. (a) The output and (b) the transfer characteristics of a 0.12 μm/0.08 μm size nMOS transistor



(a)



(b)

Figure 2.5. (a) The output and (b) the transfer characteristics of a 0.12 μm/0.08 μm size pMOS transistor

2.2. Inverter

The inverter is the nucleus of all digital designs. Figure 2.6 shows the circuit diagram of a Complementary MOS (CMOS) inverter. When V_{in} is high and equal to *VDD*, the nMOS transistor is on and the pMOS transistor is off. A direct path exists between V_{out} and the ground node, pulling the voltage to 0 V. On the other hand, when the input voltage is 0 V, nMOS and pMOS transistors are off and on, respectively. A path exits between *VDD* and V_{out}, yielding a high output voltage. The principal input and output signals of an inverter are shown in Figure 2.7.

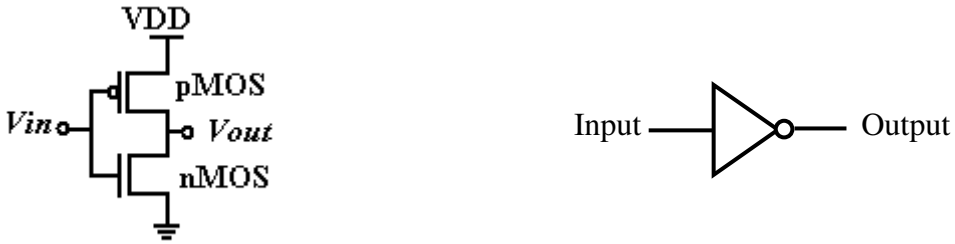


Figure 2.6. CMOS inverter (a) Transistor based (b) Logic symbol

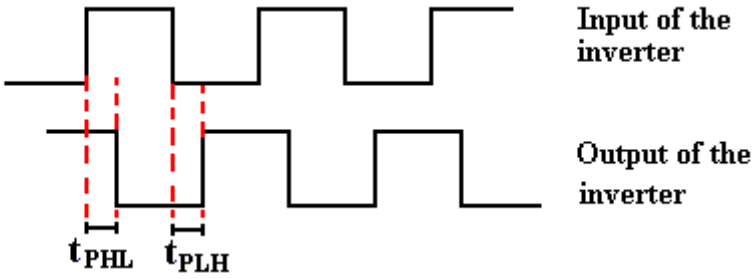


Figure 2.7. The principal input-output signals of an inverter

2.3. Delay chain

The delay chain is a circuit that has delay elements connected one to another. The principal of a delay chain is shown in Figure 2.8. These delay elements are mostly inverters or buffers and the inverters are the most basic structures in digital circuits. A buffer is constructed by connecting two inverters in series.

Each gate has a delay and the overall delay of the chain is the sum of the delay of each gate and the propagation delay caused by the interconnects. There is no need for clock signal. An applied pulse stream to the input of the chain is conveyed to the output of the chain with a certain amount of delay.

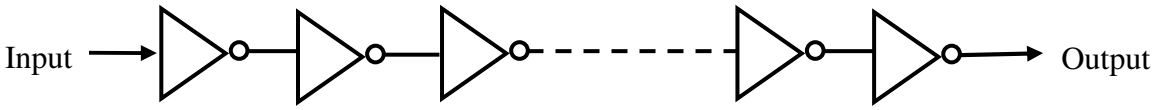


Figure 2.8. Principal of a delay chain (each symbol denotes one inverter)

There are 56 delay chains integrated on the DUT. We picked one of these delay chains to investigate the effect of laser stimulation on the propagation delay of an inverter. The delay time between the rising edges of the input and the output signals for that particular delay chain is seen in Figure 2.9 without laser. The delay decreases with increasing supply voltage up to a level of 1.5 V. According to the simulations, the delay for a small size inverter is around 20 ps.

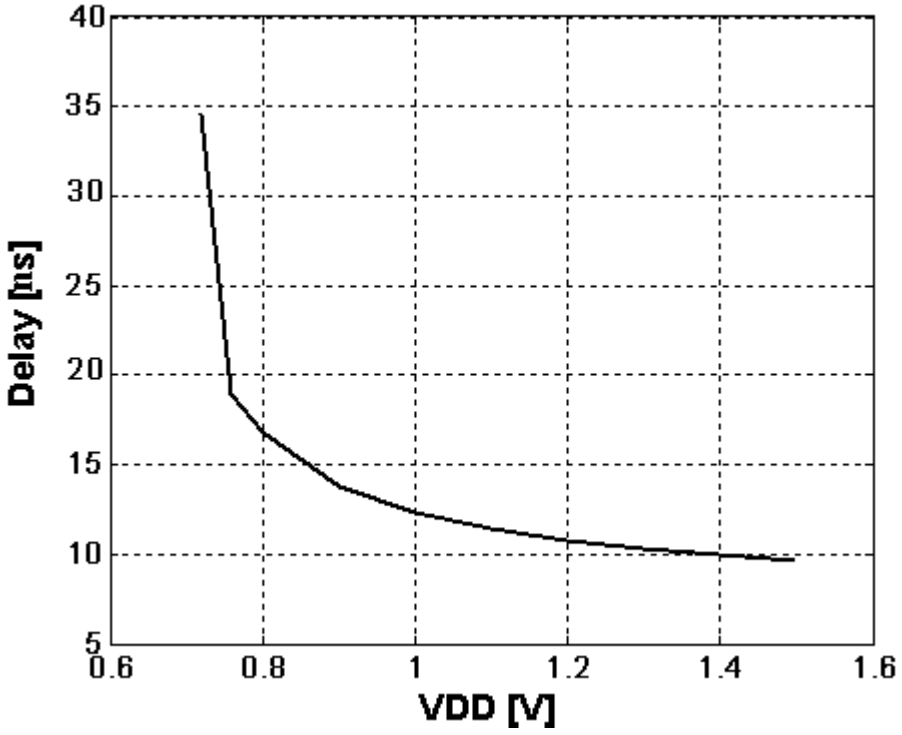


Figure 2.9. The delay versus supply voltage graph for the delay chain under investigation (measured from the rising edge of the input to the rising edge of the output)

2.4. Investigated Buffers

A buffer consists of two inverters. The output of the first inverter is connected to the input of the second one. The schematics of the investigated buffers are shown in Figure 2.10. Buffer1 consists of inverters and transmission gates. Transmission gate is made by the parallel combination of an nMOS and pMOS transistor with the control signal at the gate of one transistor being complementary to the control signal at the gate of the other. In Buffer1 circuit, there are two paths and *S1* and *S2* signals select one of these paths. The path from *Input1* to *Output* is active while the path from *Input2* to *Output* is inactive. The second buffer (Buffer2) is a simple buffer without any select circuitry and it is the next buffer in the chain

after Buffer1. We mostly selected an area, which includes Buffer1 and Buffer2 for investigations. The reflected light images of the selected area are shown in Figure 2.11 and the layouts of Buffer1 and Buffer2 are seen in Figure 2.12.

The important structures, the buffers and the filling structures, are pointed in the layout. The filling structures are some p-implanted active areas, which are connected to the ground line. Those structures are situated between the logics and they are important for photoelectric laser stimulation. Because, they cause a delay shift when they are illuminated, which will be discussed in more detail Section 8.3. Since the delay chain is produced with a triple-well technology, the whole area is covered with buried n-well.

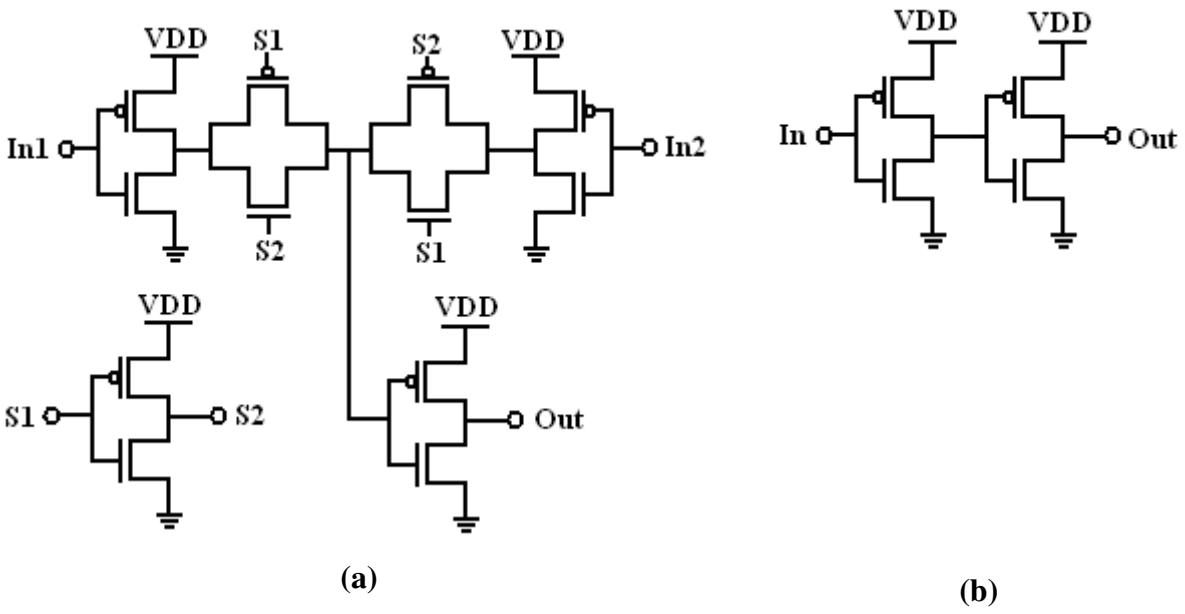


Figure 2.10. The schematic of the localized buffers (a) Buffer1 (b) Buffer2

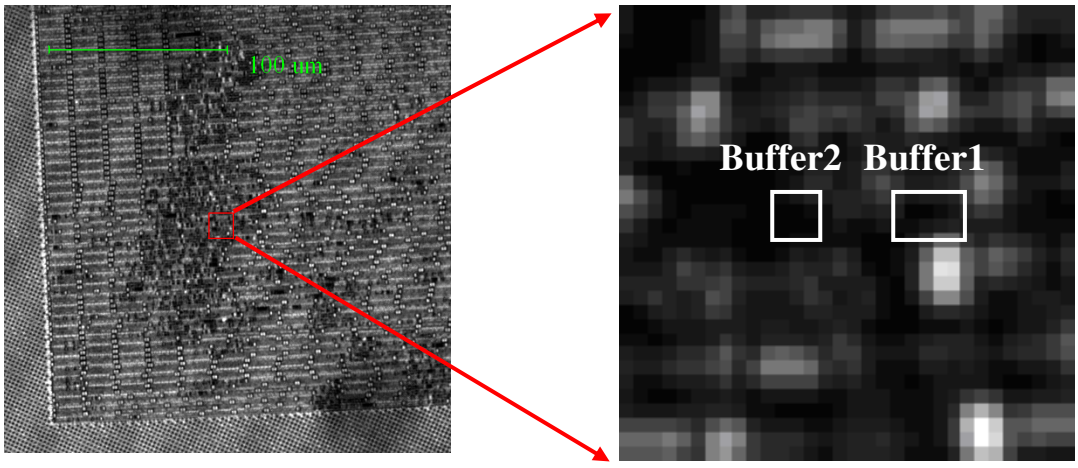


Figure 2.11. (a) The reflected laser micrograph of a 512x512 pixel area on DUT obtained by 50x objective lens (b) The reflected laser micrograph of the scanned area (32x32 pixel)

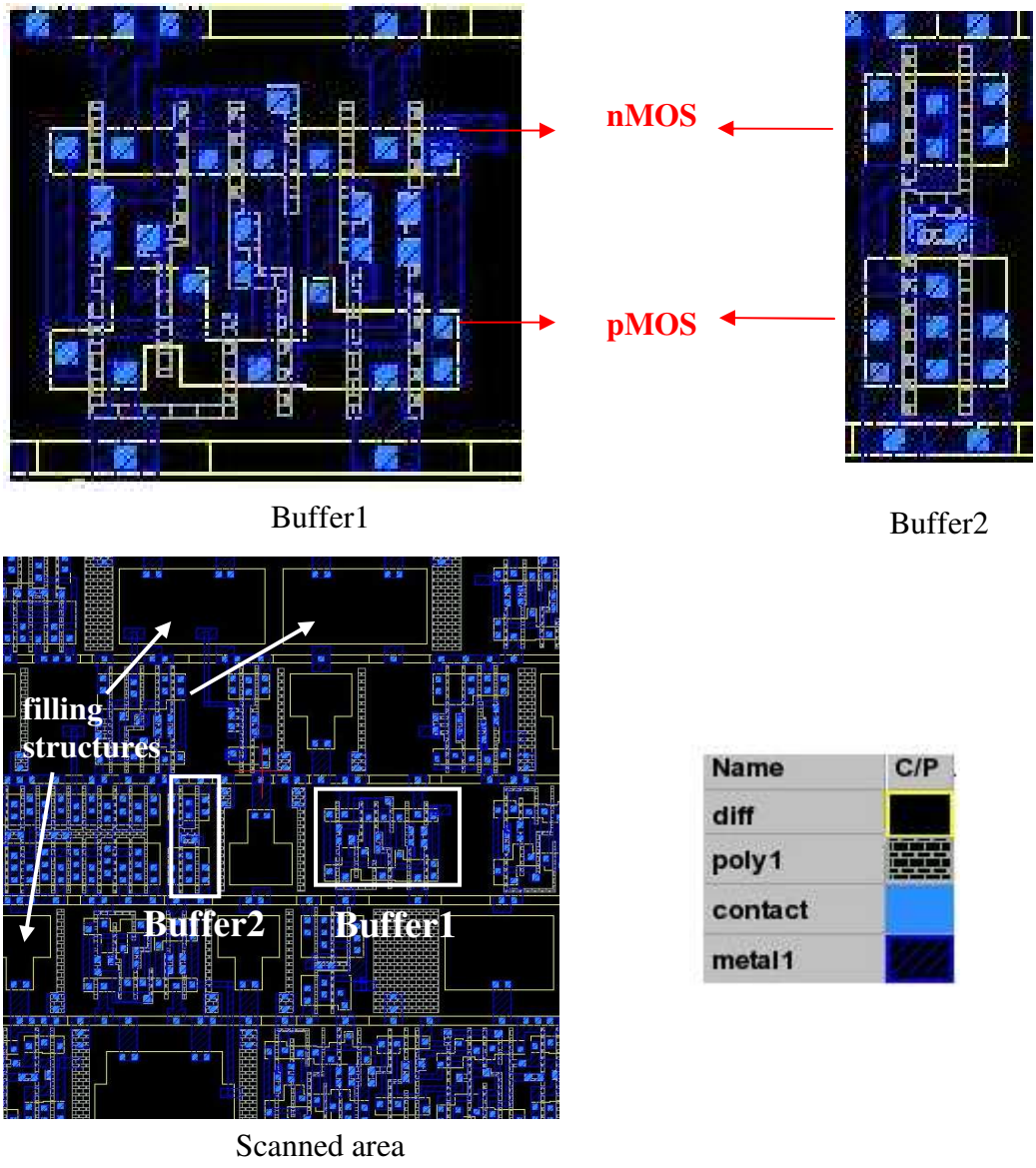


Figure 2.12. The layout of the localized buffers, Buffer1 and Buffer2 (Metal1, Contact, Poly, active areas are highlighted only)

2.5. Scan Chains

For structural logic testing, fifty scan chains are implemented in the test device and one of these scan chains, which has 259 scan cells, was investigated for our experiments.

The schematic of one scan flip-flop is seen in Figure 2.13 in logic level and in Figure 2.14 in transistor level. The layout of the scan flip-flop is presented in Figure 2.15. It is a rising edge triggered two stage master-slave flip-flop. At the input stage, there is a multiplexer selecting either data or test-in input by a *Test Enable* (*TE*) pin. A multiplexer is a device that selects one of many input signals and forwards this selected input signal to the ouput. The master stage is driven by the inverted clock signal, while the second stage is driven by the clock signal. Therefore, the first stage is low level sensitive, while the second stage is high level sensitive. The function table is shown in Table 2.2.

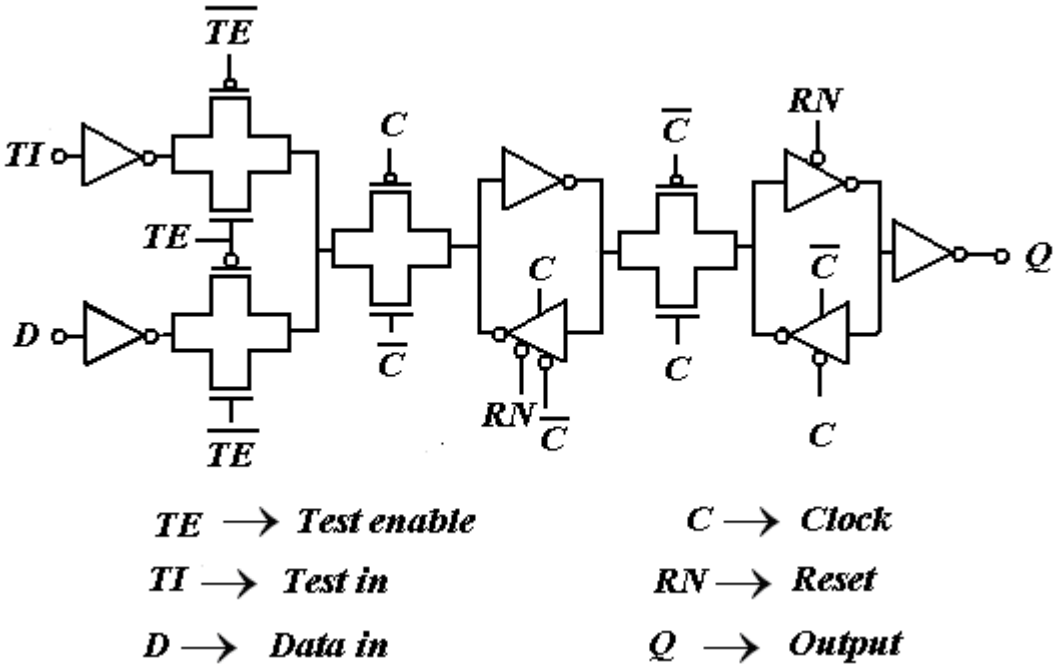


Figure 2.13. Schematic of the scan flip-flop in logic level

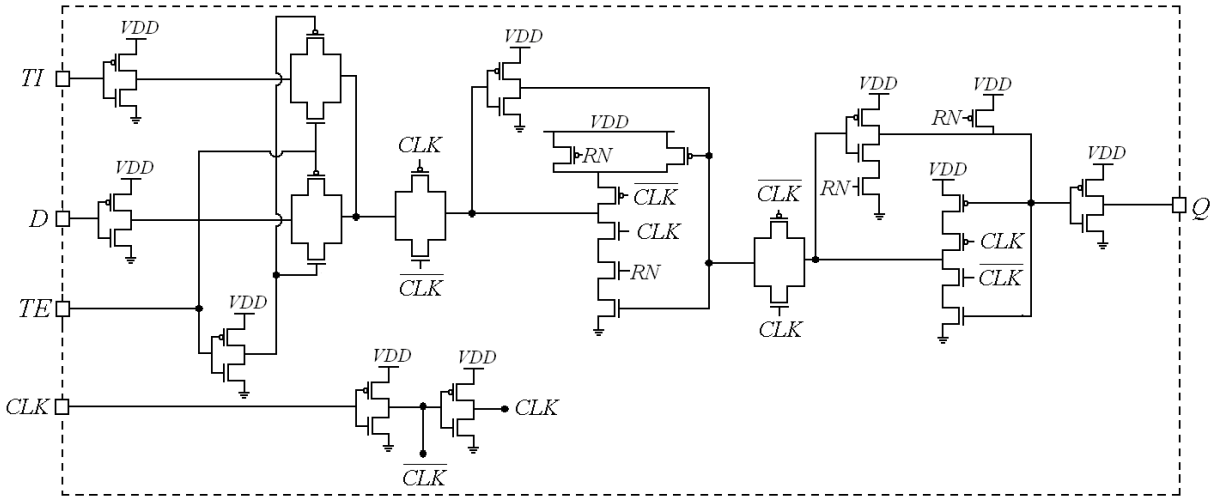


Figure 2.14. Schematic of the scan flip-flop in transistor level

Table 2.2. Function table of the scan flip-flop

TI	D	TE	CLK	Q
x	0	0	↑	0
x	1	0	↑	1
0	x	1	↑	0
1	x	1	↑	1
x	x	x	0	last Q
x	x	x	1	last Q

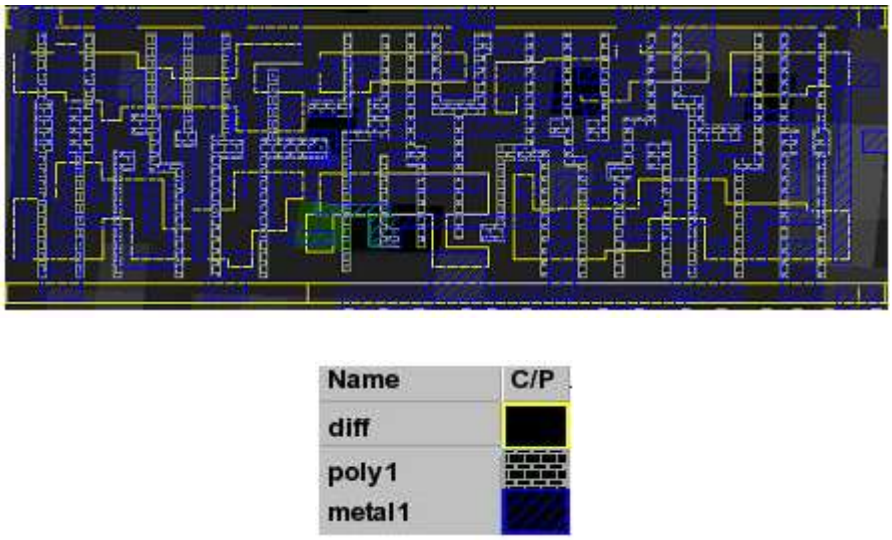


Figure 2.15. The layout of the scan flip-flop (diffusion, metal1 and poly1 are only highlighted)

3. Design for Test (DFT)

Testing is one of the most expensive steps in manufacturing. The testing complexity increases the cost of test generation, the cost of fault simulation and the time to locate and isolate a fault [BA00]. These costs can be lowered by increasing the testability of a device. Testability measures how easy it is to create a vector pattern to test the design quality of a device. The three key parameters for testability are controllability, observability and predictability. Controllability is the ability to drive an internal node to logic '0' or '1' by circuit's inputs and observability is the ability to determine any internal node state by observing circuit's outputs. Predictability is to obtain expected values when a known input stimulus is applied.

DFT refers to some special design styles and added logic to increase the testability of the device. Some examples to DFT techniques are built-in self-test (BIST), boundary scan test and scan design (also referred to as internal scan).

BIST technique adds some extra hardware and software features to allow ICs to test their own circuitry thereby increases a design's testability. The increased complexity of integrated circuits requires mixed signal testers, which have special digital and analog testing capabilities. However, the fast-rising costs of ATE testing and the growing complexity of the devices forces the designers to develop new BIST structures. BIST can be used to run tests with the aid of additional on-chip testing circuits, which eliminates or minimizes the need for an ATE. BIST can also test the circuits, which are not connected to the external pins of the device, for example embedded memories. Comparing with ATE, self-testing might be the best solution for high-speed chips in the near-future.

Boundary scan-test concept is used in complex chips or printed circuit boards (PCBs) in order to isolate one module from the other during testing. Boundary scan adds scan circuitry onto the PCB to make internal circuitry accessible via an interface. It tests the interconnects between integrated circuits on a board without using physical test probes. In boundary scan-test, a scan cell that includes a multiplexer and a latch is added to each pin on the device. For normal operation, the boundary scan cells are set so that they have no effect on the circuit. However, when the circuit is set into a test mode, a data stream is shifted into the serial data (scan) path, which forces data onto pins. Scan cells can also capture data from pin or core logic. The captured data is serially shifted out and compared with the desired results. Boundary-scan allows direct access to the each sequential element, thereby eliminating the need for large number of test vectors, which are normally needed to initialize sequential logic. The advantages of boundary scan testing are shorter test times, increased diagnostic capability and lower equipment cost [CL05].

Among all the others, the most extensively used DFT technique is scan-design testing. Testing a sequential circuit is a difficult task compared with a combinational circuit testing. To establish a desired internal state in a sequential circuit requires many cycles and the response of a sequential circuit is a function of its initial state and the inputs. Similarly, in order to observe the results of the test, the results must propagate from internal nodes through sequential elements to the output pins. Moreover, the complexity of the test generation task increases with the sequential depth of the circuit. Being a DFT technique, scan design makes a sequential circuit easier to test by making each register controllable and observable, which reduces the test problem to the combinational logic test between the registers.

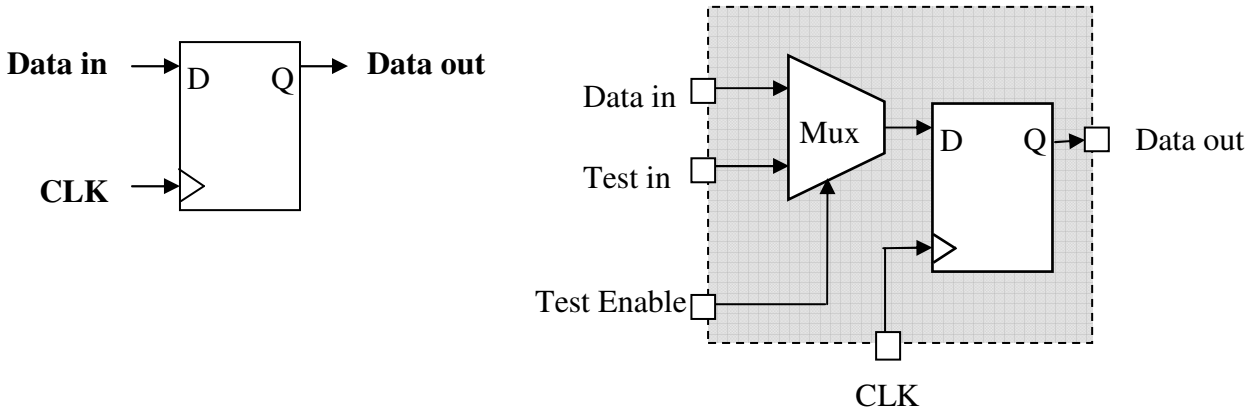


Figure 3.1. (a) Logic symbol of a D flip-flop (b) Circuit design of a scan flip-flop

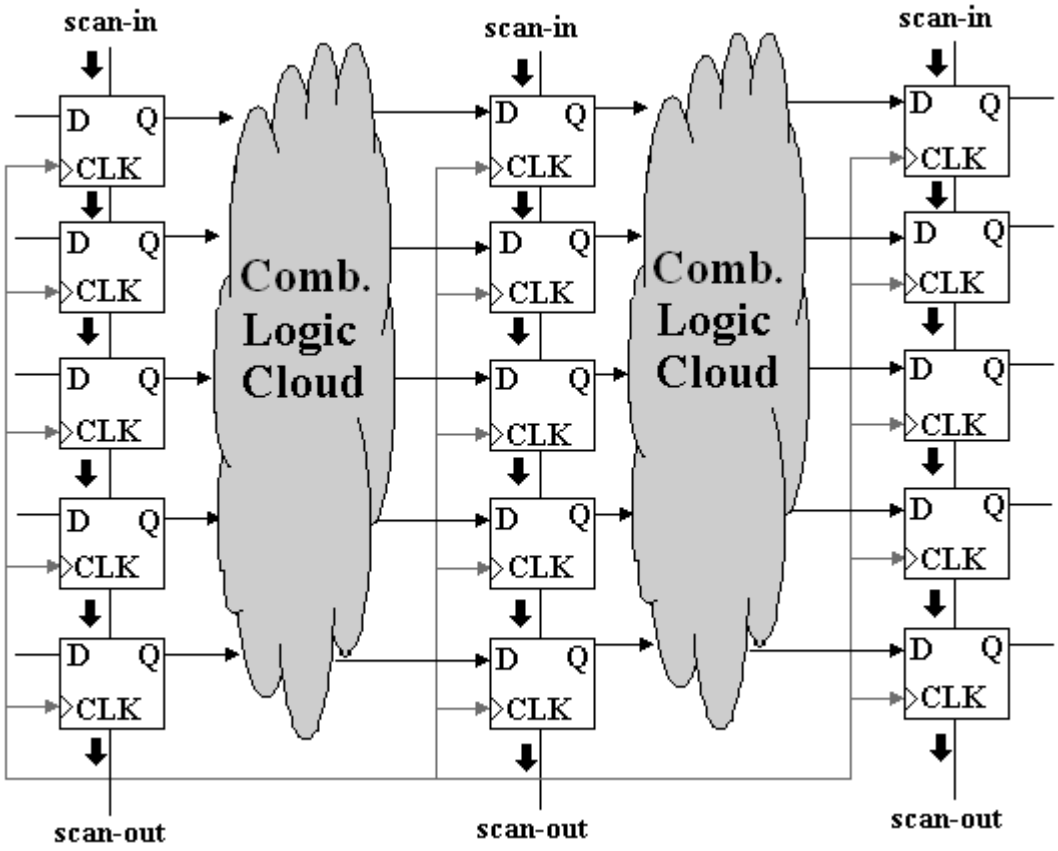


Figure 3.2. Scan design integrated circuit

As it was mentioned above, if a sequential circuit should be tested in a given state, it may require many clock cycles for the circuit to enter the desired state. However, adding a load pin can solve this problem. As it is seen in Figure 3.1, each flip-flop can be converted into a scan flip-flop by adding a multiplexer, thereby making flip-flops observable and controllable.

The sequential elements are replaced with scannable sequential elements (scan cells) and stitched to together into scan registers. These serially-connected scan cells can be used to shift data in and data out when the design is in scan mode. The *Test Enable* pin switches the scan flip-flops between normal functionality or scan (test) mode. In normal mode, scan flip-flops

behave as usual and in test mode, they form a shift register (see Figure 3.2).

The operation of the scan circuitry is as follows:

1. In order to initialize the scan cells, set the scan chain to test mode to allow shifting.

2. Hold the scan clocks off and apply data to the primary inputs.

3. Measure the outputs.

4. Switch to normal mode and pulse the clock to capture new values into scan cells from combinational logic. This cycle is called *sample cycle*.

5. Set the scan chain to test mode to unload and measure the captured values while simultaneously shifting in new data values.

3.1. Scan-Based Testing

There are four different kinds of vector patterns that can be used for testing. These are scan-shift pattern, scan logic pattern, transition delay pattern and path delay pattern.

Scan shift and scan logic are DC test patterns and they are based on stuck-at fault model. The faulty node behaves as if it is tied to either VDD (stuck-at-1) or GND (stuck-at-0).

On the other hand, transition delay and path delay test patterns include the concept of timing, thus known as AC tests. The only difference between DC and AC tests is the clocking.

AC tests observes slow transition signals at a gate or abnormal large propagation delays on interconnect signals via making the clocking critical. And then, the timing failures are converted into Boolean failures. Therefore, these methods are able to determine the speed of the device and identify the high-speed related defects.

1) Scan shift pattern:

The purpose of the scan shift pattern is to test the scan chain integrity. Before testing the combinational logic between the scan registers, one should make sure that scan flip-flops function properly [Cro05]. The defects within the scan registers cause an incorrect loading and unloading of the scan registers which makes the testing of combinational logic through the scan registers impossible. Therefore, problems in the shift process require scan-based debug and diagnosis before the scan chain can be used for testing. The root cause of the shift problems can be stuck-at faults which sets the scan-out pattern to a fixed value, invalid scan enable signals which is caused by improper connections or hold time problems.

For scan shift testing, the scan flip-flops work always in test mode. The scan flush test pattern, ‘00110011…’, which has all the transitions, is applied to the scan-in input. The length of the test pattern should be equal to the length of the chain. If the chain works properly, the applied pattern is shifted out unchanged.

If there is a single stuck-at fault blocking the data, this fault can be isolated by applying a pattern that has a logic 0 (1) before a stuck-at-1 (stuck-at-0). Another way is to apply a sample pattern, which toggles bits just before and after the break with logic values that are the opposite of the stuck-at values.

In order to isolate the hold-time bits, the scan flush pattern is applied. If a hold-time problem is present, a bit comes out early in shift out process and the whole pattern is shifted by one bit.

In case of multiple hold-time problems or stuck-ats, more complex algorithms are required for debug and diagnosis.

2) Scan logic pattern:

The purpose of a scan logic pattern is to test the system logic between the scan registers. It consists of a shift-in operation; one or multiple capture clock cycles and a shift-out operation. The timing for the scan logic pattern is shown in Figure 3.3.

The test procedure is as follows:

1. Set the DUT in test mode and load the desired state by shifting the data in through the scan flip-flops
2. Apply the test pattern to the combinational logic by switching *Test Enable* pin from test mode to normal mode and measure the primary outputs.
3. Trigger the scan clock and capture combinational logic output values to scan flip-flops.
4. Switch back to the test mode and shift out the responses and compare with the expected values.

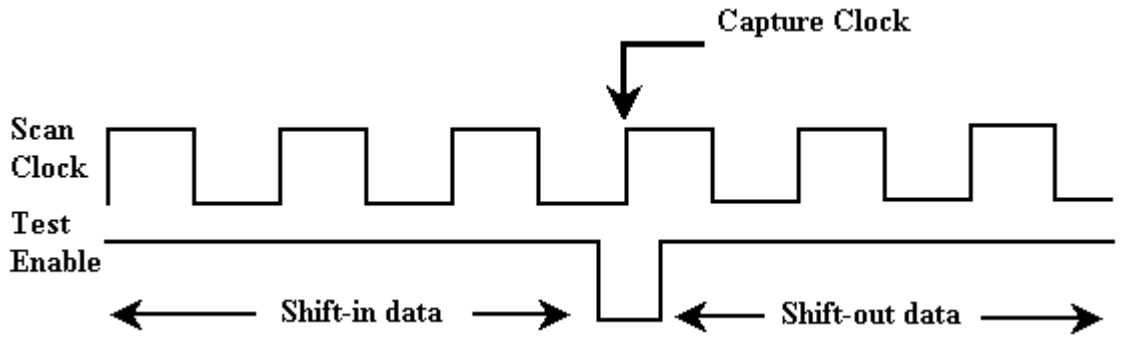


Figure 3.3. Scan operation for the scan logic pattern

3) Transition delay pattern:

The *slow-to-rise (slow-to-fall)* transition fault models a device pin that is defective because its value is slow to change from 0 to 1 (1 to 0). In order to detect transition faults, it is needed to make a transition on a node and propagate this transition to the observable outputs. Therefore, two vectors are needed to perform transition delay testing. The first one initializes the value of the node, the second vector makes a transition on the node and sensitizes the transition to the observable outputs. Commercial ATPG tools support two kinds of delay test sets: Launch-on-Shift (LOS) and Launch-on-Capture (LOC). Figure 3.4 and 3.5 presents the timing diagram of LOS and LOC patterns.

LOS transition delay pattern is very similar to scan logic test pattern. The only difference is that capture (sample) cycle has a higher frequency than shifting cycle, which is called cycle shrinking or cycle switching. The at-speed cycle can be performed without changing the frequency. The launch clock is triggered late in the period and the capture clock is set early in the next period, then the technique is called clock-chopping or duty-cycle modulation. The test enable is triggered between last shift clock and capture clock which makes test enable pin a critical signal. In this method, test enable signal should be distributed very carefully around the device like a clock signal to avoid any racing problem between the cells.

This problem is solved with the LOC pattern. It conducts two sample cycles instead of one (see Figure 3.5). Since the test enable is triggered one cycle before the sample cycle, it is not critical anymore. However, this method brings some restrictions: it suffers from large test

size and low fault coverage compared to LOS test set. In addition, it requires more ATPG computation which means longer run time.

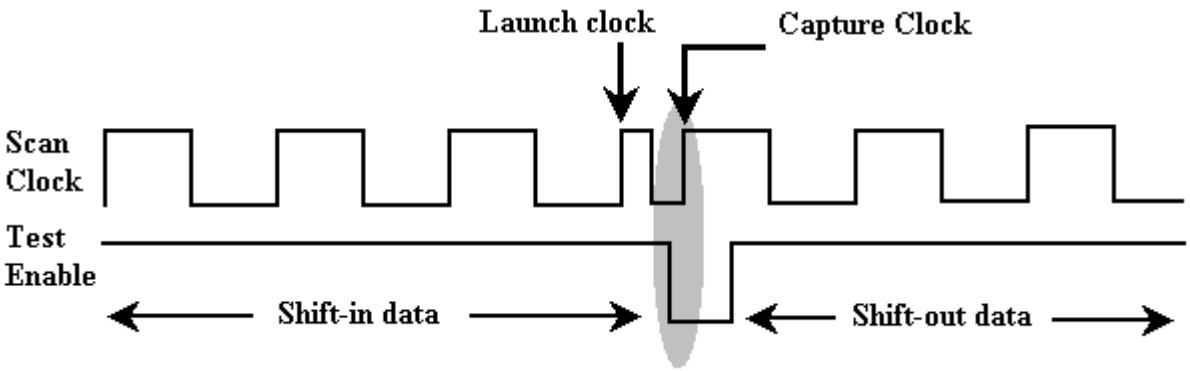


Figure 3.4. Scan operation for LOS pattern

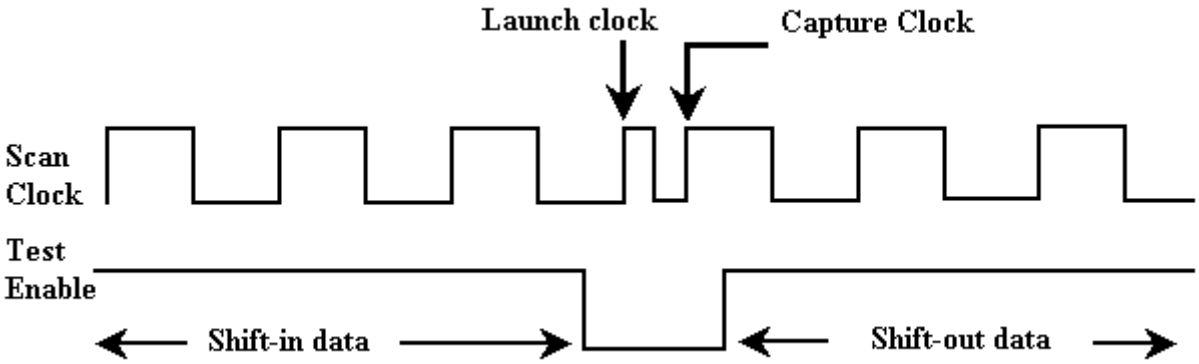


Figure 3.5. Scan operation for LOC pattern

4) Path delay pattern:

While the transition delay models slow transitions and large delays on gates or routes, path delay test pattern is applied to detect slow transition signals over a complete critical path made of nets, nodes and gates. Path faults can be viewed as the accumulation of a collection of transition delay faults. LOS or LOC methods same as transition delay are used to form the path delay pattern.

4. Tools

In order to run dynamic laser stimulation analysis of the devices, it is required to have a Laser Scanning Microscopy (LSM) and an ATE. ATE (also called tester) drives the inputs of the DUT and it is also used to measure the device outputs for some applications. LSM is employed to scan the devices with the laser beam.

4.1. Automatic Test Equipment

Once an IC is in production, testing is essential to provide the customer with working devices and to guarantee low defect levels. Since users of the chips assume that the chip functions, we need to have some method to sort out the functional chips from the bad ones. This process is called production test, and is a critical part of manufacturing. During production test, ATE is used to reduce the test times by automating the test process. The basic purpose of a tester is to drive the inputs and to monitor the outputs of a DUT. Most important features for a tester are speed, timing accuracy and number of input/output pins available. We use Agilent 83000 tester in our laboratory. The specifications are shown in Table 4.1.

Table 4.1. The specifications of the Agilent 83000 Tester

Test Processor-Per-Pin Architecture:	
Maximum clock rate	330 MHz
Maximum channel (pin) count	256
Timing Specifications:	
Edge placement resolution	10 ps
Edge placement accuracy	150 ps
Edge-to-edge clock signal skew	75 ps
Overall timing accuracy (for 150 mV swing)	± 300 ps
AC Performance of the Drivers:	
Transition time into 1 kΩ ‖ 5 pF load CMOS level (0V to 3 V, 10% to 90%)	1.1 ns
DC Performance of the Drivers:	
High level range	-1.5 V to 7.0 V
Low level range	-2.0 V to 6.5 V
Output swing	0.5 V to 6.0 V
Level resolution	10 mV
Level accuracy	± 20 mV
AC Performance of the Comparators:	
Minimum detectable pulse width	1.5 ns
DC Performance of the Comparators:	
High threshold range	-1.9 V to 7.0 V
Low threshold range	-2.0 V to 6.9 V
Minimum high/low threshold distance	100 mV
Threshold resolution	10 mV
Level accuracy	160 mV

ATE has four major components (Figure 4.1):

- The mainframe, which contains vector memory, timing generator and sequencing circuits, as well as device power supply modules (DPS).
- The testhead containing pin electronics cards: drive, receive, loads, relays and Parametric Measurement Units (PMUs). It can be positioned by the aid of a manipulator and has huge degrees of freedom.
- The cooling unit, which circulates chilled water to cool the IO boards in the mainframe and the pin electronics boards in the testhead.
- The workstation, which serves as the controller for the system, communicating via 2 sets of fiber optic cables (one to the mainframe and one to the testhead) and HP-IB.

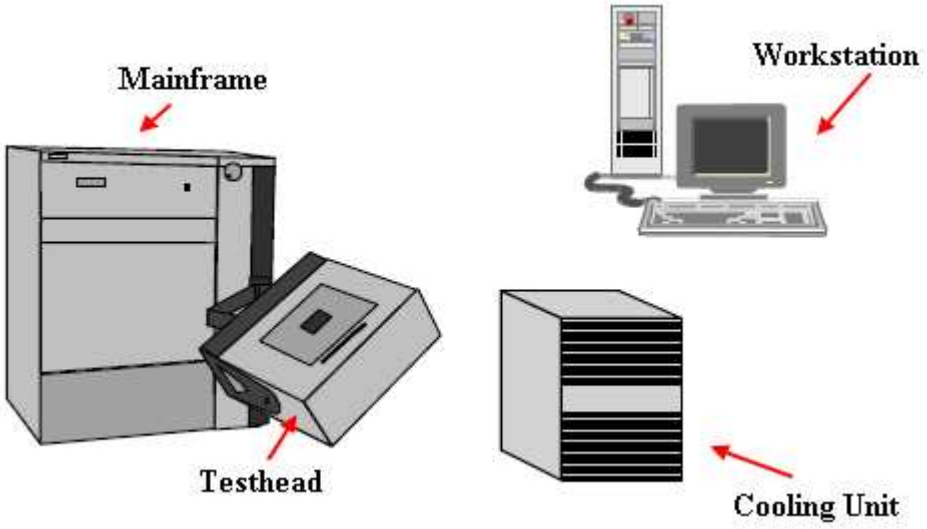


Figure 4.1. Major components of an ATE [Agi00]

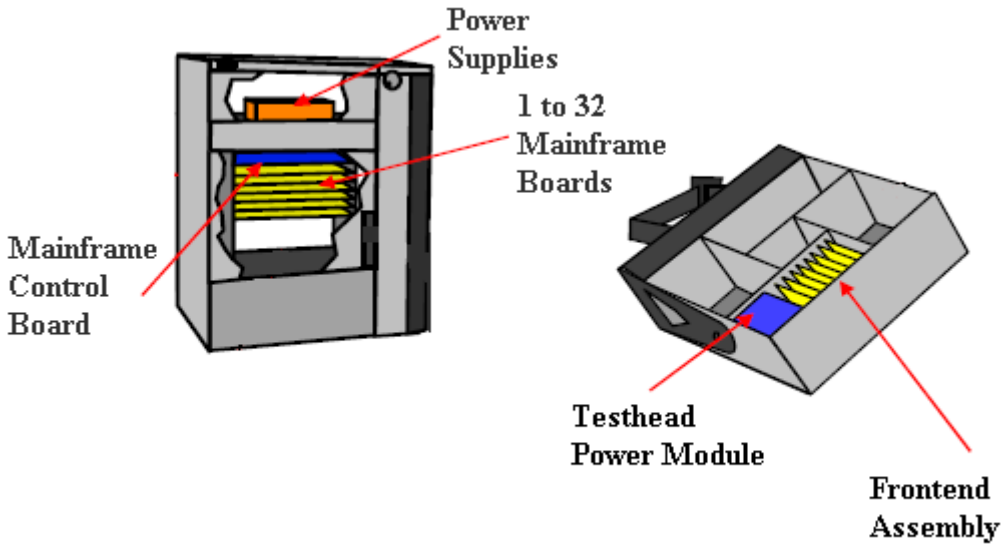


Figure 4.2. Inside the mainframe and the testhead [Agi00]

Figure 4.2 shows the contents of the mainframe and the testhead. The mainframe includes the mainframe control board, the mainframe boards and the power supply units. The

mainframe control board provides many resources shared by all the IO channels, such as:

- a data compression/decompression circuit for vector data
- a master clock generator
- a reference voltage generator
- a fiber optics interface

Additionally, it handles the communication between the workstation and the IO channel cards (via a pair of fiber optics cables).

Each Mainframe Board contains 8 Test Processor Modules. Each of them supports 2 channels. Inside the Test Processor Modules, there are SRAM Vector Memories, test processors, waveform and timing edge generators. Test processor provides the tester per- pin capability of the system. It uses the vector memory to store all the vectors, the sequencer commands, the timing sets, and the waveforms needed for a test.

Pin electronics cards, which are on the testhead, contain drive, receive, loads, relays and PMUs. PMU is used for voltage force/current measurements or current force/voltage measurements.

Figure 4.3 shows an illustration of per pin architecture. Besides, the vector memory and the test processor, you can see the timing generators and the waveform composer to form the signals. There are also driver, receiver and PMUs [Agi00].

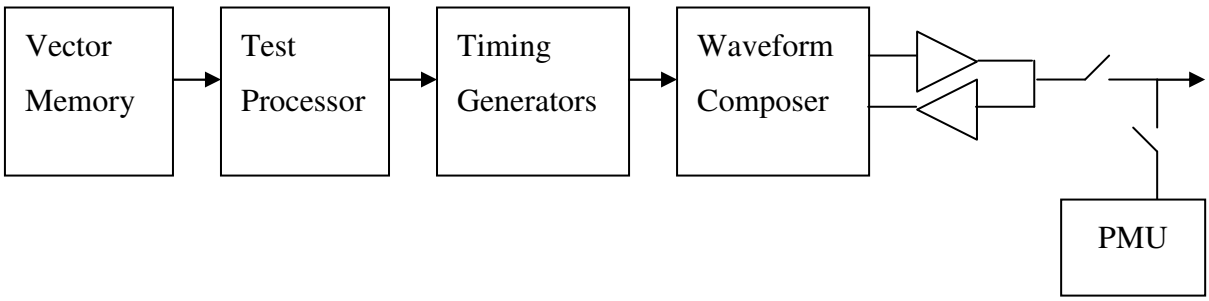


Figure 4.3. Per pin architecture [Agi00]

4.1.1. Steps to Test a Device

In order to start device testing, there are some steps that should be accomplished. They are shown in Figure 4.4.

First of all, a DUT board should be designed in order to interface the DUT pins to the tester channels. The best is to have a 50 Ω environment wiring on the DUT board to avoid the reflections.

The second step is the pin configuration setup. After deciding how the DUT board wiring allocates the tester channels to the DUT pins, this information should be entered into the software using the Pin Configuration editor. It means that, the tester channels that are connected to the DUT pins are defined in the pin configuration. The pin configuration addresses the DUT pins.

In the level setup, logic ‘0’ and logic ‘1’ levels for input signals and the thresholds of the comparator for the DUT outputs are defined. In the level setup, the supply voltage levels are

defined, as well.

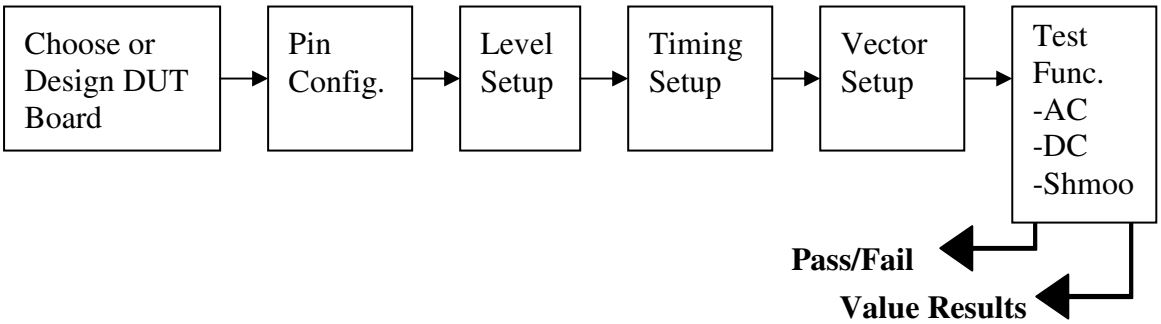


Figure 4.4. Steps to test a device [Agi00]

With the timing setup, the waveforms can be created to define the drive actions for the input pins and the compare actions for the output pins. The waveforms can be defined with and without data parameters. If a pin uses only fix values, such as a fixed pulse, it does not require any data from the vector setup.

The drive waveform actions that can be applied to a pin are:

- Fixed: '0' or '1'
- Parametric: D1 or D2 (should be defined in vector setup)
- Complementary: D1 and D2 (should be defined in vector setup)

For receive edges, one can choose between edge compare and window compare. For edge compare only one edge is used and for window compare, two edges are used. The first edge starts the compare action and the second edge determines the end of compare. A fixed action or a parametric action can be assigned for edge and window compare. The values can be:

- 0 : compare to 0
- 1 : compare to 1
- X: "do not care"
- I : compare to intermediate
- U: compare to unstable, only for window compare

If a waveform has a parametric drive or receive data, you need to supply this data in the vector setup. For such an action, the vector setup finally defines what the signal should look like. The vector pattern, which is used for testing the device's functional behavior, is defined in vector setup. Each pin can be set to a logic value to build a vector pattern.

After the vector setup is completed, the DUT is ready to be tested.

4.1.2. Test Functions

Figure 4.5 illustrates the basic principle of a digital testing. A test vector is applied to the inputs of the circuit. The response of the circuit is compared with the expected ones. The circuit is considered good if they match. Generation and evaluation of the test patterns is an important engineering issue. A certain amount of fault coverage should be reached. Generally, the test patterns are generated by Automatic Test Pattern Generation (ATPG) tools. There are

different types of test, such as functional test and structural test.

The objective of functional (or behavioral) test is to validate the correct operation of a system with respect to its functional specifications [ABF90]. It can be also called “design verification test”. It verifies the correct operation of all necessary functions. Timing specification can be also a part of functional test. In this case, the device is tested, whether it conducts the correct behavior to a timing standard.

The structural test verifies the topology of a chip and proves that all connections in the device are intact and all gate level truth tables are correct [Cro99]. Static stuck-at fault models can be used for structural testing. An input stimulus is applied to the primary inputs to toggle the suspected node to the opposite value. And then, this value must propagate to an observable output. This is accomplished by applying another set of input stimulus. If it differs from the desired value, a fault is detected. The delay fault model can be applied to verify the timing performance of the device.

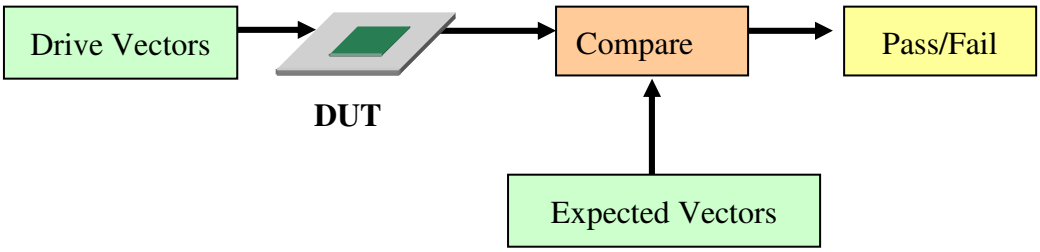


Figure 4.5. Principle of testing

4.2. PHEMOS 1000 System

PHEMOS 1000 is a semiconductor failure analysis tool that has a cooled charge coupled device (CCD) camera and two NIR laser diodes. Therefore, it is able to detect faint light emission coming from DUT by CCD camera as well as creating electrical fluctuations on the DUT by laser irradiation. It can be used for failure analysis applications in the design stage or for defective products, such as logic, power or memory device analysis. The general view of a PHEMOS 1000 system is seen in Figure 4.6. It includes four main parts:

- The black box covering the optics, CCD camera and the probe station
- Power supply unit
- PC that controls PHEMOS 1000 system and constructs the images
- Cooling unit that cools the CCD camera down to -54^0C

Besides, there is a Dynamic Analysis of Laser Stimulation (DALS) Kit, which makes dynamic analysis possible and an integrated amplifier for static applications to amplify small electrical changes for detection. In static mode, it is enough just to power up the device while it is fully activated in dynamic mode. Dynamic and static laser stimulation techniques are introduced in Chapter 6.

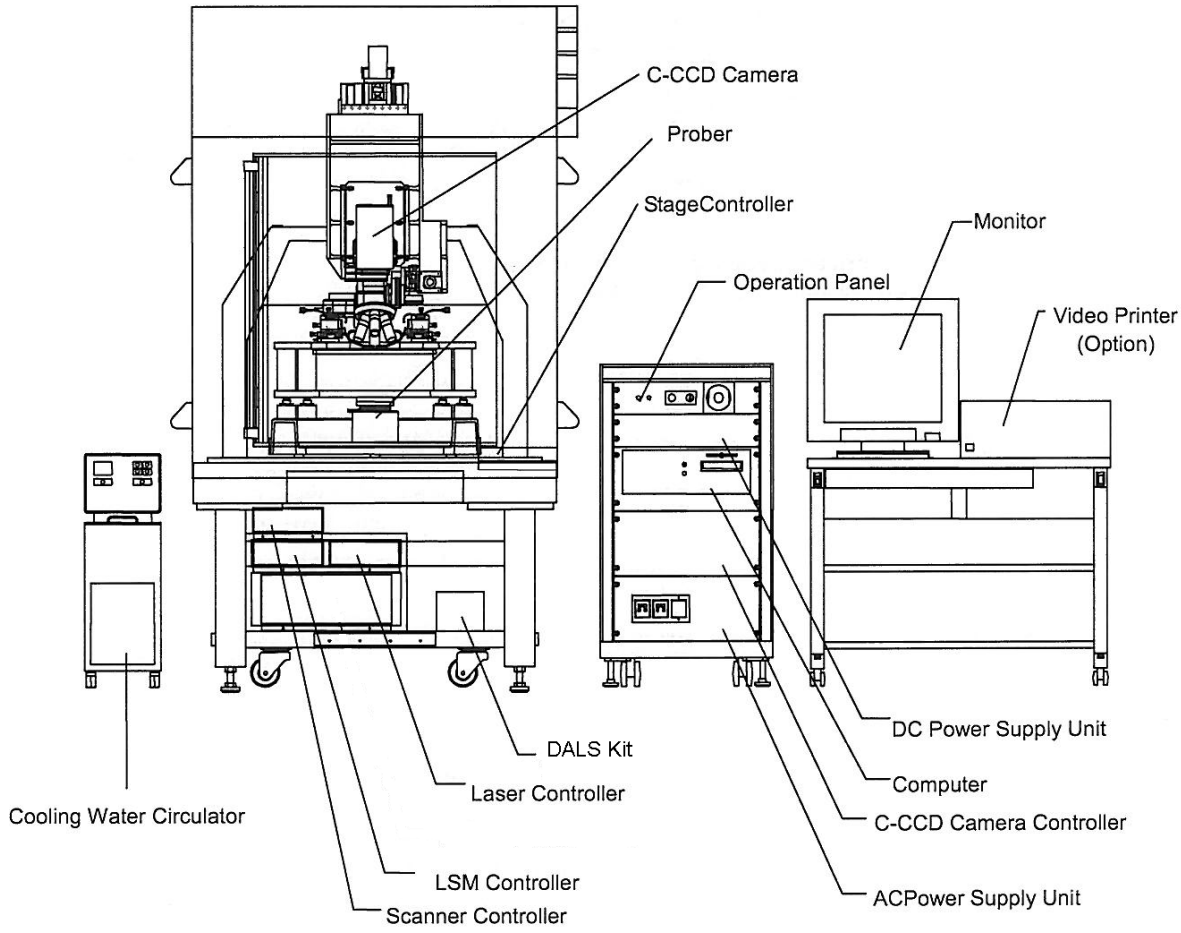


Figure 4.6. PHEMOS 1000 system [Ham]

The PHEMOS 1000 system is equipped with Süss PM 8 probe station. The DUT (a wafer or a packaged device) is placed on the probe station chuck. The probe station utilizes manipulators, which allow the precise positioning of thin needles on the surface of the wafer.

The PHEMOS 1000 system incorporates an IR-confocal laser microscope that allows obtaining a pattern image with high resolution up to 1024x1024 pixel size as well as stimulating the selected area on the device by laser beam.

4.2.1. Laser Stimulation System

The stimulation can be photoelectric or thermal depending on the energy of the laser beam. Regardless of the stimulation type, the principle of the laser scanning system is the same. There are two laser diodes installed in PHEMOS 1000 with two different laser wavelengths: 1.064 μm and 1.3 μm.

Both laser diodes can operate both in CW and externally triggered pulsed modes. 1.064 μm laser diode has two different pulsing options for pulsed mode:

1. **Impulse mode:** A laser pulse can be produced with a pulse width of 50 ns to 200 ns and 10 kHz repetition rate. The maximum laser power is 1W.
2. **Pulse mode:** Pulses with a minimum pulse width of 50ns and a repetition rate of

50 kHz can be generated. The maximum laser output power is 200mW.

Basic Laser Stimulation Setup

The basic laser stimulation set-up is seen in Figure 4.7. Main components of such a system are the laser scanning microscope, a laser source and an IC-driver. The IC-driver may be a tester, a power supply or a pulse generator depending on the application and the device. For static measurements, Hamamatsu amplifier and for dynamic measurements DALS kit may be utilized. A measurement unit, such as a parameter analyzer or an oscilloscope can be employed for some applications if it is needed.

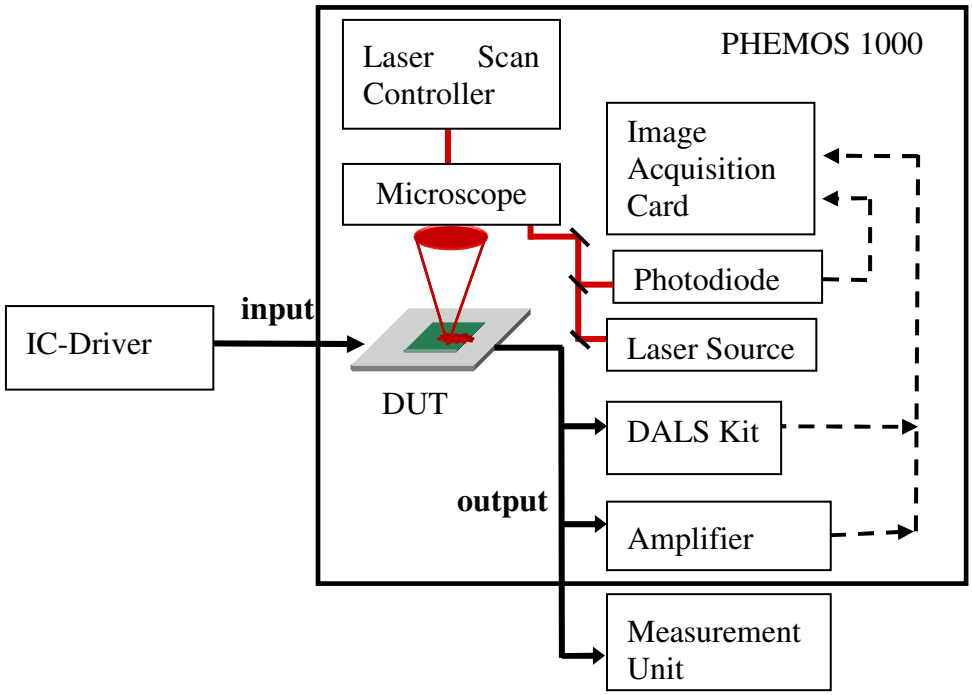


Figure 4.7. Basic laser stimulation setup

Laser Scanning Microscope (LSM)

The confocal laser scanning microscopy is utilized to scan the DUTs with the laser beam from the backside or the frontside. Confocal laser scanning microscopy is a technique for obtaining high-resolution optical images with depth selectivity [Paw06]. The key feature is its ability to acquire in-focus images from selected depths. Images are acquired pixel-by-pixel and reconstructed with a computer.

In a confocal laser scanning microscope, a laser beam is focused by an objective lens into a small focal volume within or on the surface of a specimen. The light reflected from the illuminated spot is re-collected by the objective lens. The intensity of the reflected light is then detected by a photo-detection device (photo-diode), transforming the light signal into an electrical signal that is recorded by a computer [FD07].

Five objective lenses are installed in the optics of PHEMOS 1000 whose numerical aperture (NA) values are given in Table 4.2. 0.5x macro lens cannot be used for laser stimulation purposes.

Table 4.2. Numerical Apertures of the objective lenses

Lens	**0.5x**	**5x**	**20x**	**50x**	**100x**
Numerical Aperture	0.74	0.14	0.40	0.76	0.50

The numerical aperture of a microscope objective is a measure of its ability to gather light and resolve fine specimen detail in the focal plane. The numerical aperture of an objective lens is defined by

$$NA = n \sin \theta \tag{4.1}$$

where n is the index of refraction of the medium in which the lens is working (1.0 for air, 1.33 for pure water, and up to 1.56 for oils), and θ is the half-angle of the maximum cone of light that can enter or exit the lens as shown in Figure 4.8.

A lens with a larger numerical aperture will be able to visualize finer details than a lens with a smaller numerical aperture. Lenses with larger numerical apertures also collect more light and will generally provide a brighter image.

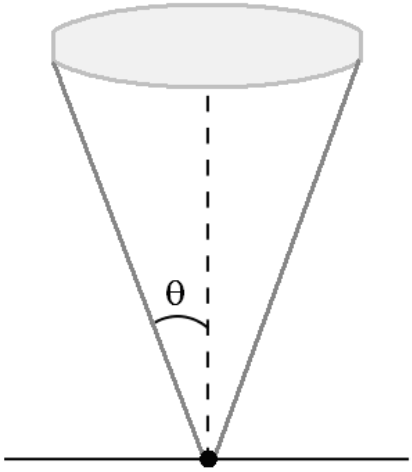


Figure 4.8. The objective lens and the specimen

The size of the scanning volume is determined by the spot size of the optical system because the image of the scanning laser is a three-dimensional diffraction pattern. The size of this diffraction pattern and the focal volume is defined by the numerical aperture of the system's objective lens and the wavelength of the laser used. The spot sizes are calculated according to $S = 1.22 \cdot \lambda / NA$ where S is the spot size and λ is the wavelength of the light (given in Table 4.3). There are several equations that have been derived to express the relationship between numerical aperture, wavelength and spot size. These equations are based upon a number of factors (including a variety of theoretical calculations made by optical physicists) to account for the behavior of objectives and condensers, and should not be considered an absolute value of any one general physical law.

Table 4.3. The spot sizes for two different laser wavelengths and objective lenses (LD=Laser Diode)

Objective		**5x**	**20x**	**50x**	**100x**
Spot size [μm]	**1.3 μm LD**	11.3	4.0	2.1	3.2
	1.064 μm LD	9.3	3.2	1.7	2.6

The maximum laser power outputs for 1.3 μm and 1.064μm laser diodes are 100 mW and 200 mW, respectively. But, it drops off after passing through the optical path. The laser powers after optical path (on the surface of the specimen) for each diode and different lenses are given in Table 4.4.

Table 4.4. The laser power for each objective lens on the surface of the specimen measured by Coherent # 1098292 Fieldmate laser power meter

Objective		**5x**	**20x**	**50x**	**100x**
Laser Power [mW]	**1.3 μm LD**	23.7	18	20.6	3
	1.064 μm LD	41	24.5	24	2.1

Mainly, 50x lens is used for the experiments in this work. Because, it has a high NA and a small spot size.

Scanning

The beam is scanned across the sample in the horizontal plane by using mirrors whose motions are controlled by the scan controller unit. As the laser scans over the area of interest, a whole image is obtained pixel-by-pixel, whereas the brightness of a resulting image pixel corresponds to the relative intensity of detected light. The reflected light intensity depends on the material that is illuminated. Metal lines reflect more light.

The scan speed can be varied. Slower scans provide a better signal-to-noise ratio, resulting in better contrast and higher resolution.

PHEMOS 1000 system employs a laser scan controller that synchronizes the position of the laser beam with the electrical variations of the device under test.

Scan operation is performed pixel by pixel. For backside analysis, the scanning starts at the top left corner of the selected area and moves vertically to the bottom as it is shown in Figure 4.9. When one line is finished, it moves to the next line at the top. The line signal repeats at each single column and frame signal repeats when a whole image is acquired.

The scanned area is larger than the selected area. It means that some scanned pixels are not used for image acquisition. These pixels are called dummy pixels. The voltage levels of the line and the frame signal are logic high if the scanner is on a dummy pixel, otherwise they stay at logic low.

There are various options for scanning of the devices. These modes are:

- Normal mode: Full two-dimensional scanning. The available options are: 512x512 or 1024x1024 pixel size.
- Slit V: Full vertical – narrow width scan. For example, 64x512, 32x512 pixel size.
- Slit H: Full horizontal – narrow vertical width scan. For example, 512x64 or

512x128 pixel size.

- Area: A small size square area. For example, 128x128, 64x64 or 32x32.
- Line H: Only one line scan in the horizontal direction (512x1).
- Line V: Only one line scan in the vertical direction (1x512).
- Point: The laser beam is placed on one pixel. It does not move.

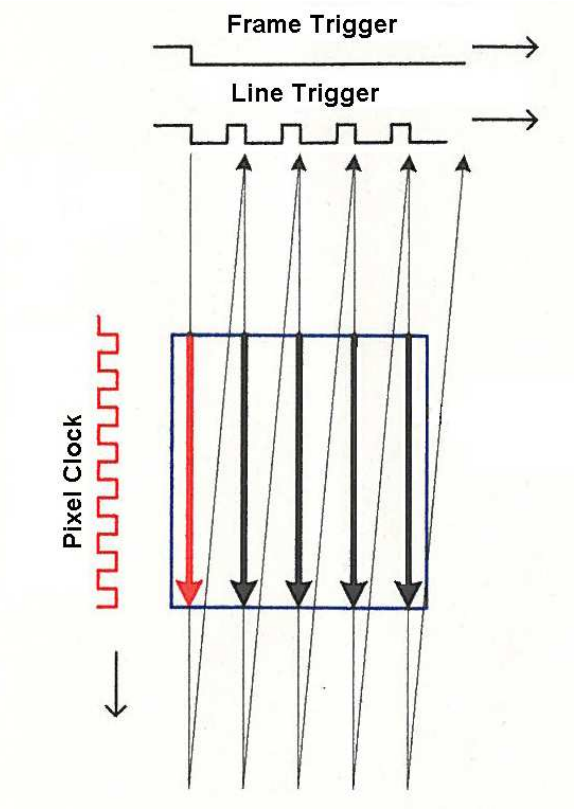


Figure 4.9. The illustration of the scanning function [Ham]

Image Processing

While reflected laser light image is obtained, the photodiode detects the reflected light from the DUT at each pixel and the electrical signal at the output of the photodiode is converted into a digital signal by an A/D converter. This digital signal is used for image processing by PHEMOS 1000 software.

The laser stimulation image is produced in a similar way. In this case, the photodiode is not in use. However, depending on the application, the Hamamatsu amplifier or DALS kit is utilized in order to process the output signal from the DUT and to grab a laser stimulation image.

This processed electrical signal variation (voltage or current) is converted into a numerical value pixel by pixel by the image acquisition card of the PHEMOS system. These numerical values are saved together with the (x,y) position information of the laser beam. Later on, Hamamatsu software constructs an image by using both of intensity and position data. The reflected laser light from the DUT of the same area is also collected (visual image of the DUT) and these two images are superimposed for localization.

Hamamatsu Amplifier

It is used for static analysis. Electrical changes caused by laser stimulation are usually very weak and therefore need to be amplified before it can be used for imaging purposes. The main function of Hamamatsu amplifier is to amplify small electrical changes like current or voltage. The Phemos1000 system has the signal-amplifying unit consisting of voltage and current sources and amplifier. This means, that the DUT can be electrically driven by current and voltage modes. Main characteristics of OBIRCH amplifier are presented in Table 4.5.

Table 4.5. Basic properties of the Hamamatsu amplifier

Amplification Type	Voltage Range	Current	Sensitivity	Amplification gain	Bandwidth
Constant Voltage	10 mV to 10 V	<100 mA	10 nA	10^6 V/A	100 kHz
Constant Current	10 mV to 10 V	<100 mA	50 μV	10^2 V/V	100 kHz

If a static analysis is performed, the DUT output is connected to the Hamamatsu amplifier and it amplifies very small electrical changes like voltage or current. If a current change at the output of a device is subject to investigation, the output current is converted into voltage and amplified by the Hamamatsu amplifier. The output voltage of the amplifier serves as input for the data processing to the image acquisition card.

DALS Kit

A DALS Kit is an interface that is implemented to PHEMOS 1000 in order to realize dynamic laser stimulation on the ICs. It makes the connection to a tester possible and it simplifies the dynamic analysis of the DUTs. The construction of the synchronization of tester with PHEMOS 1000 and feed-through of PASS/FAIL information from tester to PHEMOS 1000 are accomplished by the DALS Kit.

The scan function can be controlled by producing synchronization signals by tester and feed them to PHEMOS 1000 through DALS Kit connectors.

It has eight connectors. Four of them assigned to manage the synchronization. These are:

- ***Pixel Clock*:** At each move from one pixel to the next one, the system produces a pulse. The frequency of the signal depends on the scan speed.
- ***Line Trigger*:** When one line scan in vertical direction is over, the laser beam moves to the next line. At each start of a line scan, a pulse is produced in *Line Trigger* output signal.
- ***Frame Trigger*:** When the scan of the selected area (frame) starts, *Frame Trigger* signal is pulled from logic '1' to logic '0' and stays at this level until the scan of one frame is over.
- ***Point Pixel:*** This is a connector for outputting point pixel trigger signal of scanner in single point mode. The position of the laser beam is set in the software by the user. The point pixel signal produces a pulse when scanner reaches to this set point. If the point to be illuminated is set to (0,0) pixel where the scanning starts in normal mode, the *Point Pixel* is the same as *Frame Trigger* signal.

The *Pixel Clock*, *Line Trigger* and *Frame Trigger* signals and their relation to the scanned area are seen in Figure 4.9.

A larger area than selected is scanned. These dummy pixels are located around the selected area. If the laser is on a dummy pixel, either *Line Trigger* or *Frame Trigger* signals or both of them are at logic '1' level. The relation between *Line Trigger* and *Pixel Clock* and *Frame Trigger* and *Line Trigger* are shown in Figure 4.10 and 4.11, respectively.

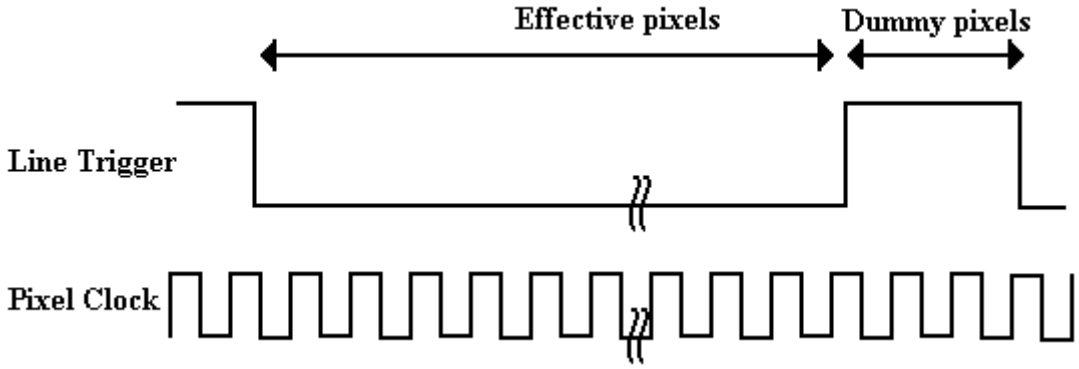


Figure 4.10. *Line Trigger* and *Pixel Clock* signals during laser scan

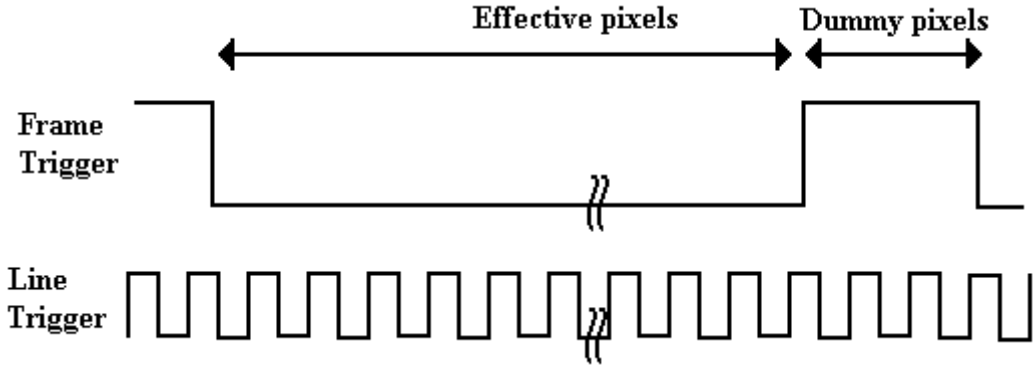


Figure 4.11. *Frame Trigger* and *Line Trigger* signals during laser scan

The other four connectors are for construction of the image, which characterizes the DUT upon laser stimulation (P/F signal from tester or the output of the DUT). These are:

Tester Signal: This is a connecter to evaluate the DUT output signal through an analog integrator. The analog integrator integrates the signal applied to its input connector. Therefore, it is able to detect electrical changes in the signal, such as amplitude, frequency, or duty cycle. For example, if the frequency decreases at a certain pixel, the output of the analog integrator will decrease, too. As a result, the related pixel will be darker with respect to gray background.

The *Tester Signal* connection can be also used to input the pass/fail signal coming from the tester. If a pulse is recognized at this input connector within a pixel clock period, the related image data will be logic high (red) for this pixel. If not, the image data will be logic low (green). A channel at tester can be allocated for pass/fail data. If the tester is programmed to produce a pulse in case of a fail, the sites where a change from pass to fail occurs can be detected. Therefore, it can detect the locations within a DUT where changes from pass to fail

or from fail to pass occur.

One of these two different options of DALS Kit functions can be selected through PHEMOS 1000 software.

Loop Trigger: It is used, when the scan is made externally. The synchronization signal produced by the tester is applied to this connector. At each pulse applied to *Loop Trigger*, the laser beam moves to the next pixel.

Tester Trigger: This connector can be used optionally for pass/fail evaluation. It can be used to latch the data applied to the *Tester Signal* connector. If the *Tester Trigger* is detected within a pixel and the *Tester Signal* is logic high, image data will be also logic high (red). If *Tester Signal* is at logic low or no tester trigger detected within a pixel, image data will be logic low.

Modulation Trigger: The *Modulation Trigger* signal, which pulses the 1.3 µm laser diode is applied to this connector.

4.2.2. Photon Emission Microscopy

Photon emission microscopy is an extensively used technique, in order to detect very faint light, generated through an electroluminescence process.

Basic Photon Emission Setup

The photon emission microscope consists of optical microscope, image intensifier, CCD camera and image processor that is shown in Figure 4.12. The conventional optical microscope is used to obtain the normal reflection image of the DUT and in a highly sensitive mode, the photon emission of the DUT. Image processing of the signal is carried out with a frame grabber card and a personal computer [Chi00]. After the photon emission image is obtained, it is overlaid onto the reflected light image. Optionally, reflected laser light image can be used in our system. For backside analysis, a Si detector is not well suited for the transmitted light that shifts to the infrared due to absorption of bulk silicon. Highly sensitive photon emission microscopes for backside analysis employ detectors of a smaller bandgap material.

When a DUT is placed under a photon emission microscope, the emission from the DUT is detected and stored to produce an electronic image. The reflected light from the DUT is also collected and emission image is overlaid on reflected light image. As a result, light emitting spots are localized.

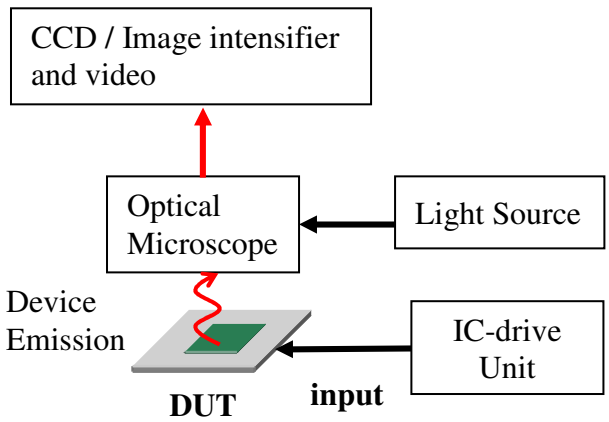


Figure 4.12. A typical photon emission microscope basic setup

CCD camera

The structure of a CCD is based upon a MOS capacitor. The gate electrodes are made of a highly conductive material such as metal or polysilicon. The oxide layer is SiO_2 and the channel is a semiconductor. Groups of electrodes form a pixel, in our case 3 electrodes. As shown in Figure 4.13, a potential well is created beneath the electrode, with a positive voltage at P2 terminal.

The photons emitted from the DUT generate electron-hole pairs in space charge regions of CCD. These minority carriers are separated by the electric field and collected by the most attractive electrodes. The collected charge is proportional to the intensity of the radiation and the integration time. A two-dimensional array, used in video and still cameras, captures a two-dimensional picture corresponding to the scene projected onto the focal plane of the sensor. By adjusting electrode voltages in the appropriate time sequences, the signal charge is sequentially transferred towards output. The last capacitor in the array dumps its charge into a charge amplifier, which converts the charge into a voltage. The output voltage amplitude is proportional to the intensity of the imaged scene. Basically, A CCD camera is a photon-charge-voltage converter.

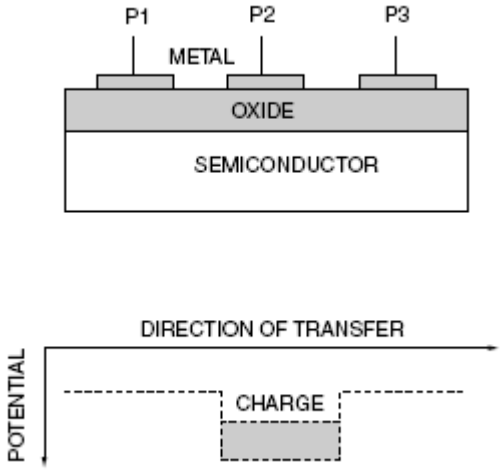


Figure 4.13. CCD area image sensor and operating principle [Ham03]

4.3. Lock-in Amplifier

A lock-in amplifier provides a DC output proportional to the AC signal under investigation. AC to DC conversion is performed by a PSD (Phase Sensitive Detector), which is also known as demodulator or mixer. It rectifies only the signal of interest while suppressing the effect of noise or interfering components, which may accompany that signal. In order to ensure that the instrument will track any changes in the signal of interest, the detector must be supplied with a reference signal, which has the same frequency as the input signal and with a fixed phase shift to that of the signal. The detector operates by multiplying signal of interest and reference signal together. If the reference signal amplitude is maintained at a fixed value, and the reference phase is adjusted to ensure a relative phase-shift of zero degrees, then the output of the Lock-in Amplifier will be equal to the mean value of the input amplitude.

The block diagram of a typical lock-in amplifier is shown in Figure 4.14. In the signal channel, the input signal, including noise, is amplified by an adjustable-gain, AC-coupled amplifier, in order to match it more closely to the optimum input signal range of the PSD. The

performance of the PSD is usually improved if the bandwidth of the input signal is filtered with a bandpass filter. That can be for example a bandpass filter, which is centered at reference frequency. In the reference channel, reference signal is passed through a phase shifter, which is used to compensate for phase differences that may have been introduced between the signal and reference inputs by the experiment, before being applied to the PSD. Those two signals that are coming from the input channel and reference channel are multiplied by PSD. There are currently three common methods of implementing the PSD, these being the use of an Analog Multiplier, a Digital Switch or a Digital Multiplier. In real applications the signal will be accompanied by noise. This noise, which has no fixed frequency or phase relationship to the reference, is also multiplied by the reference signal in the demodulator, but does not result in any change to the mean DC level. Noise components at frequencies very close to that of the reference do result in demodulator outputs at very low frequencies, but by setting the low-pass filter to a sufficiently low cut-off frequency, these can be rejected. Hence the combination of a demodulator and low-pass output filter allows signals to be measured even when accompanied by significant noise [Sig00].

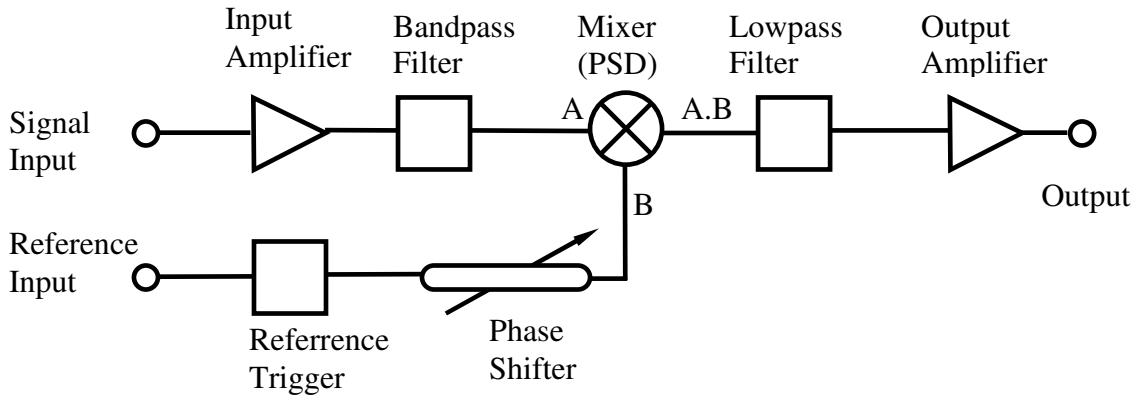


Figure 4.14. The block diagram of a lock-in amplifier [Sig00]

5. The Properties of the Optical Beam and the Interaction of the Laser Light with the Semiconductor

Laser stimulation and photon emission techniques makes use of the transparency property of Si for infra-red light. Light as an electromagnetic radiation in general and laser light properties in particular as well as light – matter interaction is presented in this chapter.

5.1. Light

Light is an electromagnetic radiation, which exhibits both wave-like and particle-like properties. This concept is called wave-particle duality.

According to wave-particle duality principle, light waves could be treated as particles, which are referred to as photons. An important aspect of light is its wavelength or frequency, which also determines the energy of the photons. The energy of a photon is:

$$E = h \cdot f \tag{5.1}$$

where h is Planck's constant and f is the frequency.

5.2. Laser Light

Laser stands for **L**ight **A**mplification by **S**timulated **E**mission of **R**adiation. It is a highly monochromatic (small bandwidth of wavelengths), coherent (temporal and spatial) and collimated (highly directional beams) light.

In order to analyze the laser light propagation and related optical systems, gaussian curve approaches are widely used. In optics, a gaussian beam is the beam of electromagnetic radiation whose electric field and intensity distributions are explained by gaussian functions. When refracted by a lens, a gaussian beam is transferred into another gaussian beam; therefore it is very convenient to utilize this model.

For a gaussian beam propagating in medium (see Figure 5.1), the spot size $w(z)$ will be at a minimum value (w_0) at one point along the beam axis, which is called beam waist. As the beam goes off the beam waist, the intensity of the beam gets smaller while the spot size gets larger. The distance, which the beam travels from the waist until the beam diameter increases by $\sqrt{2}$, or until the beam area doubles is known as Rayleigh range (z_R).

All the important parameters of the gaussian beam can be related to the waist spot size and Rayleigh range.

$$w(z) = w_0 \sqrt{1 + \left(\frac{z}{z_R} \right)^2} \tag{5.2}$$

$$z_R = \frac{\pi \cdot {w_0}^2}{\lambda} \tag{5.3}$$

At a distance from the waist equal to the Rayleigh range, the width of the beam is

$$w(z_R) = w_0 \sqrt{2} \tag{5.4}$$

R(z) is the radius of curvature of the wavefronts comprising the beam.

$$R(z) = z + \frac{{z_R}^2}{z} \tag{5.5}$$

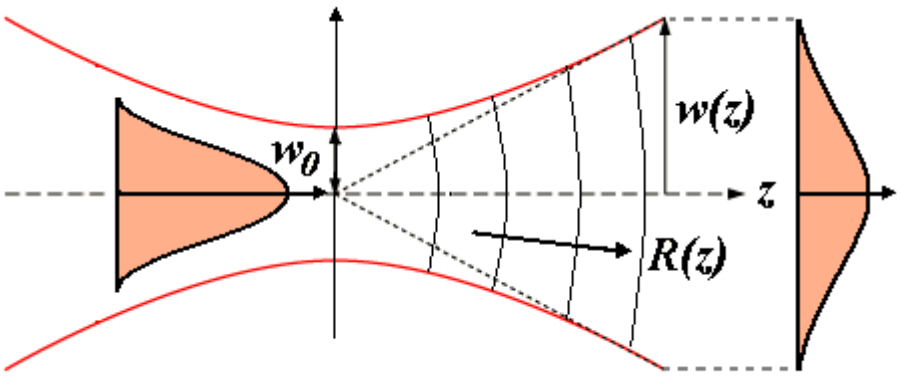


Figure 5.1. Illustration of a gaussian beam diverging away from its waist [Sie86]

5.3. Interaction of Light and Matter

When a light wave is incident on a plane, the light can be reflected and refracted. The refracted part may be transmitted through the medium or absorbed by the medium. Absorption of electromagnetic radiation is the way by which the energy of a photon is taken up by matter. Thus, the electromagnetic energy is transformed to other forms of energy, for example to heat.

When a photon hits a piece of Si, there are a few possibilities that can happen in the means of Si-photon interactions.

1. The photon can reflect off the surface
2. If the photon energy is less than the bandgap of Si (1.12 eV), the photon can be absorbed by the Si and the photon energy is converted into heat via lattice vibrations – called phonons.
3. If the photon energy is greater than the bandgap energy of Si, the photon is absorbed by the Si, but in this case the photon energy is given to an electron in the valance band which excites it into conduction band. Such a process generates electron-hole pairs and creates excess carriers.

Therefore, for the photon energies higher than 1.12 eV (wavelengths lower than 1107 nm) electron-hole pair generation is likely to occur. The penetration depth in Si and the generation rate is a strong function of absorption coefficient.

5.4. Optical Absorption

The absorption coefficient determines how deep light or any electromagnetic radiation penetrates into a material before it is absorbed. In other words, it is the relative number of photons absorbed per unit distance. It depends strongly on the material and on the energy of the light. Semiconductor materials have a sharp edge in their absorption coefficient, since light, which has energy below the band gap energy, does not have a sufficient energy to raise an electron across the band gap. The absorption coefficient for several semiconductors materials is shown in Figure 5.2 and for intrinsic Si versus photon energy in Figure 5.3 and versus photon wavelength (in vacuum) in Figure 5.4. The absorption coefficient increases

very rapidly for the energies higher than the band gap energy of the material. If the absorption coefficient is large, the photons are absorbed over a relatively short distance. The absorption coefficients are very small for the photon energies lower than the band gap energy of the semiconductor, so the semiconductor appears transparent for this energy range.

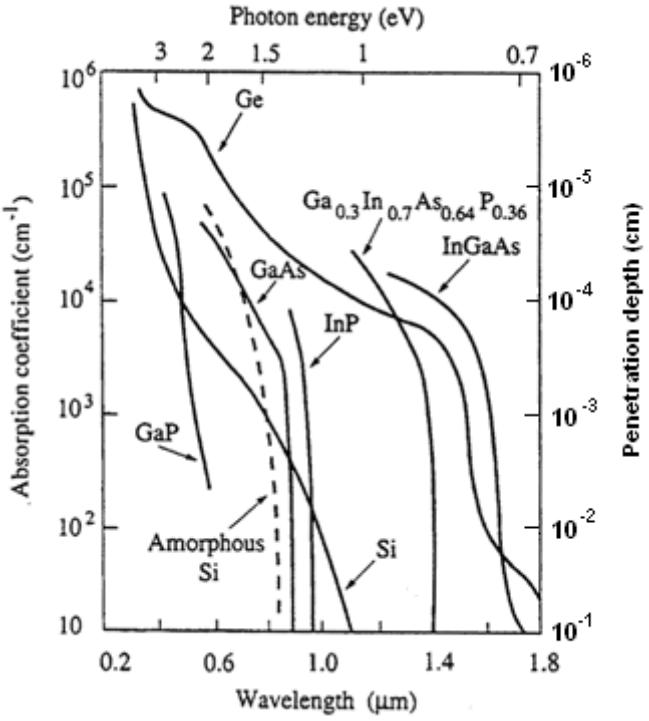


Figure 5.2. Absorption coefficient as a function of wavelength and photon energy for several semiconductors [Shur90]

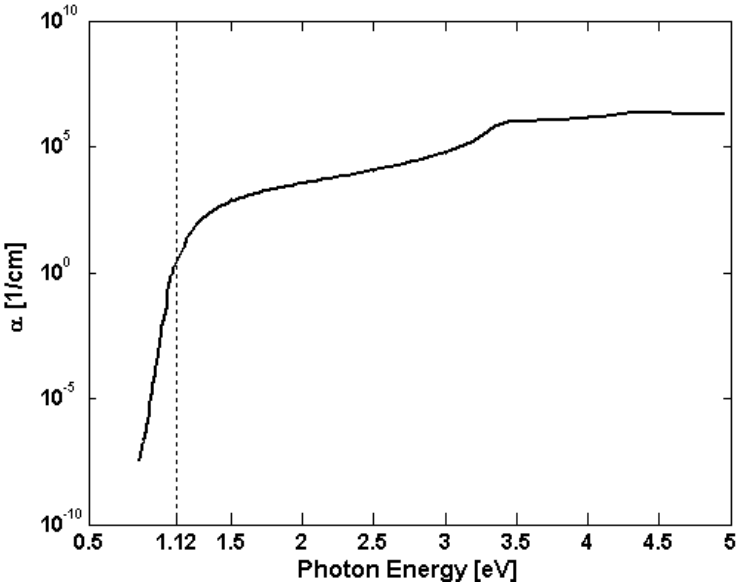


Figure 5.3. Absorption coefficient of intrinsic silicon as a function of the photon energy on semi-logarithmic scale (source: [Ble86])

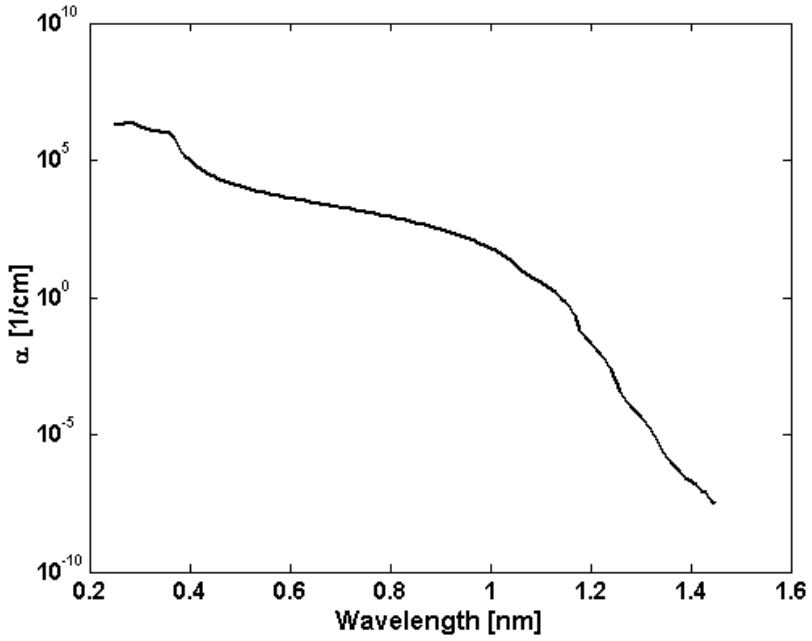


Figure 5.4. Absorption coefficient of silicon as a function of the wavelength on semi-logarithmic scale (source: [Ble86])

The relation between absorption coefficient and photon flux is given by

$$\frac{dI_v(x)}{dx} = -\alpha \cdot I_v(x) \tag{5.6}$$

where $I_v(x)$ is the intensity of the light [energy/cm^2], α is the absorption coefficient [cm^{-1}] and x [cm] is the distance from the surface. If the initial photon flux is given as I_{v0}, solving the differential equation (Equation 5.6), leads to

$$I_v(x) = I_{v0} e^{-\alpha x} \tag{5.7}$$

which denotes that the intensity of the light inside a material falls off exponentially from the surface.

Penetration depth is defined as the depth at which the intensity of the radiation inside the material falls to 1/e (about 37%) of the original value at the surface. If δ denotes the penetration depth, we have,

$$\delta = \frac{1}{\alpha} \tag{5.8}$$

5.4.1. Carrier Generation Mechanism

Excess carrier generation mechanism plays an important role in the photoelectric laser stimulation technique.

Electrons in the valance band may be excited to the conduction band by an external stimulus (i.e. by photons), which produces excess carriers in the semiconductor. This phenomenon is called inter band absorption. The transition may be direct (with conservation of momentum energy) or indirect (with change in electron momentum because of emission or absorption of phonon or heat particle) – see Figure 5.5 and 5.6.

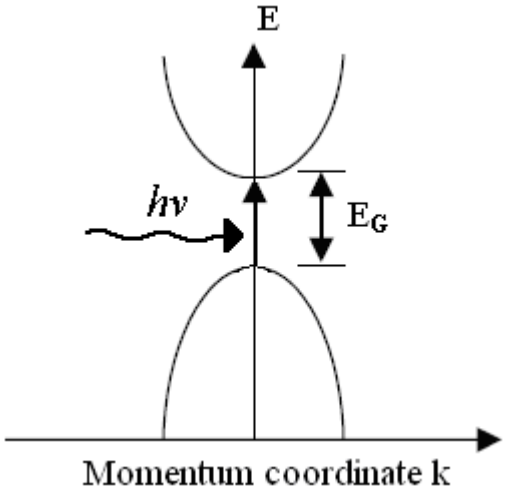


Figure 5.5. Direct photo-excitation

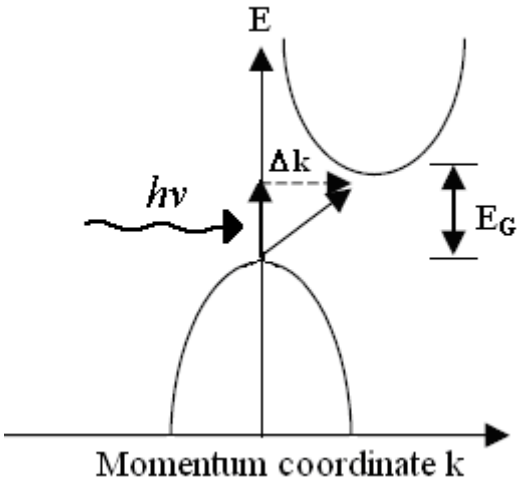


Figure 5.6. Indirect photo-excitation

In direct transition, the generation of an electron-hole pair is triggered by a photon, which transfers its energy to a valence band electron. The excited electron moves to the conduction band leaving a hole behind. The photon energy has to be at least of the magnitude of the band-gap energy for this process. In indirect transition, a momentum change (Δk) is required. Since a photon carries little momentum, it cannot be provided by the photon itself. Momentum transfer is acquired via lattice vibrations (phonons).

Since the conservation of momentum can only be satisfied with the interaction of phonons for indirect semiconductors (Si, Ge), the process is less probable in comparison to direct semiconductors (GaAs). This is the reason, why the slope of the absorption coefficient versus the energy for indirect semiconductors is not as steep as the slope of direct semiconductors (see Figure 5.2).

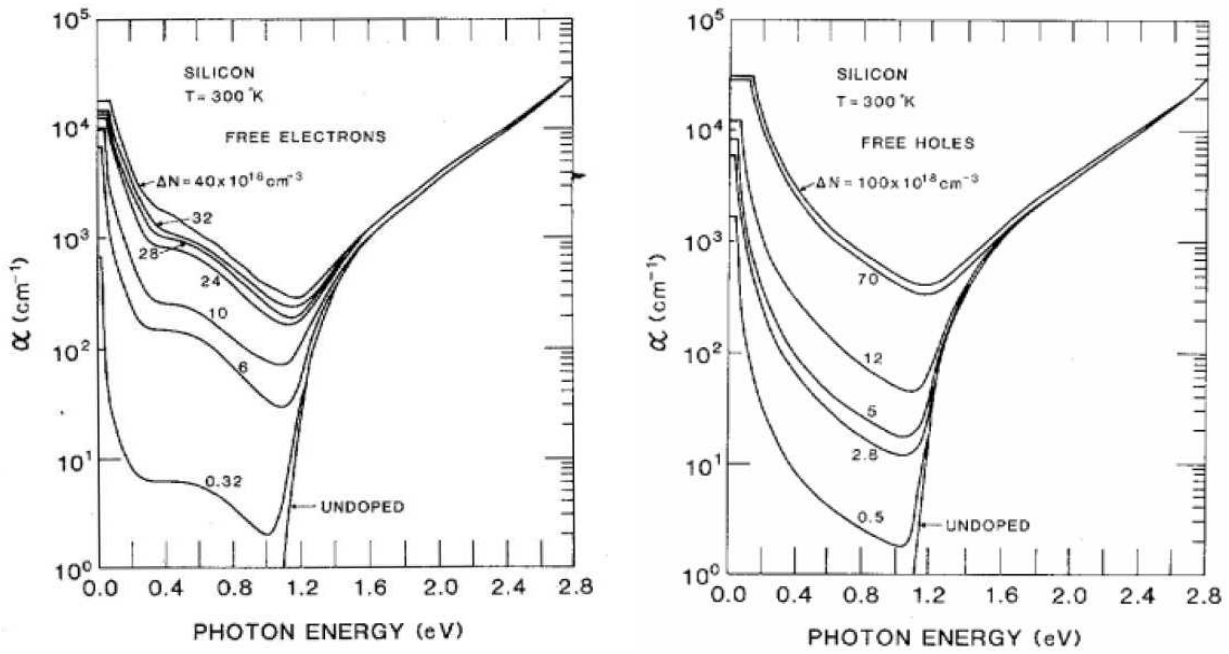
5.4.2. Free Carrier Absorption

Free carrier absorption mechanism involves the absorption of a photon by the interaction of a free carrier within a band, which is consequently raised to a higher energy. The free carrier density in a semiconductor can be changed either by injection of charge carriers into an undoped sample or by the removal of free carriers from a doped sample. Free carriers can also be injected depending on the voltage that is applied to the device.

The change in the absorption coefficient introduced by the free carriers can be expressed as [SB87]

$$\Delta\alpha = \frac{\lambda^2 q^3}{4\pi^2 c^3 \varepsilon_0} \cdot \frac{\Delta N}{n m^2 \mu} \tag{5.9}$$

where ΔN is the change in the concentration of charge carriers, n is the refractive index of the semiconductor, q is the electronic charge, ε_0 is the permittivity of free space, m is the conductivity effective mass of the electrons or holes and μ is the electron or hole mobility. The absorption coefficients versus photon energy graphs for different doping concentrations are seen in Figure 5.7. According to these graphs, high concentration of free electrons and holes increases the absorption coefficient [SB87].



(a) (b)

Figure 5.7. Optical absorption spectra of Si showing the influence of various concentrations of (a) free electrons and (b) free holes [SB87]

5.5. Recombination of electron-hole pairs

Carrier recombination mechanisms in semiconductors are shown in Figure 5.8. Direct band-to-band recombination is a radiant process. An electron loses its energy and moves from conduction band to the valance band releasing its energy in the form of photon. In Shockley-

Read-Hall recombination (indirect), an electron from the conduction band is trapped by a defect or an impurity in the lattice and this electron moves to the valance band and recombines with a hole. The excess energy during recombination is transferred to the crystal lattice (phonon). In Auger processes, three carriers are involved. The transition of an electron from the conduction band to the valance band is made possible by transfer of the energy to another free electron or hole. This process is called Auger recombination. Auger process is a non-radiant process. The band-to-band and trap assisted recombination mechanisms can be better understood by the illustration of the transitions in Energy-Momentum diagram as shown in Figure 5.9. The transition of the carrier should conserve momentum. It is accomplished by phonons. For an indirect semiconductor like Si, band-to-band recombination is less probable than the trap assisted recombination.

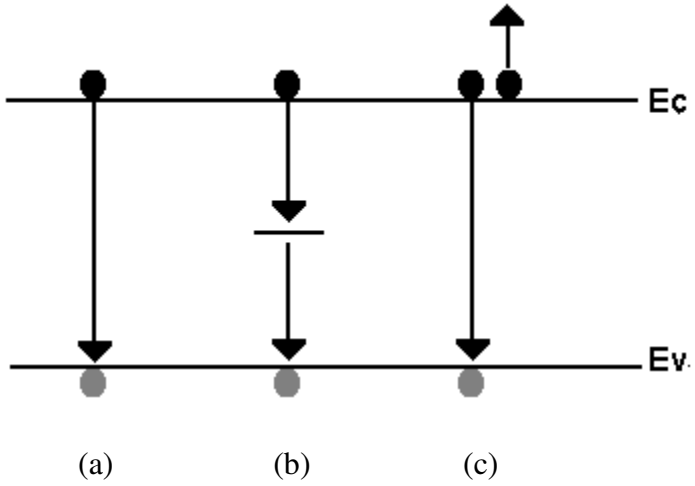


Figure 5.8. Carrier recombination mechanisms in semiconductors (a) Band-to-band recombination (b) Shockley-Read-Hall recombination (c) Auger recombination

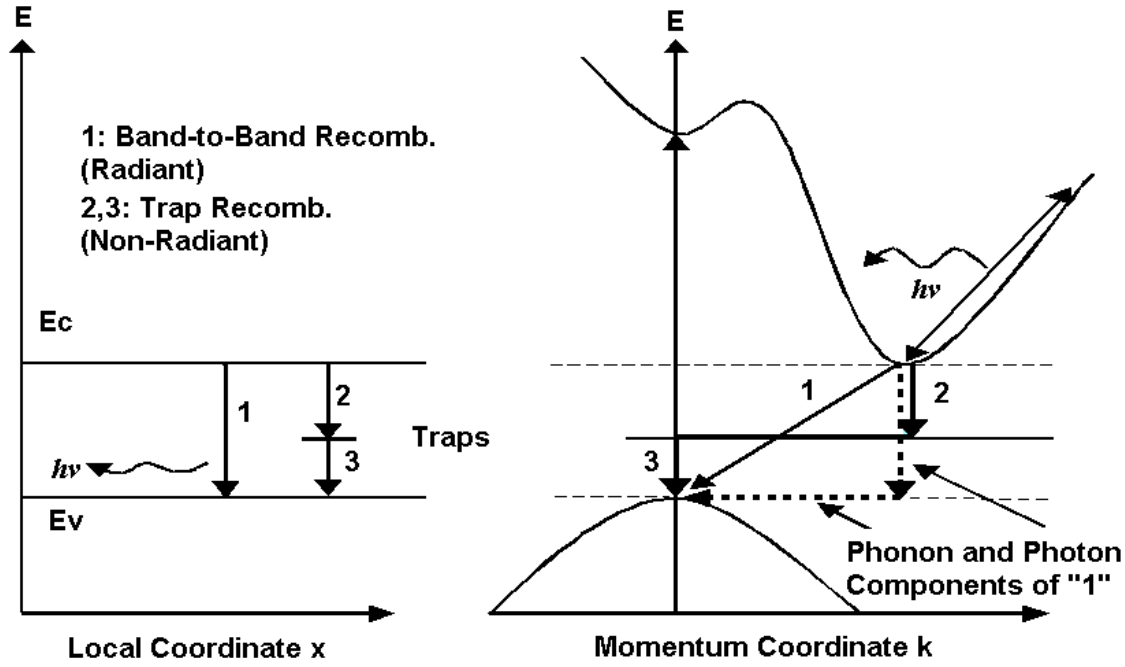


Figure 5.9. Si band structure and illustration of recombination mechanisms in local coordinate and momentum coordinate [Boi04]

5.6. Generation Rate

Generation rate is defined as the rate at which the electron-hole pairs are created. Assuming that the absorption of each photon causes the generation of one electron-hole pair, the generation rate G at a distance of x from the semiconductor surface is determined by

$$G = \frac{\alpha(\lambda)}{E_{ph}} \cdot \frac{P_{opt}}{A} \cdot e^{-\alpha(\lambda)x} = \alpha(\lambda)\Phi_0(\lambda)e^{-\alpha(\lambda)x} \tag{5.10}$$

which is in units of 1/cm^{3}s. α is the photon wavelength dependent absorption coefficient, P_{opt} is the laser power, E_{ph} is the photon energy, A is the laser beam area and Φ_0 is the photon flux at the surface (photons/cm^{2}s). The reflection of the photons from the surface can be also taken into account. Let $R(\lambda)$ be the fraction of the photons reflected from the surface, the Equation 5.10 can be rewritten as

$$G = \alpha(\lambda) \cdot \Phi_0(\lambda)\left[1 - R(\lambda)\right] \cdot e^{-\alpha(\lambda) \cdot x} \tag{5.11}$$

5.7. Recombination Rate

When the injected carriers are fewer than the majority carriers, the recombination rate for an n-type semiconductor can be given by

$$R = \frac{p_n - p_{n0}}{\tau_p} \tag{5.12}$$

where p_{n0} is the equilibrium minority carrier concentration, $p_n = \Delta p + p_{n0}$ and τ_p is the minority carrier life-time [Sze98].

5.8. Characteristics of Excess Carriers

There are some basic equations, which explain the static and dynamic behavior of excess carriers in the presence of external perturbation by photons or electric fields. The net flow of electron and holes will create a current in the semiconductor. This is called transport and it has two basic components: drift and diffusion.

Carrier drift is caused by the electric field applied to the semiconductor. Since both electrons and holes contribute to the drift current, the total drift current density can be written as follows:

$$J_{drift} = q\mu_n nE + q\mu_p pE \tag{5.13}$$

where n is the electron and p is the hole concentration, E is the electric field and μ_n and μ_p are electron and hole mobilities, respectively.

The second mechanism that induces current is diffusion. Diffusion is a net transport of carriers from a region of higher concentration to one of lower concentration by random motion, which produces a net flux of carriers flowing. We may write for a one dimensional case in the form of

$$J_{dif} = qD_n \frac{\partial n}{\partial x} - qD_p \frac{\partial p}{\partial x} \tag{5.14}$$

where D_n is the electron and D_p is the hole diffusion coefficients.

The total current density is the sum of these four components. For one dimensional case,

$$J_{tot} = q\mu_n nE + q\mu_p pE + qD_n \frac{\partial n}{\partial x} - qD_p \frac{\partial p}{\partial x} \tag{5.15}$$

The continuity equation describes the transport of carriers. A change in carrier density over time is due to the difference between the incoming and outgoing flux of carriers plus the generation and minus the recombination, which is given below:

$$\begin{aligned} \frac{\partial n}{\partial t} &= G_n - R_n + \frac{1}{q}\nabla \cdot J_n \\ \frac{\partial p}{\partial t} &= G_p - R_p - \frac{1}{q}\nabla \cdot J_p \end{aligned} \tag{5.16}$$

For one-dimensional case under low injection and low electric field condition, the continuity equation becomes [Sze98]

$$\frac{\partial n_p}{\partial t} = G_n - \frac{n_p - n_{p0}}{\tau_n} + n_p\mu_n \frac{\partial E}{\partial x} + \mu_n E \frac{\partial n_p}{\partial x} + D_n \frac{\partial^2 n_p}{\partial x^2} \qquad \text{p-type Si} \tag{5.17}$$

$$\frac{\partial p_n}{\partial t} = G_p - \frac{p_n - p_{n0}}{\tau_p} - p_n\mu_p \frac{\partial E}{\partial x} - \mu_p E \frac{\partial p_n}{\partial x} + D_p \frac{\partial^2 p_n}{\partial x^2} \qquad \text{n-type Si}$$

5.9. Charge Separation

Induced excess carriers, which are created by a localized light source on a semiconductor, will recombine. However, if the carriers are generated within a space charge region or in the distance of diffusion length from the space charge region, the carriers will be swept out of the depletion region by the electric field and the electrons will drift to the n-region and the holes to the p-region. The illustration of charge separation is seen in Figure 5.10. If the recombination is neglected, the induced photo current within the depletion region is given by [Nea03]

$$J = q\int_0^W G dx \tag{5.18}$$

where G is the generation rate of the excess carriers and W is the depletion region thickness. If G is constant throughout the depletion region, the equation becomes

$$J = qGW \tag{5.19}$$

Excess carriers in the distance of diffusion length from n and p regions are also collected. If we assume that there is no electric field and G is constant, the photocurrent density for the n-region and p region can be written as

$$\begin{aligned} J_n &= qGL_n \\ J_p &= qGL_p \end{aligned} \tag{5.20}$$

where L_n and L_p are the diffusion lengths for electrons and holes, respectively. The total photocurrent is given by

$$J_{tot} = qGW + qGL_n + qGL_p \tag{5.21}$$

In case of a MOSFET, the generated electron-hole pairs in the space charge region (SCR) under the channel and in the SCR of the reversed biased drain/bulk and source/bulk junction

will be collected by the transistor.

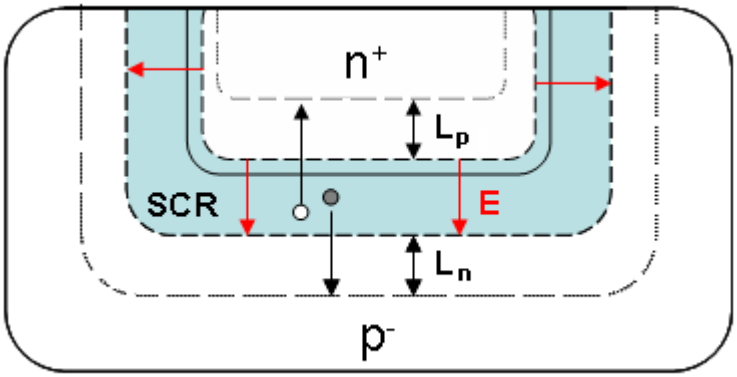


Figure 5.10. Charge separation in a pn junction

5.10. Laser Pulse

If we assume that a laser pulse generates excess carriers in an n-type semiconductor, the transport equation for the minority carriers after the pulse is given by setting *G=0* and *dE/dx=0*:

$$\frac{\partial p_n}{\partial t} = -\frac{p_n - p_{n0}}{\tau_p} - \mu_p E \frac{\partial p_n}{\partial x} + D_p \frac{\partial^2 p_n}{\partial x^2} \tag{5.22}$$

If an electric field is applied, the solution to this differential equation will be:

$$p_n(x,t) = \frac{N}{\sqrt{4\pi D_p t}} \exp\left(-\frac{(x-\mu_p E t)^2}{4 D_p t} - \frac{t}{\tau_p} \right) + p_{n0} \tag{5.23}$$

where N is the number of electrons or holes generated per unit area. In the presence of a constant electric field the excess carriers moves with the drift velocity $\mu_p E$ [HS51]. At the same time, they diffuse outward as shown in Figure 5.11.

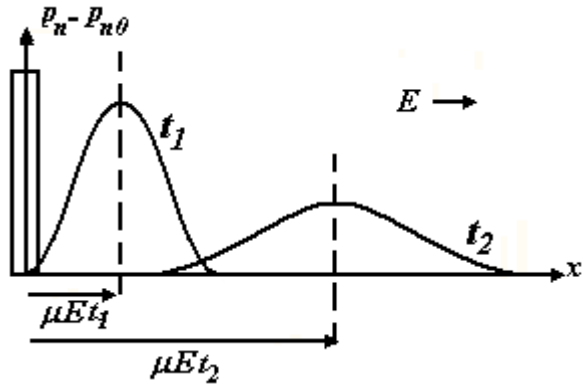


Figure 5.11. Carrier diffusion after the laser pulse in the presence of constant electric field [Sze98]

5.11. Backside Analysis

Frontside failure analysis is facing severe limitations with the growing use of multi-level metal processes, which prevents the laser beam from reaching the internal nodes of the devices under investigation or emission from reaching the detector. Therefore, the physical failure analysis techniques access the chip from the backside. For this purpose, the device is polished down to 50/200 µm so that the laser stimulation and photon emission analysis can be performed from the backside. Backside analysis takes the advantage of the fact that Si is transparent to NIR light with photon energies smaller than Si energy bandgap of 1.12 eV, corresponding to wavelengths of 1.1 µm and higher. Figure 5.12 shows the light transmission for different doping concentrations.

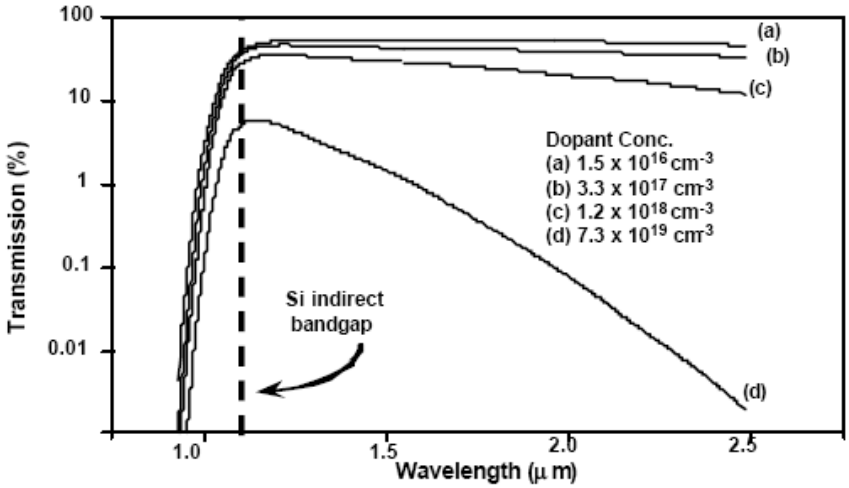


Figure 5.12. Light transmittance of p-Si with different doping concentrations [Chi00]

5.12. Laser Stimulation from the backside in the means of the laser wavelength

The time-averaged intensity distribution formulas for front-side and back-side are shown below [LPB01].

$$I(r,z) = I_0\left(\frac{w_0}{w(z)}\right)^2 \cdot \exp\left(\frac{-2r^2}{w^2(z)}\right) \cdot \exp(-\alpha \cdot z) \qquad \text{for front-side} \qquad (5.23)$$

$$I(r,z) = I_0\left(\frac{w_0}{w(z-d)}\right)^2 \cdot \exp\left(\frac{-2r^2}{w^2(z-d)}\right) \cdot \exp(-\alpha \cdot z) \qquad \text{for back-side} \qquad (5.24)$$

In these formulas, r is the radial distance from the center axis of the beam, α is the absorption coefficient, d is the sample thickness where the laser beam propagates, w_0 is the beam waist, z is the distance from the center of the axis beam, z_R is the Rayleigh range, $w(z)$ is the spot size at a distance of z and I_0 is the intensity at the center of the beam at the waist.

The propagation of the light beam in a 42 µm thick Si substrate with a doping concentration of 10^{16} cm^{-3} for 500 nm, 800 nm and 1064 nm wavelengths are calculated in MATLAB and shown in Figure 5.13, 5.14 and 5.15, respectively. The intensity distribution formula for backside is used for the calculations and I_0 is taken as unity. For 500 nm and 800 nm laser wavelengths, the light intensity drops severely and light can penetrate only in small distances. On the other hand, the loss of intensity can be neglected for 1064 nm wavelength

laser. This difference in the intensities for the different wavelengths is due to their different absorption coefficients.

According to Figure 5.12, the absorption coefficient has a sharp edge close to the bandgap energy of Si (1.1 µm). The photon energy for 1064 nm laser wavelength is 1.165 eV, which is slightly higher than the Si bandgap energy, leading a low absorption coefficient. Therefore, the absorption is weak, the charge density in the track is low and the track can penetrate deep into the semiconductor. Conversely, photons with energies much greater than the bandgap are strongly absorbed (for example 500 nm and 800 nm laser wavelengths), produce a very dense charge track and penetrate only short distances into the semiconductor

1064 nm wavelength laser light is able to create electron-hole pairs in Si substrate while the absorption coefficient for this wavelength is low enough for laser beam to penetrate deep into Si substrate. Therefore, this wavelength seems to be a good choice for laser stimulation applications.

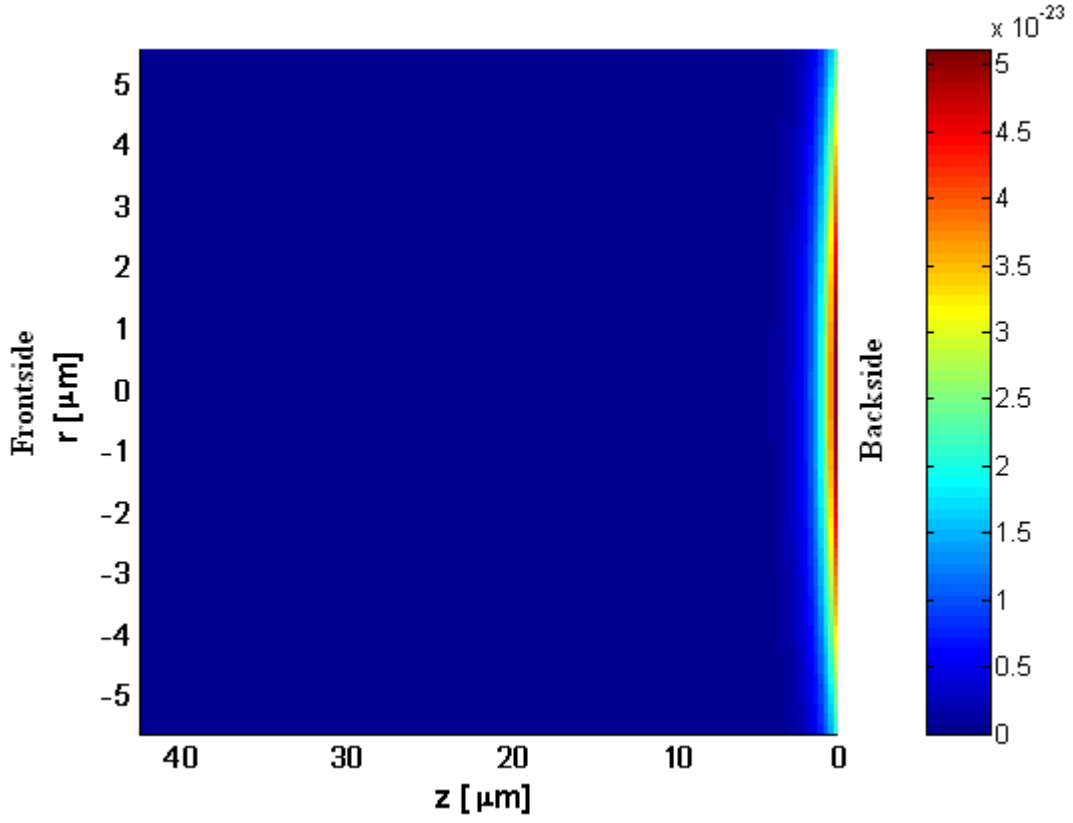


Figure 5.13. The propagation of 500 nm wavelength laser beam from the backside of a 42 µm Si sample (*d=42 µm, w_0=0.79 µm, z_R=4.0 µm*)

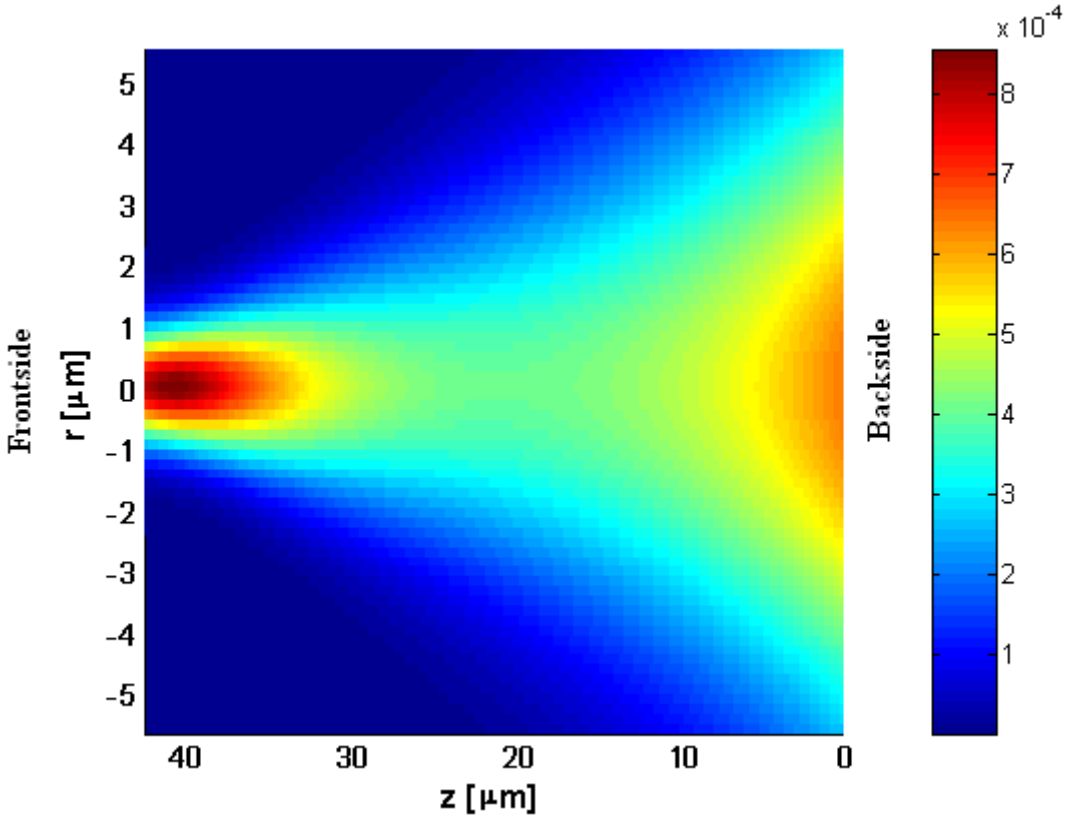


Figure 5.14. The propagation of 800 nm wavelength laser beam from the backside of a 42 µm Si sample ($d=42\mu m$, $w_0=1.27\mu m$, $z_R=6.4\ \mu m$)

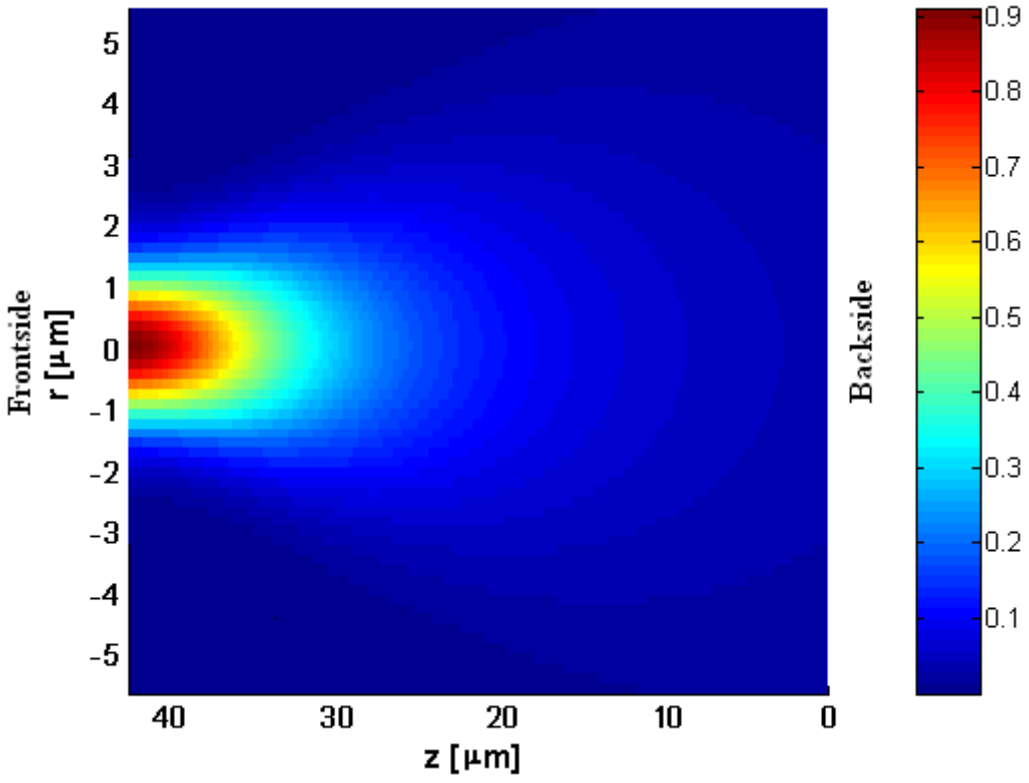


Figure 5.15. The propagation of 1064 nm wavelength laser beam from the backside of a 42 µm Si sample ($d=42\ \mu m$, $w_0=1.7\ \mu m$, $z_R=8.53\ \mu m$)

6. The Techniques

Failure Analysis and debug of integrated circuits can be tackled using optical techniques. These techniques make use of the silicon optical transparency in the near infrared (NIR) wavelength range. We have used two main optical techniques for the investigation of the devices: laser stimulation and photon emission.

Laser stimulation techniques use laser beams to stimulate devices that are sensitive to carrier generation or temperature. The laser beam injects photons locally onto an area of the device. Depending on the laser wavelength, these photons may induce current (electron-hole generation) or heat. If the laser energy is smaller than the band gap energy of Si, no carriers are generated. But, it causes local heating on the device. If the laser beam energy is greater, both carrier generation and heat-build up occurs. In this case, the dominant effect is the electron-hole generation.

Photon emission microscopy is based on electro-luminescence and detection is achieved with a photon-sensitive camera. Emission of light from semiconductor devices results from the transformation of electrical energy into photons by excited radiative systems. Emission of radiation generally occurs when charge carriers occupying a higher energy state jump to an empty lower energy state and all or most of the energy difference of the two states can be emitted as electromagnetic radiation. This light must be detected very sensitively by a CCD camera.

6.1. Laser Stimulation Techniques

Laser stimulation techniques can be divided into two groups: static and dynamic. While statical laser stimulation requires just a power supply to power up the device, dynamic laser stimulation requires full activation of the DUT, mostly by an ATE.

In deep sub-micron designs, it is impossible to find all the faults by using only statical approach. Many defects can be found in a static configuration, but some failures are only activated in a dynamic mode. For example, failures caused by racing between the logic blocks can be only found in dynamic mode. The induced current or heat may perturb the switching time of the devices, slowing down or speeding up the devices. Dynamic laser stimulation technique is able to monitor these slight timing variations.

The device perturbations can be induced by two different wavelength laser diodes: 1300 nm thermal effect and 1064 nm carrier generation. Many techniques have been invented that are based on the laser stimulation principle. Table 6.1 gives an overview of the most common techniques for statically and dynamically driven circuits.

Table 6.1. Laser stimulation techniques

Techniques		Source	Read-out	Device Analysis
TLS	**OBIRCH**	Voltage	Current	Identification of short circuits
	TIVA	Current	Voltage	Identification of short circuits
	SEI	No bias	Voltage	Identification of opens
	RIL and SDL	Dynamic	Pass/Fail - Functional Test	Identification of soft defects
PLS	**OBIC**	Voltage	Current	Localization of p-n junctions
	LIVA	Current	Voltage	Read out of signals sensitive to logical state
	LADA	Dynamic	Pass/Fail - Functional Test	Localization of speedpaths
	DVM	Dynamic	Propagation delay	Delay variation mapping
	PVM	Dynamic	Phase variations	Phase variation mapping

6.1.1. Thermal Laser Stimulation

Using 1.3 μm laser beam, the device is heated in order to localize the abnormal resistive issues. This technique is called TLS. Laser beam energies lower than Si band gap energy (1.1eV) causes a local heating effect on the materials. During the interaction of the laser beam and the medium, the laser photons are absorbed by the material and converted into phonon to induce a local temperature variation. The induced temperature can change some electrical properties of an IC through the modification of the resistivity of the heated material, thermoelectric voltage generation (Seebeck Effect Imaging-SEI) or heat transfer to the active areas of the devices.

Resistivity Change of the Heated Interconnects

Temperature changes the resistance of the heated material. If the material is a conductor, the resistance will increase with temperature, which is modelled as follows:

$$\Delta\rho = \rho_0 \alpha_{TCR} (T - T_0) \tag{6.1}$$

where ρ_0 is the resistivity, α_{TCR} is the thermal coefficient of the material and T is the temperature.

There are two LSM techniques, which use resistivity change signature: Optical Beam Induced Resistivity Change (OBIRCH) and Thermally Induced Voltage Alteration (TIVA).

OBIRCH is an imaging technique which uses a laser beam to induce thermal changes in the metallic, polysilicon elements and highly doped substrate areas through modification of the resistivity of the heated material. A constant voltage is applied to the DUT. As the laser locally heats an area on a metal line, which is carrying a current, the resulting resistance changes can be detected by monitoring the input current to the device. The current change ΔI is linearly proportional to resistivity change ΔR, which can be expressed as

$$\Delta I = -(\Delta R / R^2) V_{\text{sup}} \tag{6.2}$$

In contrast to OBIRCH, TIVA displays voltage shifts by using constant current biasing approach, which can be expressed as

$$\Delta V = \Delta R \cdot I_{\text{sup}} \tag{6.3}$$

OBIRCH and TIVA are useful for detecting short circuits. Because, as the resistance of the short site changes, the power demand of the IC changes.

Seebeck Effect Imaging (SEI)

Seebeck effect is the direct conversion of temperature differences to electric potential gradients with typical values on the order of mV/K. The temperature gradient on a material causes charged carriers to diffuse from the hot side to the cold side. Accumulation of the carriers at one side will result in an electric field.

The strength of the seebeck voltage is determined by the seebeck coefficient. It is a measure of the magnitude of an induced seebeck voltage in response to a temperature difference across that material. It can be expressed as

$$S = \frac{dV}{dT} \quad [\mu V/K] \tag{6.4}$$

In general, when two different materials form a junction and the sides of the junction are kept at different temperatures, a potential difference is generated across the junction. In case of a laser beam scanning a thermoelectric junction (as illustrated in Figure 6.1), one side is heated first, generating a voltage difference through the thermocouple. After the laser beam crosses the junction, a voltage in the opposite direction is generated. The generated Seebeck voltage can be written as

$$\Delta V = (S_A - S_B)\Delta T \tag{6.5}$$

where S_A and S_B are the seebeck coefficients of materials , A and B.

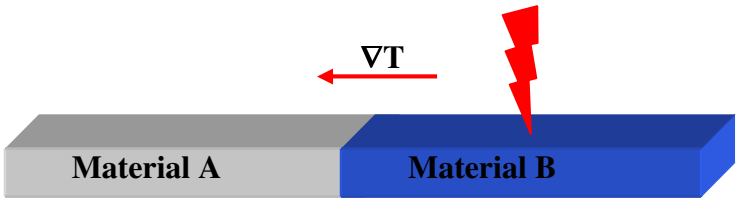


Figure 6.1. Thermal gradients induced at a thermocouple as a function of the laser beam position

Open conductors can be localized by SEI technique. When the laser changes the thermal gradient of a floating conductor, its electrical potential changes. This change in potential will change the bias of any transistors connected to the floating conductor, which affects the power dissipation of the device. SEI signatures are not always that much clear and very often not easy to interpret, mostly because of superposition of additional OBIRCH when the device is biased.

TLS on MOS devices

Temperature changes two factors in a MOSFET. These are the threshold voltage and the mobility of the carriers. The temperature dependence of mobility follows a $T^{-3/2}$ rule, indicating a carrier scattering mechanism. For Ge and Si, the mobility caused by phonon interaction can be written as

$$\mu \approx (m^*)^{-5/2} T^{-3/2} \tag{6.6}$$

where m^* is the conductivity effective mass and T is the temperature [Sze98].

Furthermore, the decrease in the well fermi potential with the temperature is responsible for the reduction in the threshold voltage. The minority carrier density increases with the temperature. Because, intrinsic carrier concentration, n_i, strongly depends on the temperature. As n_i increases exponentially with temperature, the minority carrier density increases, too.

As a result, an increase in the temperature enhances the carrier density in the channel, which results in threshold voltage drop off. Competition of this effect with a decrease in mobility for increasing temperature yields operational conditions with negative, positive or zero temperature coefficients. The change in the mobility and threshold voltage shifts the drain current. The temperature influence on the transfer characteristics of a FET is shown in Figure 6.2.

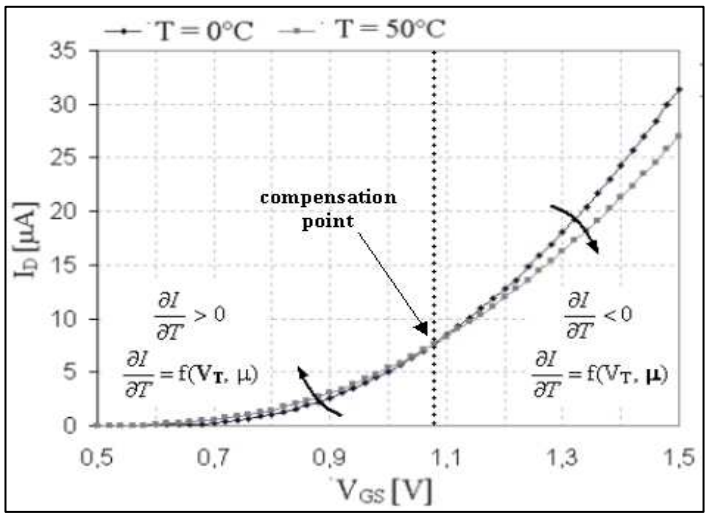


Figure 6.2. The temperature effect on the transistor transfer characteristics [GBS07]

The channel of the MOSFET is not heated directly but indirectly through heat diffusion from neighboring regions. These regions are polysilicon gates and interconnects for the frontside laser stimulation. For the backside laser stimulation, light to heat conversion occurs in drain/source active areas and buried layers, in addition to polysilicon gates and interconnects as shown in Figure 6.3 [BGB04].

Thermal laser stimulation effect from the backside is investigated on 90 nm technology nMOS and pMOS transistors. Figure 6.3 to Figure 6.7 shows the laser on and laser off transfer characteristic curves and the subtraction between the current values obtained with and without laser stimulation. The compensation voltage for both transistors where no drain current change is observed is around 0.5 V. Above compensation voltage, thermal laser stimulation decreases the drain current whereas lower voltages boosts the current flowing from drain to source contact. The maximum current reduction is around 1.7 µA for nMOS and 1.2 µA for pMOS transistor which leads to 1.84% and 3.4% of net current change, respectively.

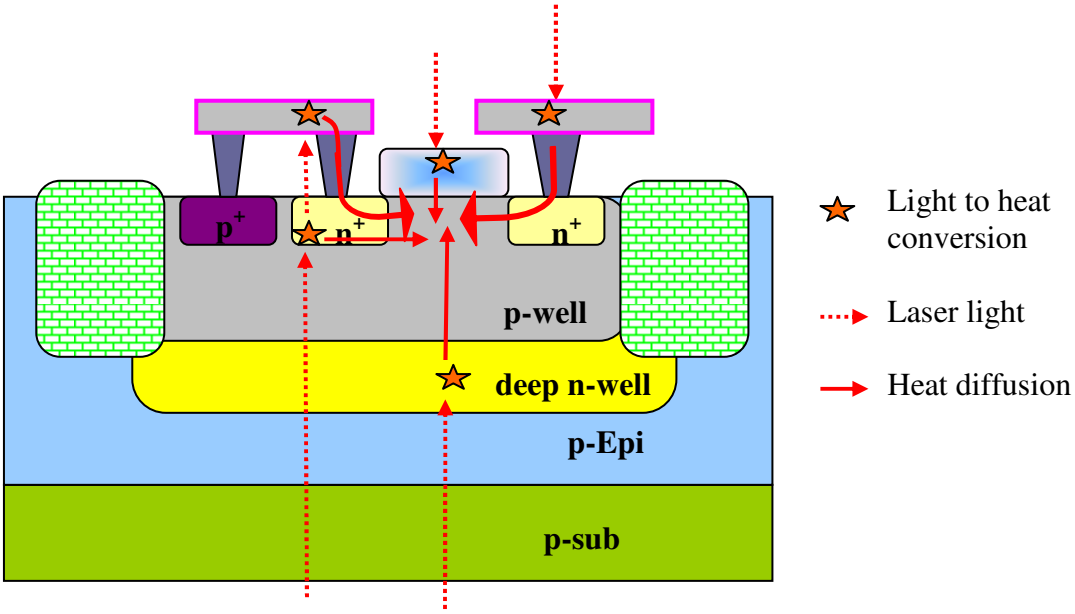


Figure 6.3. Heat diffusion for the frontside and the backside thermal laser stimulation on an nMOS transistor

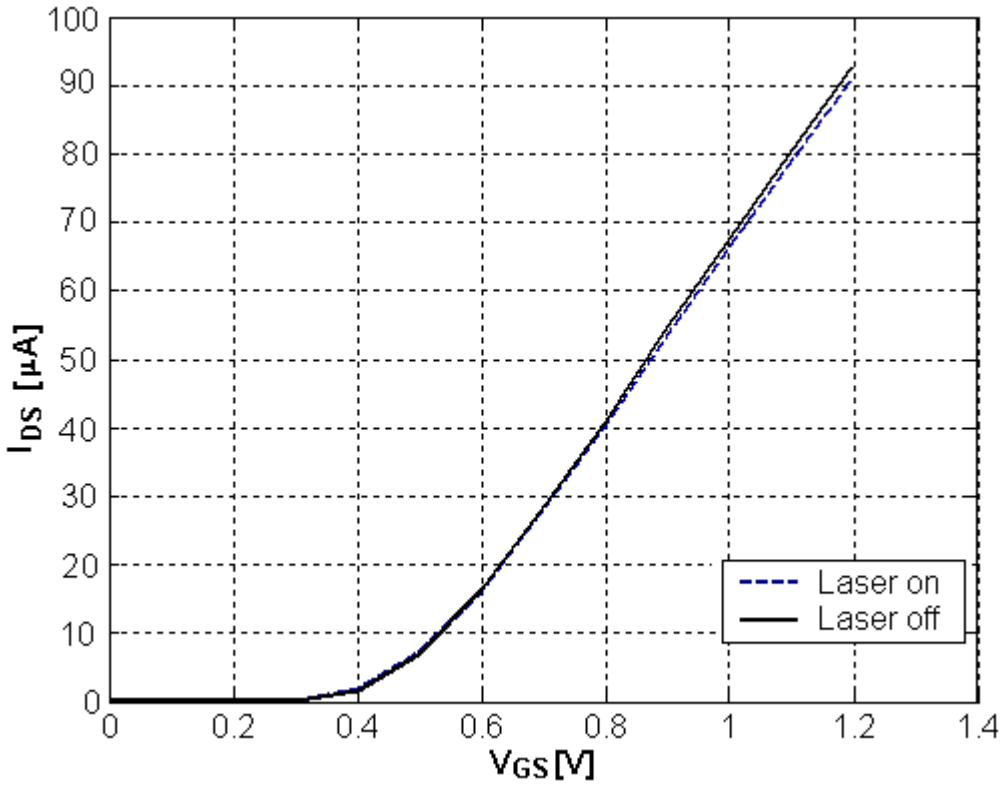


Figure 6.4. The transfer characteristic of a 0.12 µm/0.08 µm nMOS transistor with and without laser stimulation (20.6 mW laser power on the device)

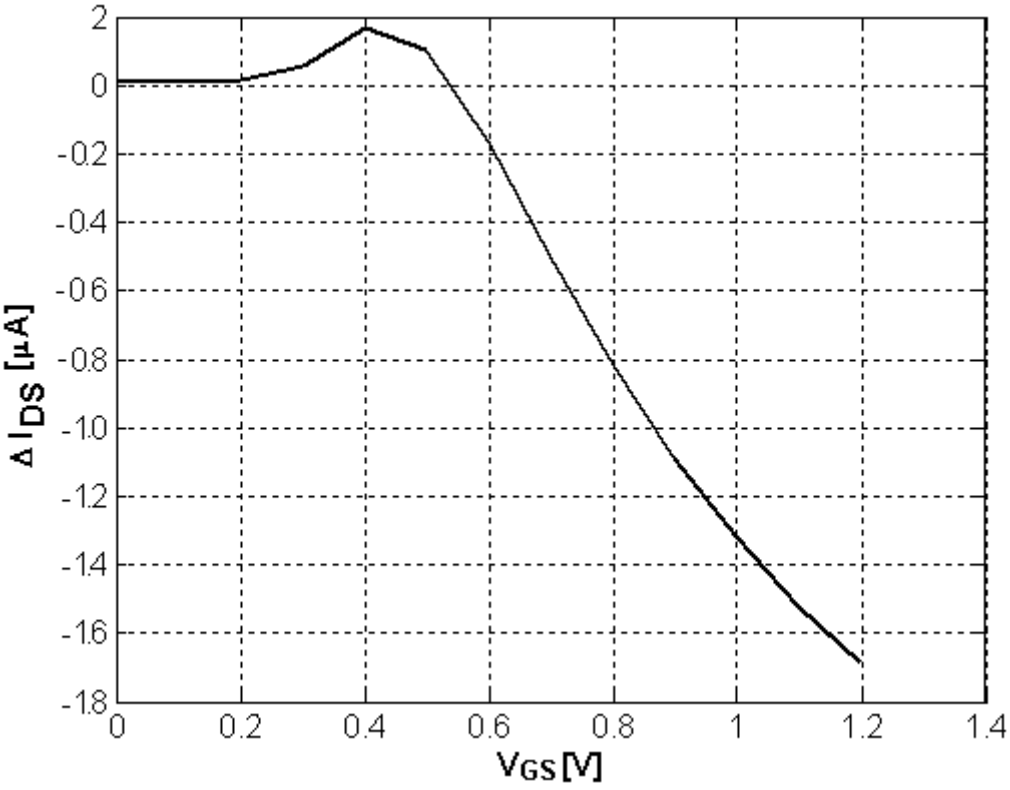


Figure 6.5. The induced current by thermal laser stimulation at the drain contact of a nMOS transistor by thermal laser stimulation (20.6 mW laser power on the device)

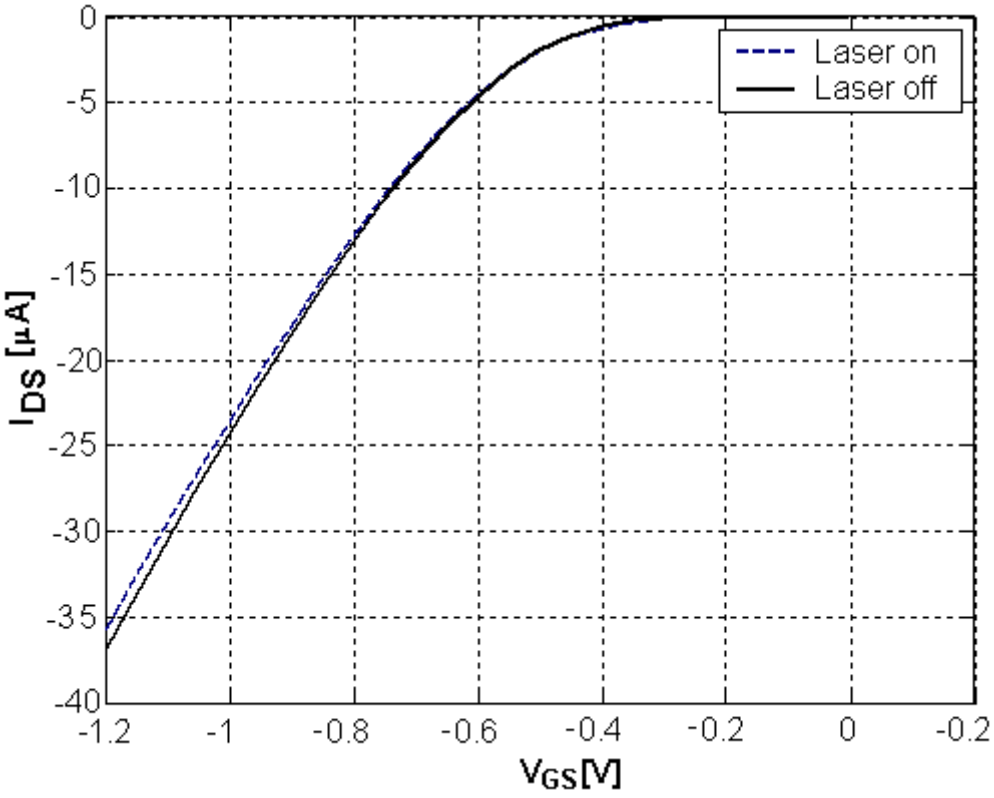


Figure 6.6. The transfer characteristic of a 0.12 µm/0.08 µm pMOS transistor with and without laser stimulation (20.6 mW laser power on the device)

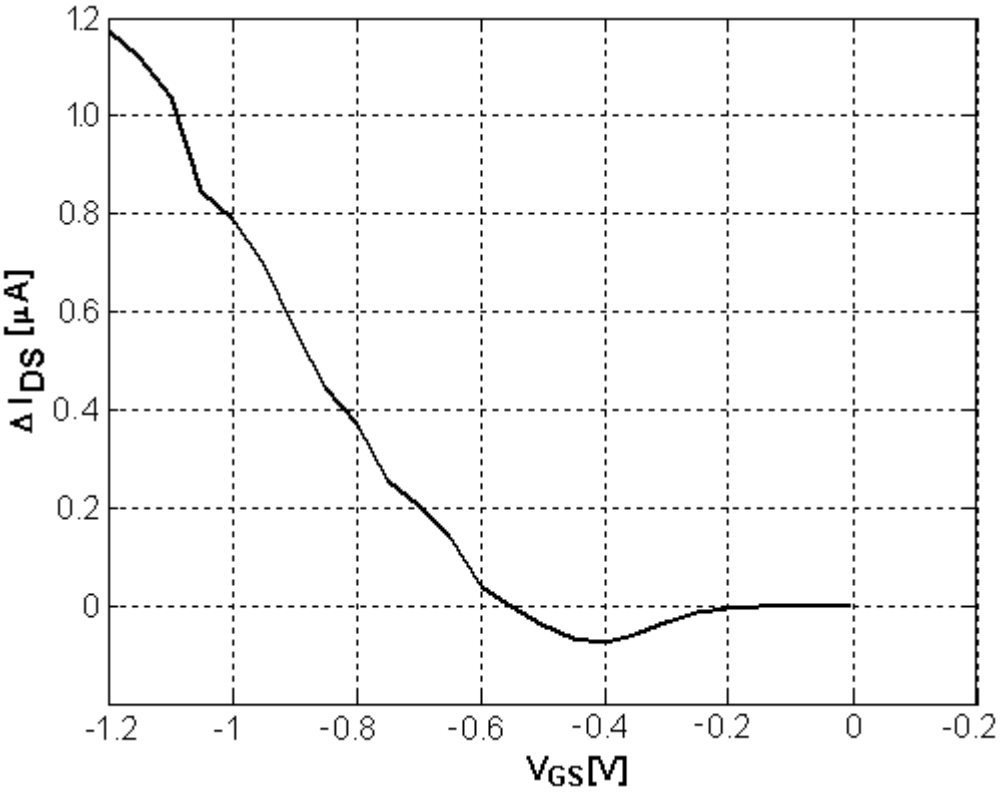


Figure 6.7. The induced current by thermal laser stimulation at the drain contact of a pMOS transistor (20.6 mW laser power on the device)

6.1.2. Photoelectric Laser Stimulation

Illuminating the device with a 1.064 μm wavelength laser beam causes two effects to occur. These effects are the generation of the electron-hole pairs and the heat build up. The generation of electron-hole pairs is related to the interband absorption of photons (as the energy is greater than the band gap energy). The photoelectric effect is much stronger than the thermal effect. The thermal effect can be neglected because; the layers of high carrier density regions are very thin. On the other hand, electron-hole pairs are generated in whole bulk material. Therefore, this technique is called Photoelectric Laser Stimulation (PLS).

It uses laser beam to generate localized photocurrent within the active regions of the semiconductor temporarily altering the device characteristics. Induced excess carriers on a semiconductor will recombine after carrier lifetime. However, if the carriers are generated within or in the diffusion length away from a SCR, the carriers will be dissociated by the junction potential and a net photo current will be generated. The photocurrent that can be measured will occur when the laser beam is on an SCR such as channel and drain/substrate or source/substrate junctions. The illustration of charge separation and the generation of photocurrent signature are seen in Figure 6.8.

Optical Beam Induced Current Change (OBIC) and Light Induced Voltage Alteration (LIVA) as PLS techniques map the SCRs. Therefore, these techniques are useful for detection of defective junctions and buried diffusion regions as well as localization of gate oxide shorts (with deprocessing). In addition to that, they are able to identify the logic states of transistors [Col04]. In OBIC applications, the DUT is biased in constant voltage mode and the current changes are mapped and it is vice versa for LIVA.

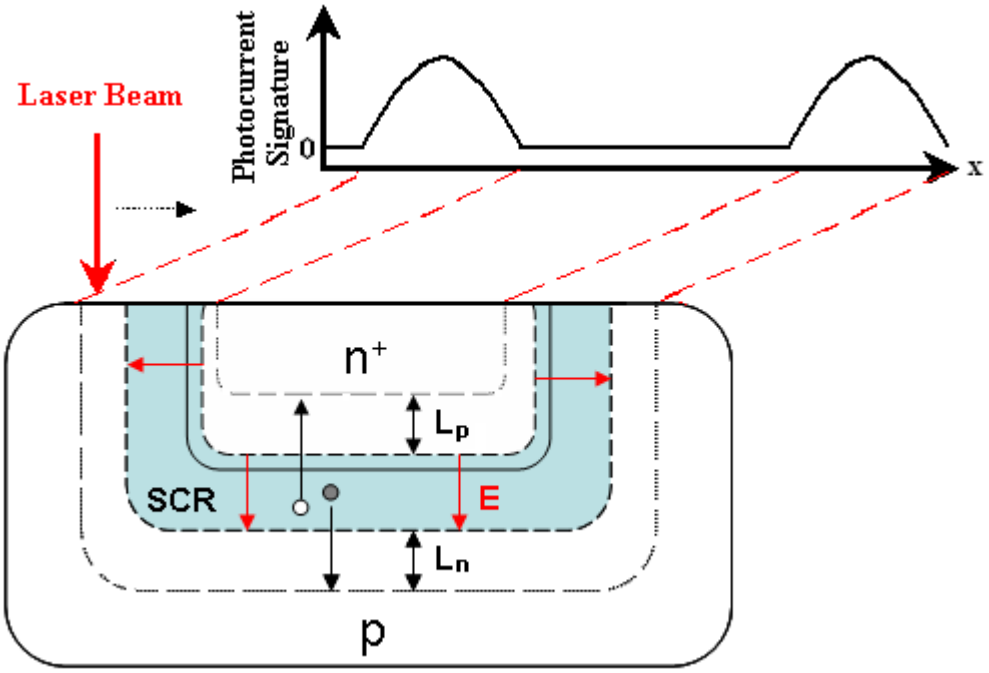


Figure 6.8. Signal generation by PLS [Col04]

If the laser incident is on the frontside, active regions of the device are masked by the interconnects. In case of backside, they are completely exposed. The illustration of the stimulation from the backside for n- and p-channel MOSFET is shown in Figure 6.9 and Figure 6.10, respectively. The electron-hole generation occurs at the SCRs of the channel, well-highly doped areas and nwell-p substrate junctions [GBS07]. For nMOS transistor, an additional electron–hole generation source is the pwell-deep nwell reversed biased junction for triple-well technologies.

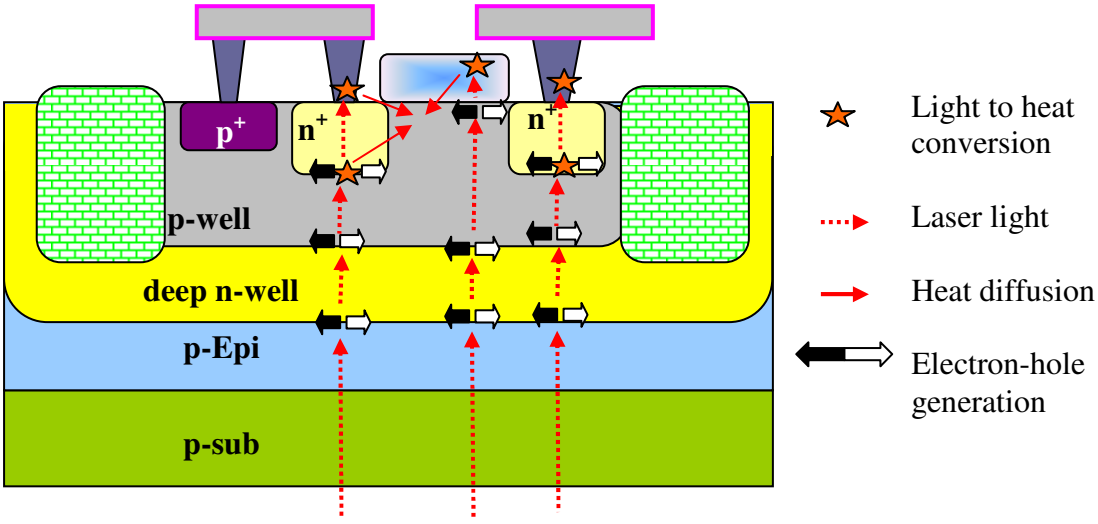


Figure 6.9. n-MOSFET backside PLS schematic

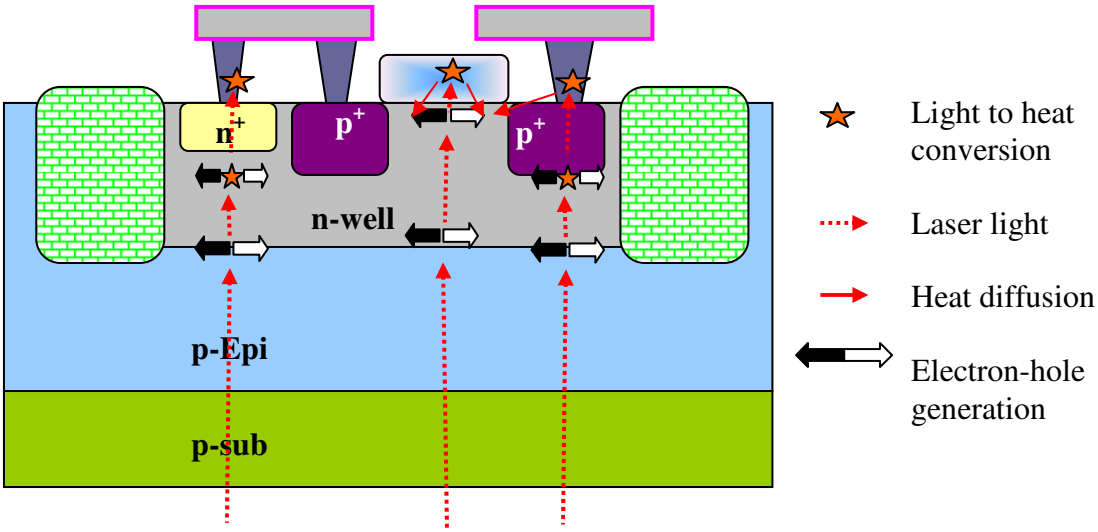


Figure 6.10. p-MOSFET backside PLS schematic

The effects of the 1064 nm wavelength laser are studied for short channel n- and p-channel transistors fabricated in triple-well CMOS process technology. In Figure 6.11 and 6.12, the laser induced drain current change and transfer characteristic with and without laser illumination are shown for an nMOSFET.

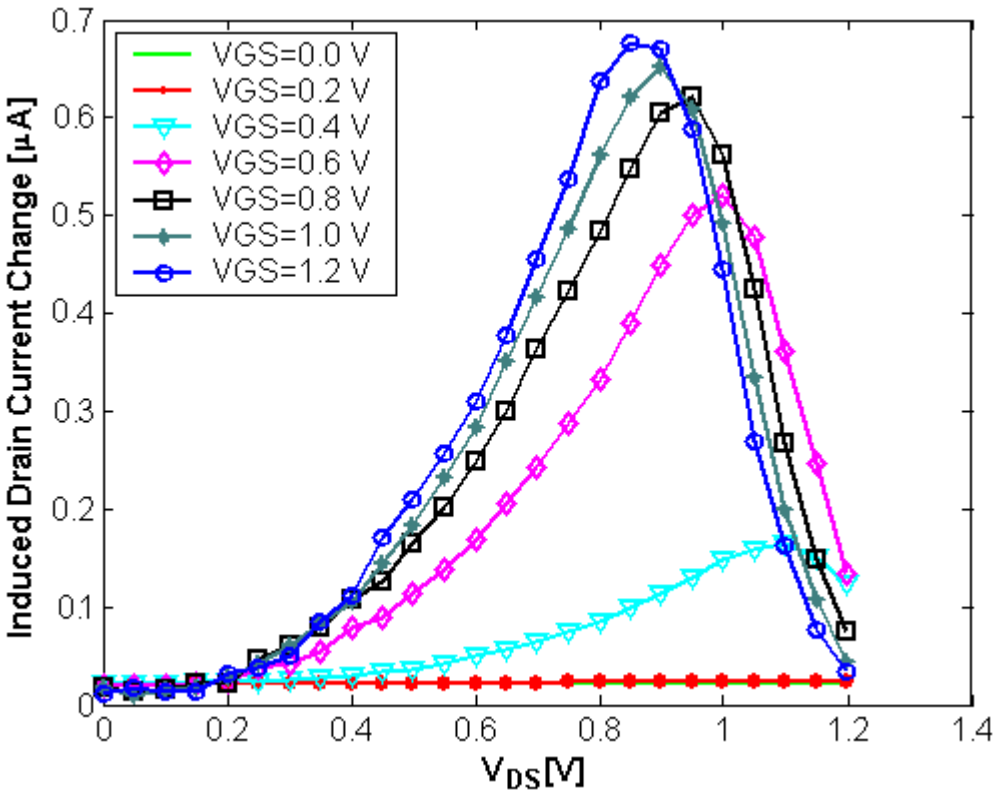


Figure 6.11. PLS drain current change introduced by the 1.064 µm laser beam on a 0.12x0.08µ^2 nMOSFET for backside analysis (laser power is 0.96 mW and 50x lens is used)

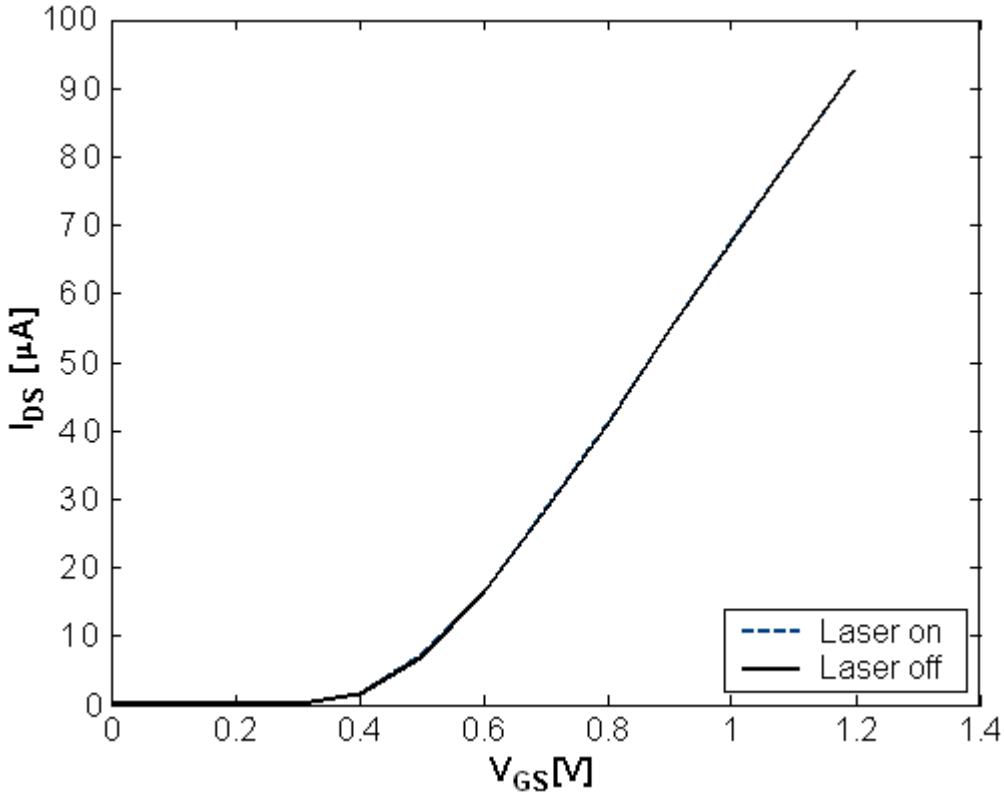


Figure 6.12. PLS bulk current change introduced by the 1.064 μm laser beam on a 0.12x0.08μ^2 nMOSFET for backside analysis (V_{DS} is 1.2 V, laser power is 0.96 mW and 50x lens is used)

When the transistor is off, the net photocurrent is constant at 20 nA independent of V_{DS} voltage. When the transistor is on, the generated charge carriers in the SCR under the channel increase the inversion layer charge and the drain current. The maximum photo current is 0.7 μA for 0.96 mW laser power (4% of maximum laser power). The generated charge carriers are collected by the transistor which results in an increase in the drain current. The photocurrent collected by the drain terminal is modulated by the applied gate and drain voltages. The laser stimulation effect on nMOSFET can be modeled as shown in Figure 6.15a. If the resistance of the p-well is neglected, it can be modeled by a current source bypassing drain and source junctions.

The output and transfer characteristics of illuminated pMOSFET are shown in Figure 6.13 and Figure 6.14, respectively. When the V_{GS} voltage is lower than -0.8 V, the induced photo current curves become erratic. It might be due to some degradation of the pMOSFET. With these devices, the experiments have been repeated later and the results could not be reproduced again. The threshold voltage shift is clear in the transfer characteristic under laser stimulation. Photocurrent generated in the bulk material acts similar as a change of bulk voltage (back gate). Due to high resistivity of bulk material, the photocurrent causes a voltage drop in V_{BS}, which results in a decrease in threshold voltage. But, there is still some additional current collected by drain-nwell junction, which is a function of V_{GS} and V_{DS} voltages. In case of an increase in V_{DS}, the space charge region expands, thus the collected current by drain terminal increases. When the transistor is off, the laser-induced current is around -0.5 nA, which is negligible. The electrical PLS effect model on pMOSFET is illustrated in Figure 6.15b. If nMOS and pMOS transistors are compared, the additional drain current on pMOSFET is approximately 4 times larger.

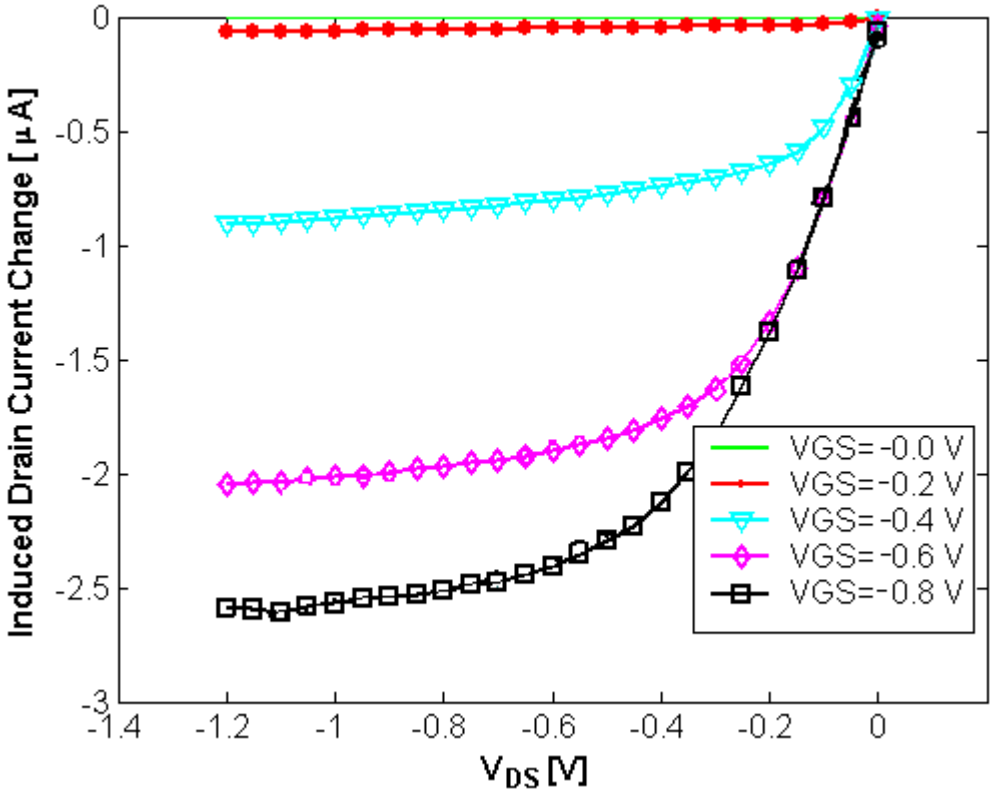


Figure 6.13. Drain current change introduced due to PLS on a 0.12x0.08μ^2 pMOSFET for backside analysis (laser power is 0.96 mW and 50x lens is used)

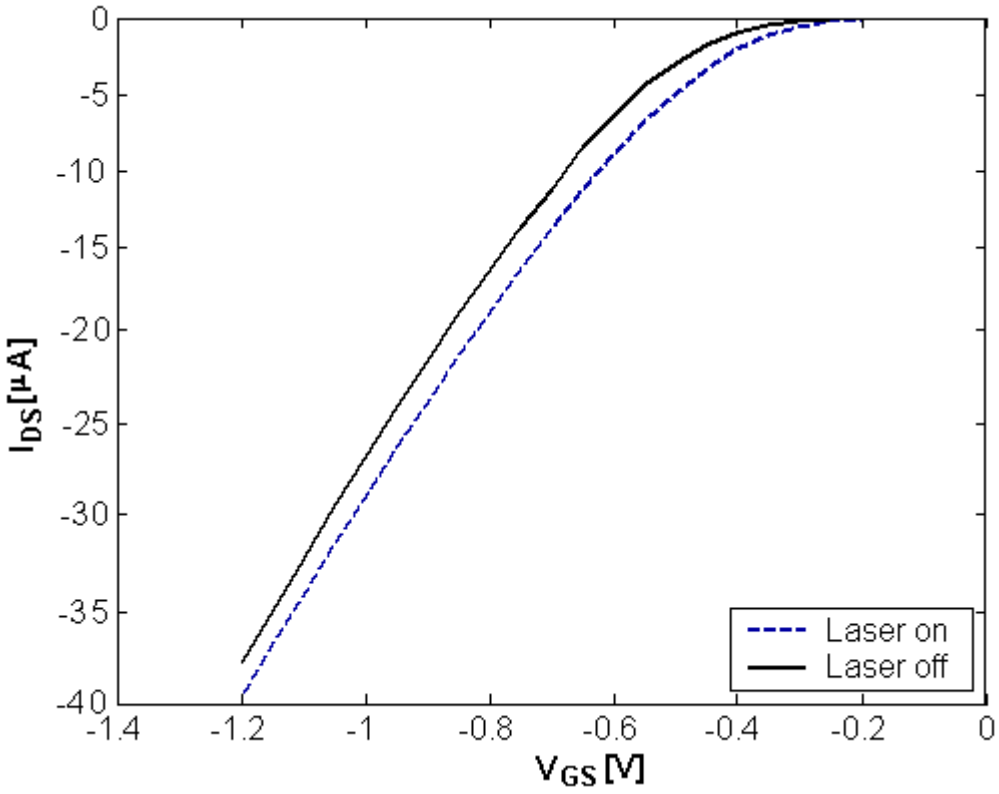


Figure 6.14. Transfer characteristic of a 0.12x0.08μ^2 pMOSFET for backside analysis with and without laser illumination (V_{DS} is -1.2 V, laser power is 0.96 mW and 50x lens is used)

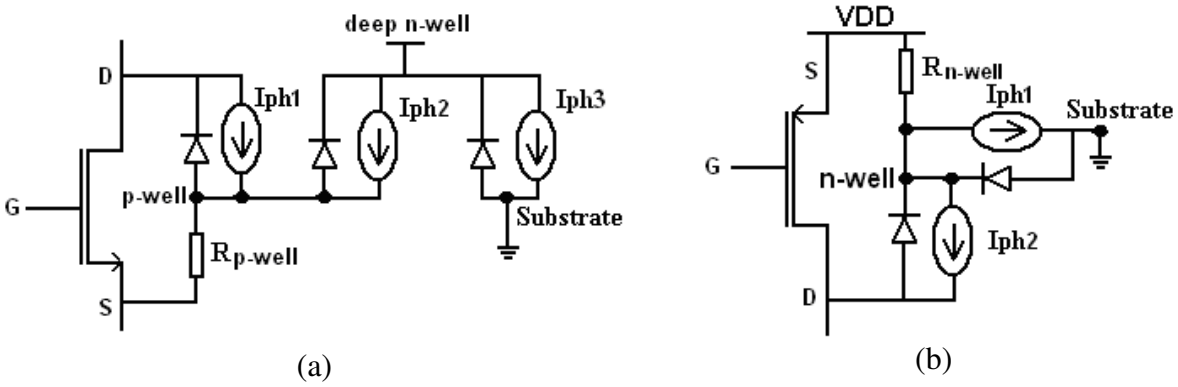


Figure 6.15. Electrical model of PLS effect on (a) nMOSFET and (b) pMOSFET

6.1.3. The effect of laser stimulation on the switching properties of the transistors

The propagation delay times of an inverter, t_{PLH} and t_{PHL} determine the input-to-output signal delay during low-to-high and high-to-low transitions of the output, respectively. t_{PLH} refers to the time required for the output to reach 50% of its final output level when the input changes from high to low. Similarly, t_{PHL} is the time delay between the $V_{50\%}$ -transition of the rising input voltage and the $V_{50\%}$ -transition of the falling output. Propagation delay formula of a CMOS inverter is derived in the literature [Rab96] and it is given in Appendix B. It can be expressed as:

$$t_{PHL} = 0.52 \frac{C_L V_{DD}}{I_{DSATn}} \tag{6.7}$$

$$t_{PLH} = 0.52 \frac{C_L V_{DD}}{I_{DSATp}} \tag{6.8}$$

where C_L is the load capacitance, and I_{DSATn} and I_{DSATp} are the saturation currents for nMOS and pMOS transistors, respectively.

The overall propagation delay of the inverter is defined as the average of the two values:

$$t_{pd} = \frac{t_{PHL} + t_{PLH}}{2} \tag{6.9}$$

VDD and C_L stay unchanged by the laser stimulation. The drain current modifies the charge or discharge time of the load capacitance. Thus, we concern ourselves with the change in the drain current induced by the laser.

Timing alteration by TLS

The drain current variations of the nMOS and the pMOS transistors under thermal laser stimulation affect the speed of the inverter. Heat build up reduces the mobility and is therefore limiting the drive strength of the transistor for nominal supply voltages. This causes a slow down at both rising and falling edges. This property of thermal laser stimulation can be used to isolate the performance limiting circuitry in a complex IC. Though, since the local heating maximum temperature rise is restricted to a few degrees, the timing variation by TLS is very weak.

Despite of this fact, it can be used for the localization of interconnects such as defective

vias, contacts and conductors, which are called “soft” defects. Actually, the delay variation induced by TLS of soft defects is mostly sufficient to affect the output switching time. For example, higher temperatures are known to slow down the device performance as the metal resistance increases.

The Seebeck effect on the gate conductor can be transferred into a change in the drain current by modulation of the channel resistivity via gate voltage modulation. The SEI signatures are visible on the MOS transistor gate conductor even if the MOS transistor is biased. When the laser hits the junctions on the gate interconnect, only Seebeck voltage is generated and no OBIRCH can occur as there is no gate current. When the laser scans across the junction in the gate conductor, the Seebeck signature occurs (as a signal of the FET output circuit) independently of the supply level [BBG05].

As a result, the induced seebeck voltage may alter the switching speed of the devices via modification of the drain current.

Timing alteration by PLS

PLS technique can be used for the timing analysis of defect-free performance limiting circuits in advanced devices. These techniques use a laser incident from the backside of the chip to perturb the timing of internal nodes by means of temporary alteration of transistor characteristics. The photocurrents affect the timing at a node directly assisting the charging or discharging of a load capacitance. The photoelectric laser stimulation effect on a DUT is illustrated in Figure 6.16. Due to different effect on pMOS and nMOS devices, PLS can be used to speed up or slow down the devices such as inverters. When it is applied to the devices in critical timing paths in a healthy device, performance-limiting circuits can be isolated.

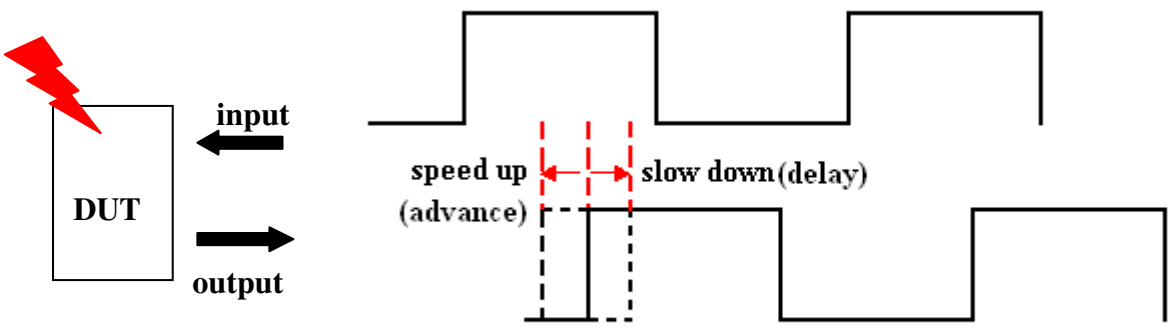


Figure 6.16. Delay variation caused by 1.064 μm laser stimulation on a delay chain

The measured laser stimulation effect at the output of the chain is a function of the rising and the falling edges, the number of the inverter in the chain and the illuminated device type. As illustrated in Figure 6.17, the illumination of pMOSFET at the rising edge of the output causes an advance. At the rising edge, pMOS transistor goes ON from OFF condition. The illumination increases the drain current thus strengthens the device driving capability. Therefore, the load capacitance at the output of the inverter is charged faster causing a speed-up. For the falling edge, pMOSFET switches off. Since the drain current is higher, pMOSFET turns off slower causing a delay. The effect on the nMOS is vice versa. At the falling edge of

the output, nMOSFET charges off the load capacitance. Due to increased drain current under illumination, charge off takes place faster causing a speed up. Similarly, on the rising edge, nMOSFET should go from ON to OFF. However, the photocurrent slows down the switch off operation.

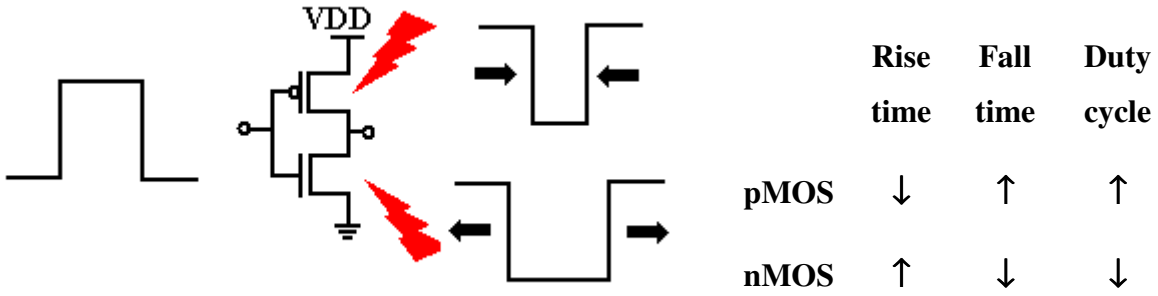


Figure 6.17. The effect of PLS on the rising and falling edges of a single inverter output

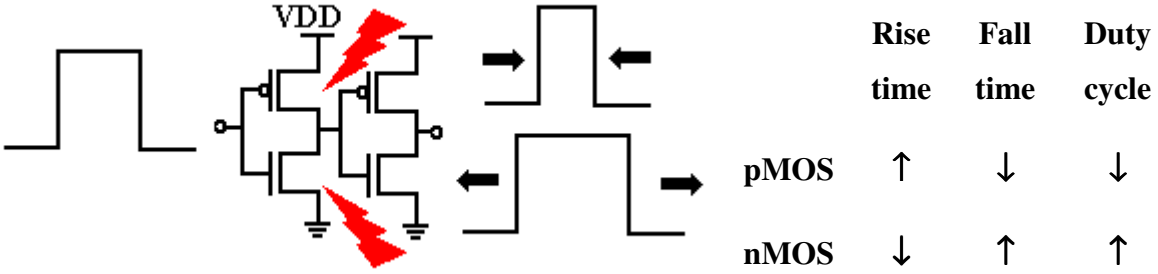


Figure 6.18. The effect of PLS on the rising and falling edges of a buffer output

The first inverter was illuminated in Figure 6.17 whereas the second inverter starting from the output side is exposed in Figure 6.18. The measured effects at the output are different for both cases. In the former case, when the pMOSFET is exposed, the output signal is advanced at the rising edge and it is delayed at the falling edge, thus a duty cycle increase is observed. In the latter case, it is the opposite. The duty cycle decreases. As a result, the illumination causes duty cycle increase on the odd-numbered inverter pMOSFETs (counted from the output side) and a decrease on the even number inverters.

The presence of induced high bulk current can affect the dynamic behaviour and alters the timing globally even if the illuminated site on the device is not a part of the investigated circuitry. Global timing shifts can be avoided by using pulsed laser instead of CW laser.

One another application of PLS is the soft fault injection into internal nodes of the illuminated devices. Fault injection phenomenon happens in memory elements, such as SRAMs or flip-flops. The induced drain current in the feedback loop of the memory element transistor increases the drain node voltage. This amplitude change is kept constantly in the loop (as long as there is illumination) and then it is conveyed to the output of the flip-flop or

to the second stage of the flip-flop.

This technique can be used for Single Event Upset (SEU) analysis if the ultra short laser pulses are used. The applied laser pulse can experimentally simulate the behaviour of a memory element as if a heavy ion enters the semiconductor. Heavy ions may corrupt a logical state and cause failures and errors.

Soft fault injection technique can also be used to define precisely the most critical point in time of a switching event in a node when a pulsed laser is used. Using a laser diode pulsed by external triggering provides a very convenient opportunity to do a time mapping of stimulation sensitivity and define the event that is closest to faulty operation which gives a deeper insight to the circuit operation.

Pulsed laser stimulation can also be used for reverse-engineering applications in golden devices, for example to localize a desired scan flip-flop in the chain.

6.2. Photon Emission Microscopy

Photon emission microscopy is a successfully used tool in failure analysis of integrated circuits. It detects the faint light emitted from a golden or faulty device. The location of the light emitting spot can be pinpointed by superimposing the photon emission image on the reflected light micrograph of the device. During the switching of the transistors, the transistors emit light and light emitting spots are detected by a CCD camera.

Photon emission in semiconductors occurs in two basic mechanisms:

1. The charge carriers that have high kinetic energy in an electric field relax and emit light (F-PE). It happens at reverse biased junctions of CMOS devices. Reverse biased currents are usually very low. The current should be increased in order to have enough emission intensity for detection. A leakage current in CMOS technologies may trigger this mechanism. Besides, MOSFET operating in saturation has high electric field combined with high current, which yields high intensity photon emission.
2. Radiative band-band recombination of electrons and holes (R-PE) causes electroluminescence. Recombination involves carriers from conductance and valence band thus it is an interband process. Since silicon is an indirect semiconductor, it rarely happens.

A MOSFET operating in saturation region is a source of photon emission. The luminescence intensity is proportional to the substrate current. For new technologies, the substrate current increases with increasing gate voltage, which gives rise to emission intensity as shown in Figure 6.19. If the gate of a transistor is pulsed, the emission intensity increases at the rising edge of the applied pulse, stays constant during pulse and falls to zero again at the falling edge, which is illustrated in Figure 6.20.

In case of a CMOS inverter, there is no photon emission at high and low states, since there is no current flowing. Photon emission occurs during switching. This mechanism is used in order to localize the buffers and inverters on the DUT.

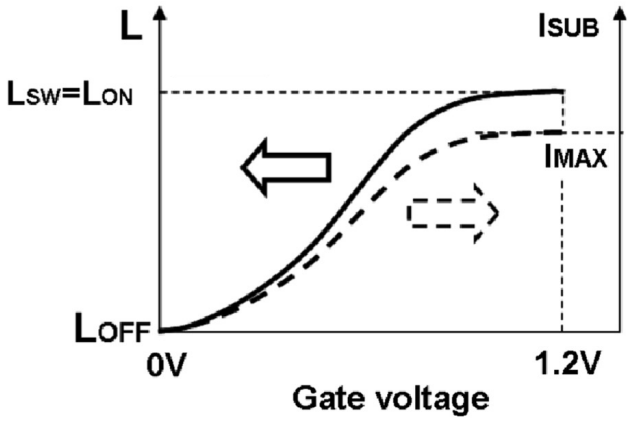


Figure 6.19. Substrate current and luminescence intensity versus voltage [LGB08]

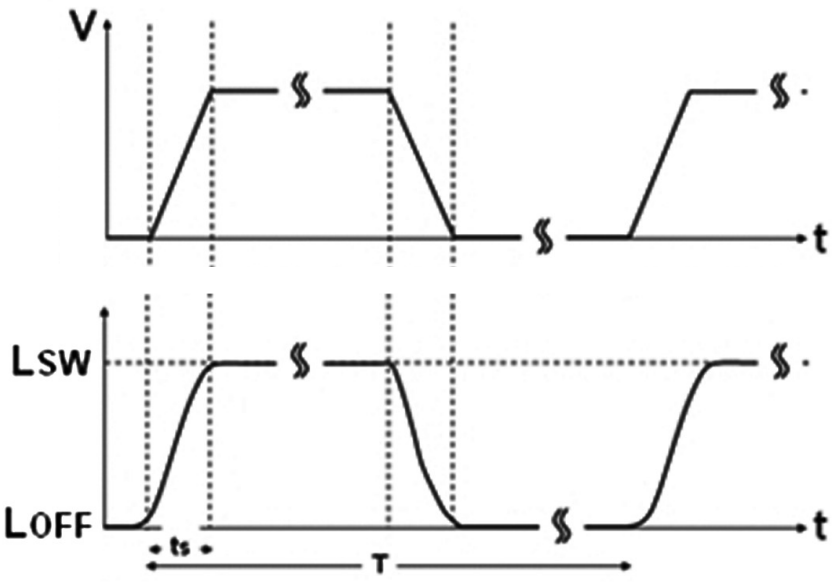


Figure 6.20. Luminescence intensity related to the device operation (a) Basic electrical force (b) Luminescence intensity [LGB08]

7. Localization of the Buffers on the Device

Initially, we didn't know where the inverters were situated on DUT. The inverters and the buffers may be spread all around the device. Therefore, it was mandatory to localize them on the chip at first. We used photon emission detection technique to find the inverters.

7.1. Photon Emission Experimental Set up

The photon emission measurements are carried out by Hamamatsu photon emission microscopy, PHEMOS 1000. The system is equipped with a cooled Si-CCD detector. The experimental set up is seen in Figure 7.1. Static Si-CCD camera is used to detect the photon emission that is integrated over time for 500s. The DUT is driven with a signal frequency of 250 MHz by a pulse generator and the device is biased at 1.5 V since the photon emission is not strong enough to be detected at 1.2 V nominal supply voltage. 50x optical lens is used for this experiment.

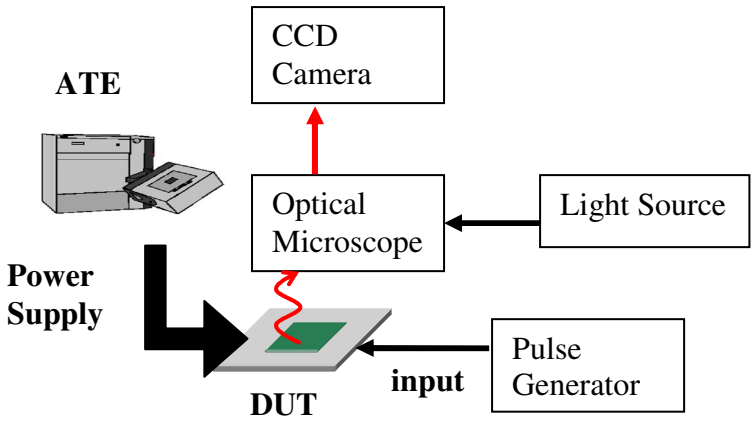


Figure 7.1. Experimental set-up for the localization of the inverters by cooled CCD camera

7.2. Results

As a result, two buffers are localized. The photon emission sites related to the buffers are shown in Figure 7.2. The emission of those that are very close to each other is merged in the photon emission image. These buffers are used for laser stimulation investigations on delay chains.

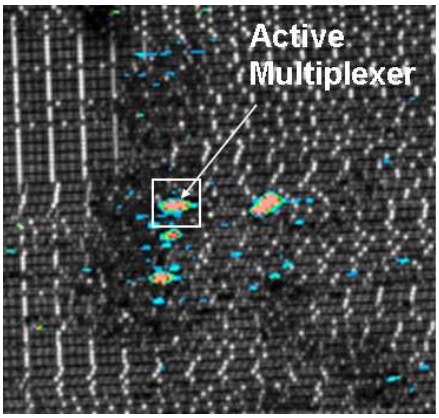


Figure 7.2. Photon emission image of the buffer acquired by 50x lens (VDD=1.5 V, f=250 MHz, integration time=500 s)

8. Laser Stimulation on the Delay Chain

Investigation of the effects of laser stimulation on basic logic gates such as buffers is carried out. Heat-build up at a gate or localized photocurrent induced by laser beam may alter the switching time of a gate through alteration of some electrical transistor characteristics, mainly the drain current. The effect of temperature and localized current on a FET transistor is discussed in detail in Chapter 6. Since the propagation delay of a gate strongly depends on the drain current of nMOS and pMOS transistors, it is possible to tune the delay of an inverter by laser stimulation.

Analysis of the timing properties of a device can be accomplished through perturbation of the transistors by laser stimulation, which alters the switching properties of the transistors. Therefore, we applied laser stimulation technique on the most basic logic structure, buffer.

8.1. Experimental Set-ups

Principally, we use PHEMOS 1000 as a laser stimulation system for our experiments. Both 1.064 μm and 1.3 μm wavelength laser diodes are used in order to evaluate the device response to thermal and photoelectric effects. 1.064 μm laser is utilized in two modes: CW and pulsed mode.

Laser stimulation of the buffers from the backside causes delay variations in the range of pico seconds that are hard to detect. Therefore, a few different set-up configurations were built in order to acquire the strongest signal with a minimum acquisition time.

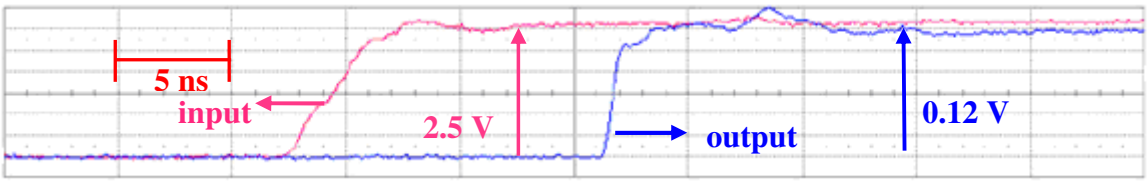
ATE is used for some set-up configurations just to power up the device and sometimes also to apply the input stimuli. We employed PHEMOS 1000 as LSM system. Basically, the DUT is driven by tester while the LSM scans pixel by pixel. Therefore, the DUT is placed in PHEMOS 1000. Data transfer between tester and PHEMOS 1000 is done by flat coaxial data cables. These cables carry input, output or control signals. Depending on the length, they have around 10 ns transmission time.

Generally, a DUT board, also called Motherboard, is required to interface the DUT to the ATE. DUT board simply connects DUT pins to the ATE channels. But, we have a separate LSM tool and it is not possible to dock the big testhead into LSM. Therefore, two HP motherboards are used, one in PHEMOS 1000 and the other one on the testhead for all of the experiments. The connection between two motherboards is done via flat coaxial cables. One of the motherboard is used as an interface between DUT and the data cables and it is placed in PHEMOS 1000. The other end of the cables is connected to the second motherboard, which is docked on to ATE testhead.

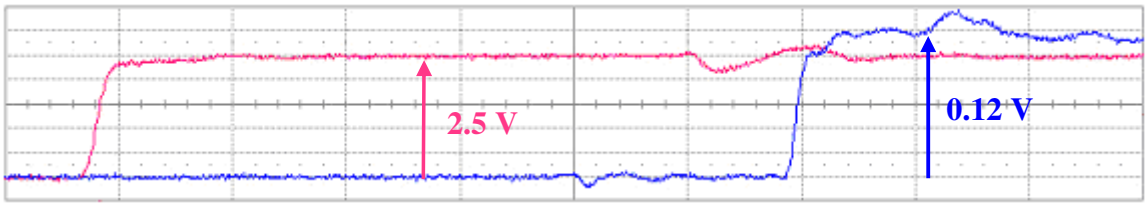
In order to detect pico second delay changes, the signal transfer from device to the measurement unit should be done carefully. The reflections should be avoided. Therefore, the input and the output cables should be terminated with 50 Ω at the receiving end or at the source. For critical signals like input and output, we used special low loss cables to improve the quality of the signals. These cables have 95 dB attenuation per 100 m at 6 GHz.

In addition, the rising and the falling times should be small. When we drive the device with the tester, a reflection at the input signal was observed. Therefore, we utilized a pulse generator which has 1 ns rising time. The comparison of the signals for tester driven and pulse generator driven cases is seen in Figure 8.1. When the device is driven by tester, the rising time of the input signal is 3.9 ns and in the case of pulse generator it is 1.08 ns, although there is no such a big difference in case of output signal. Since the input impedance of oscilloscope is 50 Ω, which is too low for device to drive, a 1 KΩ resistance is connected serial to the

output pin. The rise time for output signal is 1.5 ns. The time delay difference between two plots is caused by the delay of the cables.



(a)



(b)

Figure 8.1. Input and output of the delay chain (a) Tester driven (b) Pulse generator driven (A 1 KΩ resistance is connected between scope channel and the output of the device)

Another problem that we encountered during our experiments is that there is mechanical drift at PHEMOS 1000 laser scanner system. The microscope moves 2 pixels (1 µm) in every 5 minutes in vertical direction. It is caused by the heat generation on the motors, which control the movements of the microscope. It means that if the scanning time is greater than 2.5 minutes, the scanned area will be shifted one pixel in vertical direction. Therefore, the acquisition time should be kept low. To achieve low acquisition time, the sensitivity of the delay time measurement units should be high.

The experimental set-ups (initial and improved) for the delay variation detection and how we have overcome above mentioned problems are explained in more detail below. Initially, we used tester for the pass/fail evaluation and for measuring the delay variation induced by laser. However, the acquisition time was too slow and the detection was not sensitive enough. Therefore, we utilized a delay measurement unit that has a high time resolution, which is explained in the improved set-up section.

8.1.1. Initial Set-up

HP 83000 ATE so called tester is initially used to drive the device and to measure the output responses. Employing the tester allows taking the advantage of its great capabilities. It has many advantages in driving the device inputs and measuring the device responses.

First of all, the tester has a large number of tester channels, which means that a device with many input and output pins can be tested easily. Because of its big vector memory, big vector patterns can be applied to the devices for testing. Besides, it gives a big freedom in arranging the input waveforms. The timing of each pin can be set easily. For some laser stimulation applications, extra tester channels may be needed to produce a signal, e.g. to establish synchronization between the tester and the LSM, to feed the LSM with the pass/fail information from the tester or to trigger the laser diode. The timing of the laser pulse trigger can be adjusted easily in respect to the input or clock signals.

From the measurement point of view, the tester is able to run many different kinds of

tests. It can execute dc tests such as leakage current test, standby current test, contact test or ac tests such as functional test, propagation delay, hold-time, set-up time, jitter etc. Therefore, the analysis with the tester makes it possible to obtain detailed quantitative analysis pixel by pixel while scanning the devices. In other words, the localization and the quantitative measurements can be accomplished simultaneously.

Due to all above mentioned advantages, we employed tester to drive the input pins of the DUT and to evaluate the responses from the DUT, initially.

I. Read out of Laser Stimulation as Pass/Fail

As a beginning, we did pass/fail evaluations. It is based on the IC functional test variations (pass or fail). The device inputs are driven with an input pattern and a functional test is applied. Tester compares the real values of the outputs with the expected values and makes a pass or fail decision. The goal is to run the functional test in a loop and collect pass or fail data per laser beam position while the laser is scanning the device. Functional test compares measured outputs with expected ones for all input configurations chosen by the user. The final result is either PASS or FAIL when one or more differences are found. The supply voltage is reduced until the operating point of the device is close to the pass/fail border. If the device works close to that border in pass or fail region, a very little stimulation will be enough to change from pass to fail or from fail to pass. Actually, the hardest part of this experiment is the proper adjustment of the output strobe. Output strobe determines the moment in time, when the measurement occurs. If it is set too early, the data will not be ready and the measured value will be equal to the previous expected value, which leads to FAIL.

Under laser stimulation, a failing device may result PASS and a good device may fail if the output strobe is set to a critical point. An illustration for FAIL to PASS transition caused by laser stimulation is shown in Figure 8.2. The output strobe is set a little bit early and the status is FAIL. When the laser illumination is present, the delay gets shorter which causes final result to be PASS. Because, laser illumination has an impact on the delay of the device as explained in Section 6.1.3. If the output strobe were set too early, FAIL-to-PASS transition would not be recognized.

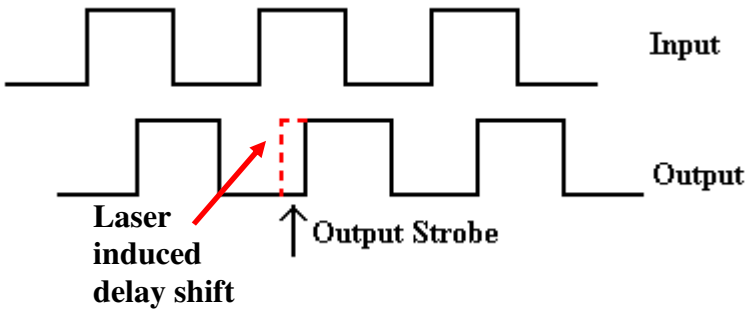


Figure 8.2. The illustration of the signals for the fail-to-pass experiment

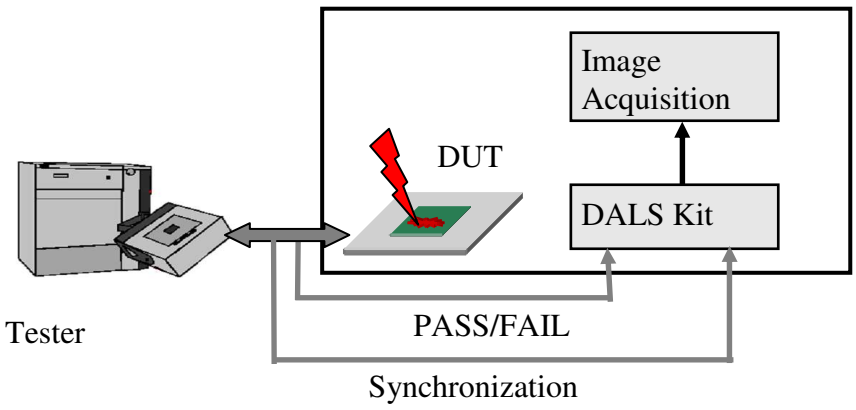


Figure 8.3. Experimental set-up for pass/fail evaluations

Transfer of Tester Pass/Fail Information to PHEMOS image processing card

The experimental set-up is shown in Figure 8.3. The DUT is driven by ATE via coaxial cables. The key is to acquire the pass/fail status from the tester and feed it to the PHEMOS 1000 video card with proper voltage levels. Therefore, two channels are allocated at the tester: one for pass/fail information and one for synchronization. Those signals are fed into DALS Kit of the PHEMOS 1000.

The test flow is seen in Table 8.1. The vector pattern runs in a loop. We used a conditional jump instruction for this loop. The beginning and the end of a group of consecutive vectors are marked with MATCH and MATCHEND commands. The system repeats the loop until a match is found. If a match is found, it goes out of the match loop and then the tester continues to execute vectors as a linear system. If there is at least one failing cycle inside the match loop, the match loop starts again until functional test inside the match loop passes. Pass/fail signal (P/F) is programmed to be logic 0 (indicating fail) inside the match loop and logic 1 (pass) outside the loop.

Table 8.1. The test flow at tester for PASS/FAIL experiments

Sync.	P/F	DIO 0 (Input)	Control Signals			DIO 56 (Output)	…...	Command line
1	0	X	X	X	X	X	…...	LOOP
1	0	0	1	1	1	0	…..	Match
1	0	1	1	0	0	1	…...	
1	0	.	.	.	.	.		
1	0	.	.	.	.	.		
.	0	.	.	.	.	.		
.	0	.	.	.	.	.		
.	0	.	.	.	.	.		
0	0	.	.	.	.	.		
0	0	.	.	.	.	.		Matchend
0	1	X	X	X	X	X	…...	Repeat, 1200
0	1	X	X	X	X	X	…...	LOOPEND

The match loop is put into another endless loop. When the test is over, the match loop starts again from the beginning so that the synchronization is ensured. The synchronization between LSM and ATE is controlled by 'Synchronization' signal. When this signal switches from logic 1 to 0, the laser scanner moves to the next pixel.

If the P/F signal is logic 1 within one pixel, then the related pixel is colored with red. The result image shows the spots where a change from pass to fail or from fail to pass occurs. Later on, this image is overlaid over the reflected image of the device and the sensitive areas are identified.

This method has some disadvantages. First of all, the accuracy of strobe placements at tester has some timing limitations. Moreover, some effort should be paid in order to find the right parameters (i.e. VDD, frequency, temperature etc.) to have P/F regions. Besides, only the significant delay variations can be detected and it is only possible to make qualitative measurements with this method.

The method of delay read out as pass/fail is a great opportunity to localize the performance limiting circuitry in a device. Despite that, measuring directly the delay gives a more detailed view of the device in the means of timing analysis.

II. Delay Recognition by Tester

HP 83000 ATE is used to drive the device and to measure the output responses for our experiments. The data transfer between tester and PHEMOS 1000 is done by coaxial cables. These cables carry input, output and control signals. The experimental set up is shown in Figure 8.4.

In order to accomplish measurements with tester, synchronization should be established between the laser scanner and the tester. While the laser is on one pixel, the tester sends the test sequence to the device, measures the input-output delay, writes the result in a file and then produces a signal to move the laser scanner to the next pixel. Finally, this file is used to acquire the delay variation graph. The test program that performs all these jobs for each pixel is written at tester. This test flow is shown in Figure 8.5.

Two tester channels are allocated for two control signals, "Synchronization" and "FrameOrLoopTrigger". The first signal is used for synchronization of PHEMOS 1000 and ATE. After the measurement is completed for a certain pixel, this control signal is pulled from logic '1' to '0', so that the scanner moves to the next pixel.

One should take dummy pixels into account while building the synchronization. The "FrameOrLoopTrigger" control signal is assigned for that job. This signal is used in order to check if the laser scanner is on a dummy pixel or not. "Line Trigger" and "Frame Trigger" signal outputs of the DALS kit can be used to recognize the dummy pixels. The voltage levels are pulled to logic '1' if the scanner is on a dummy pixel, otherwise they stay at logic '0'.

"Line Trigger" and "Frame Trigger" signals are ORed and the output of this OR gate is fed into FrameorLoopTrigger channel at the tester. If the output of the OR gate is '1', it indicates that the scanner is on a dummy pixel and the delay is not measured for that pixel. If it is '0', the delay is measured and the values are stored in a file. This file is used to build the delay variation mapping later by using MATLAB. Thus, the image processing unit of the Hamamatsu system is not used.

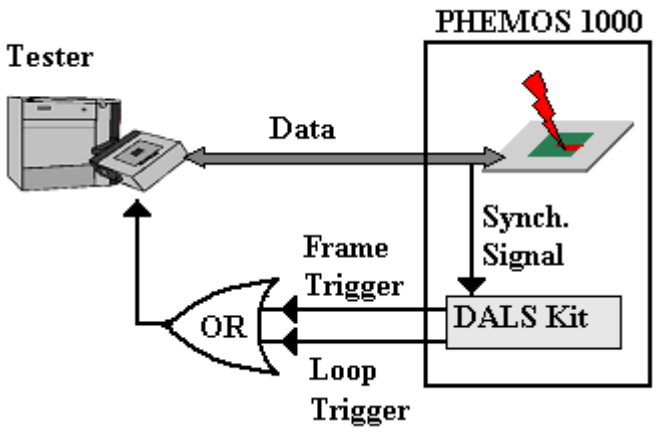


Figure 8.4. Experimental set-up for delay recognition by tester configuration

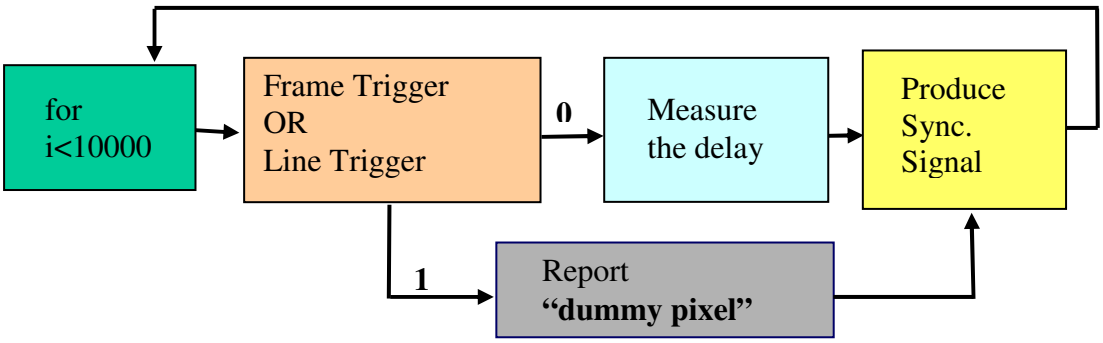


Figure 8.5. Delay measurement test flow at tester

Using tester as a measurement unit has its advantages and disadvantages. One advantage is that tester has many capabilities. One may measure not only delay, but also e.g. set-up time, hold-time, jitter, quiescent current etc. "Measure the delay" command in the test flow seen in Figure 8.5 can be replaced by any of these test functions. Moreover, quantitative results are obtained when a tester is used as a measurement unit. The amount of delay variation induced by laser and the coordinates of that spot where the delay variation occurs can be seen in the same plot.

In addition to its advantages, tester approach has also some disadvantages. Our tester has an overall timing accuracy of 300 ps. This timing resolution is not suitable for detecting tens of pico second delay changes caused by laser stimulation. In order to overcome that issue, one should scan the chosen area at least 5 times and take the average. But that will increase the acquisition time.

Besides, there is a mechanical drift at our PHEMOS 1000 system as explained above. The microscope moves 2 pixels (1 µm) in every 5 minutes in vertical direction. Therefore, the acquisition time should be as fast as possible. However, scanning speed has to be very slow while the delay read out is grabbed by tester. Because, for each pixel the tester needs around 300 ms to recognize the dummy pixels, measure the delay, save them in a file and to produce the synchronization signal. That means that it takes 25 minutes to get a 5 times integrated

32x32 pixel image. Considering the mechanical drift, the image will be shifted around 10 pixels by the end of the experiment.

8.1.2. Improved Set-up

We improved our setup in order to have fast acquisition time and high timing resolution. We built a few different set-ups, which are explained below.

The laser illumination produces very little changes on the devices, which may be hard to observe, e.g. delay variations in the range of pico seconds at the gate of an inverter. Therefore, a proper measurement setup is needed for the observations as explained above. We took a measurement unit called Time to Amplitude Converter (TAC), which makes faster evaluations and has a higher timing resolution than tester. We built two set-up configurations with the TAC. At first, we connected TAC output directly to the DALS kit of the PHEMOS 1000 system and in the second case, we used lock-in amplifier combined with the TAC in order to get a better SNR. Moreover, lock-in amplifier is employed in order to detect the phase changes between input and output.

I. Delay Recognition by External TAC

A TAC is a measurement unit, which measures the time interval between two pulses applied to its START and STOP inputs and generates an output pulse whose amplitude is proportional to the measured time interval. The threshold voltage is 2 V for both inputs. When the “Start” input signal exceeds the threshold voltage, the conversion process is triggered until a valid “Stop” signal is accepted. Therefore, TAC is sensitive only for the variations at the rising edges. We used an ORTEC Model 566 TAC unit. It can measure the time interval between pulses to its start and stop inputs with a maximum speed of 277 KHz and 5.5 ps resolution. Output amplitude - time delay graph for TAC is seen in Figure 8.6. The amplitude change is around 0.19 V/ns.

The experimental set-up is shown in Figure 8.7. The input and the output of the delay chain drive the START and STOP input terminals of the TAC, respectively. The output of the TAC is fed into *Tester Signal* connector of the DALS Kit. The circuitry in the DALS Kit takes the average of the TAC output signal per pixel and this integrated signal is used to build up a gray image by the image acquisition card of PHEMOS 1000, while the laser is scanning a certain area. In case the pulsed laser is used, the laser trigger is produced by tester and fed into the laser diode box.

Consequently, the acquisition time and the timing resolution is improved more than 90% compared to tester approach.

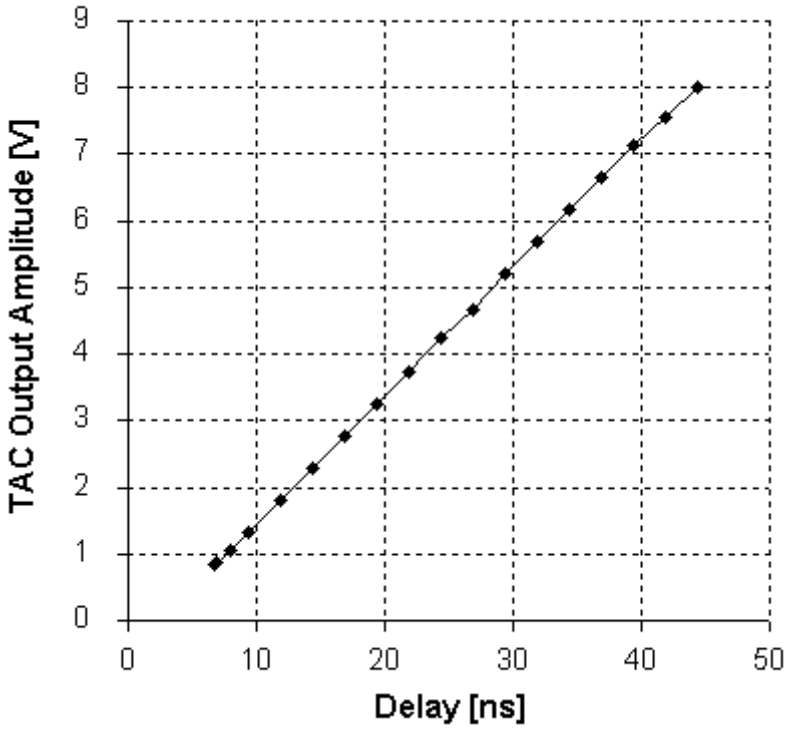


Figure 8.6. TAC output amplitude versus delay

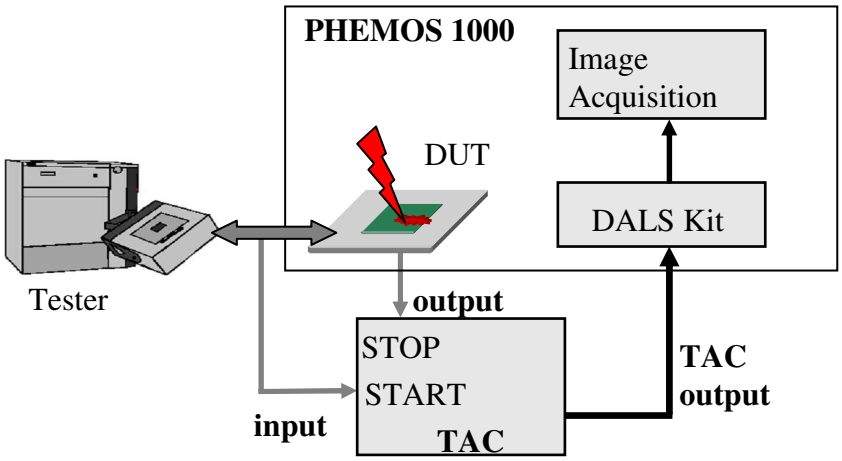


Figure 8.7. Experimental set-up with external TAC

II. Delay Recognition by TAC and Lock-in Amplifier

A high sensitivity Lock-in Amplifier (Signal Recovery 7280) together with TAC is used for delay recognition, in order to improve the acquisition time of the experiments and the sensitivity of the measurements. The experimental set-up is illustrated in Figure 8.8. The output of the TAC is fed into input channel of the Lock-in Amplifier and Valid Start signal is taken from TAC as a reference signal.

The aim of using lock-in amplifier for our experiments is to acquire the amplitude of the TAC output from the Lock-in Amplifier. Therefore, a reference signal is needed. TAC has three output terminals: TAC output, Valid Start and Valid Conversion. The duration of the Valid Start output indicates the interval from the accepted start until the end of reset. Valid Conversion occurs from the end of the internal delay after stop to the end of reset. They both have a fixed +5 V amplitude. The Valid Start signal can be taken as a reference signal, because this signal is not affected by laser stimulation at all. It depends on the start signal and the internal reset of the TAC, which are fixed values. The TAC output signal, Valid Start and the product of the two signals are scoped and shown in Figure 8.9. The product of these signals has zero phase with the TAC output that is desired and it has 5 times more amplitude than the TAC output. With this approach, the acquisition time is improved by 20% compared to only TAC set-up.

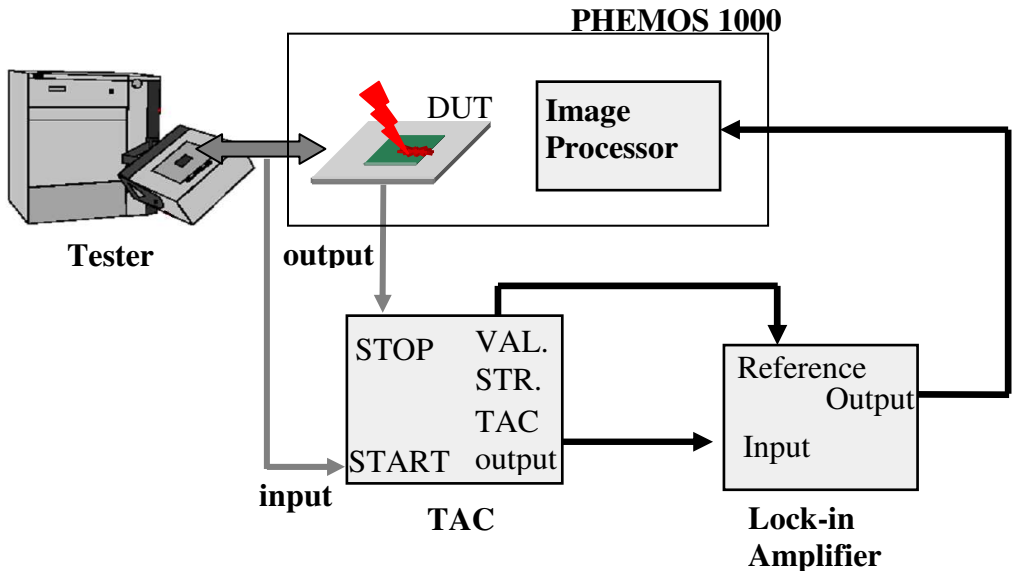


Figure 8.8. TAC + Lock-in amplifier experimental setup

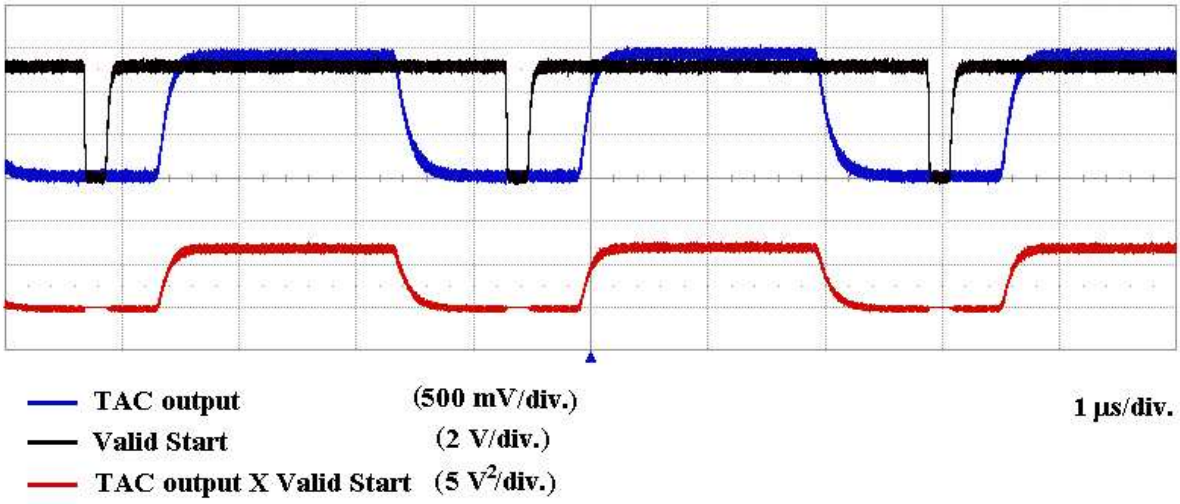


Figure 8.9. TAC output, Valid Start and the product of those two signals measured by oscilloscope

III. Delay Recognition by Lock-in Amplifier

Delay Recognition by lock-in amplifier method is based on detecting the phase changes between input and output of a delay chain that is caused by perturbation of the delay chain output.

The lock-in amplifier is a phase sensitive measurement unit. The output voltage of the amplifier consists of the phase difference between the input and the reference signal and the product of both signals amplitudes. The lock-in amplifier can be programmed to detect only the phase differences, which is implemented for these experiments. Any delay shift on the rising or falling edge of the output, induced by laser stimulation, will cause a phase shift, consequently changing the lock-in amplifier output. Therefore, this method is sensitive to the delay changes both at the rising and the falling edge. These changes can be converted into a gray image by the image acquisition card of the PHEMOS 1000 system via DALS Kit.

Consider the case where a noise-free sinusoidal signal voltage V_{in} is being detected at the lock-in amplifier,

$$V_{in} = A\cos(w \cdot t) \tag{8.1}$$

where w is the angular frequency of the signal which is related to the frequency, f, in hertz by the equality:

$$w = 2\pi f \tag{8.2}$$

The lock-in amplifier is supplied with a reference signal at frequency f:

$$V_{ref} = B\cos(wt + \theta) \tag{8.3}$$

where θ is the phase-shift between input and output channel. The detection process consists of multiplying these two components together so that the PSD output voltage is given by:

$$V_{psd} = A\cos(wt) \cdot B\cos(wt + \theta) \tag{8.4}$$

$$= 1/2\, AB\cos\theta + 1/2\, AB\cos(2wt + \theta)$$

After multiplication, this signal is applied to a low pass filter that removes the $2wt$ component. If the magnitude, B, of the reference frequency is kept constant, then the output from the phase-sensitive detector is a DC signal which is proportional to the magnitude of the input signal A and proportional to the cosine of the angle, θ, between it and the reference signal [Sig00].

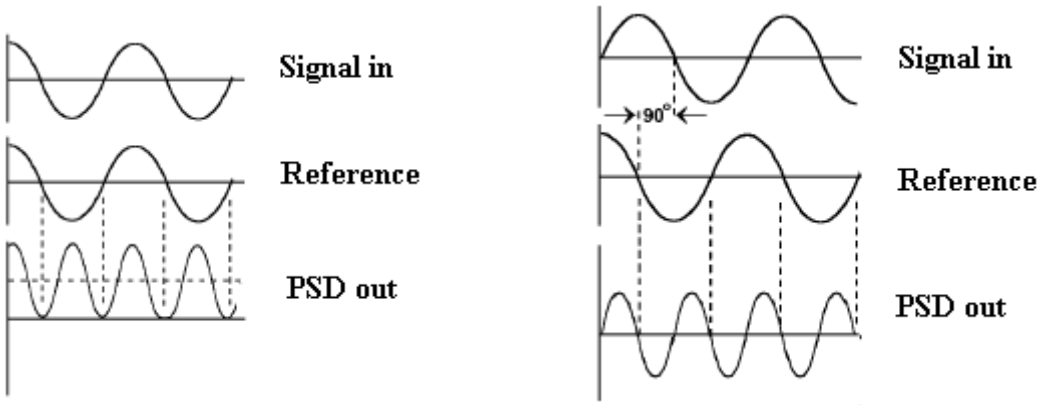
$$V_{out} = 1/2\, AB\cos\theta \tag{8.5}$$

The multiplication of two signals with 0^0 and 90^0 phase-shift are illustrated in Figure 8.10.

In our case, the signal applied to the reference and the input channels of lock-in amplifier is a pulse signal, which can be expressed by:

$$V_{IC}(t) = \sum_{n=0}^{\infty} \left[c_n \cos(nwt + \theta_n) \right] \tag{8.6}$$

Nevertheless, the lock-in amplifier produces a signal, which is proportional to the phase between input and output signals of the delay chain. This signal is used to build the grey image at PHEMOS 1000, which pinpoints the speed up and slow down structures on the DUT.



(a) (b)

Figure 8.10. Principle of Phase Sensitive Detector (PSD) for a (a) 0^0 and (b) 90^0 signal in-reference phase shift [Sig00]

The experimental set-up is shown in Figure 8.11. The input to the delay chain is used as a reference signal for the lock-in amplifier, as well. The output of the delay chain is applied to the input channel of the lock-in amplifier as a signal under investigation. Since the delay changes are very small, we need to drive the DUT with a high frequency input signal in order to collect as much information as much per pixel. Moreover, the lock-in amplifier should be able to process fast switching signals, as well. Therefore, we utilized a Stanford SR844 200 MHz digital lock-in amplifier and a Fluke PM 5786B 125 MHz pulse generator.

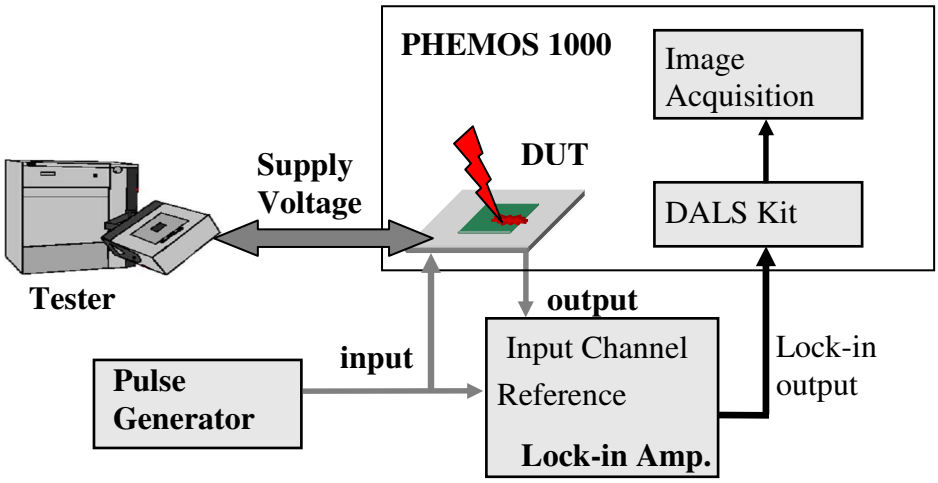


Figure 8.11. The experimental set-up with Lock-in Amplifier

If the phase shift between two multiplied signals is 90^0 or -90^0, then the output voltage of the lock-in amplifier will be 0. For the experiments, we set the phase shift between input channel and reference signal as -90^0 by using phase shifting function of the lock-in amplifier. When a perturbation occurs in the signal under investigation, the output from the lock-in amplifier will increase or decrease depending on an advance or delay. In our case, speed up in

the signal under investigation causes a brighter contrast whereas slow down darker contrast in the output image. The mixing operation of the input and the output signals, which have -90^0 phase shift is illustrated in Figure 8.12.

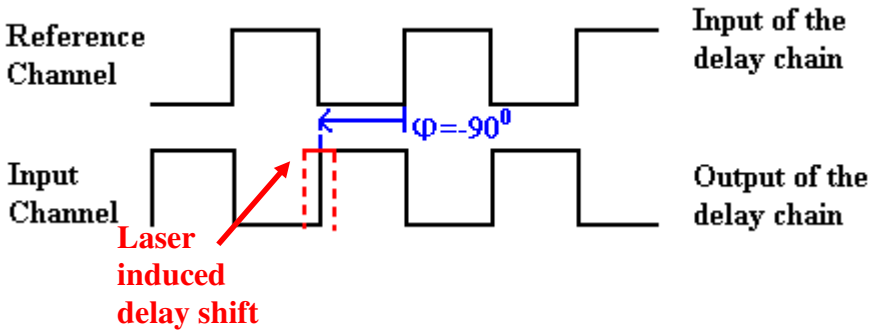


Figure 8.12. Mixing operation for input and output signals of the delay chain

IV. Delay Recognition by Oscilloscope

Although TAC and lock-in amplifier improves the timing resolution and the acquisition time of our experiments, it can be only used for qualitative measurements. The set-up with the TAC and/or lock-in amplifier is very suitable for scanning the device and grabbing the stimulation images, but it is not straightforward and accurate for quantitative measurements. In any case, an oscilloscope is needed to measure the output amplitude of TAC to have quantitative values. Thus, we used a high bandwidth, high sample rate LeCroy 8620A oscilloscope in order to measure the exact delay values. It has 6 GHz bandwidth and 20 GHz sampling rate. The input impedance of the oscilloscope channels is 50 Ω, so that the reflections are avoided. This measurement unit is a great opportunity to measure the delay variations introduced by laser. Besides, the oscilloscope allows us to study the shape of the output pulses in detail.

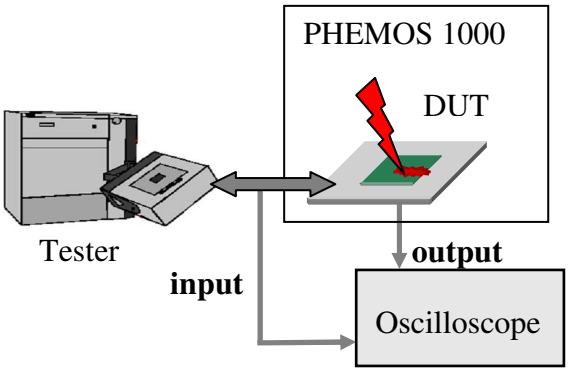


Figure 8.13. Experimental set-up for oscilloscope measurements

The experimental set-up can be seen in Figure 8.13. For oscilloscope measurements the laser is set to point mode. In order to scope the response of the device when a certain buffer or inverter in the delay chain is stimulated, the laser should be set to point mode. In this mode, the laser does not scan an area, but it can be placed on a desired point on the device. We analyzed the delay between input and output signals of the delay chain while a certain buffer

is illuminated by laser beam. In single point mode, it is possible to evaluate the laser stimulation intensity for each spot one by one. It is not possible with the scan mode.

8.2. Thermal Laser Stimulation on the Delay Chain

The delay variations caused by thermal laser stimulation are observed on 90 nm technology delay chain on DUT. PHEMOS 1000 system is used to scan the devices with laser. A few buffers and inverters are investigated under laser stimulation. But, the strongest stimulation occurred on the buffer, which includes also transmission gates in the same active area. The maximum laser power (20.6mW) and 50x objective is used for the experiments. While the device is being scanned pixel wise, the delays are measured using,

- Automatic Test Equipment
- Time to Amplitude Converter (TAC)
- TAC and Lock-in Amplifier
- High bandwidth oscilloscope

8.2.1. Automatic Test Equipment Measurements

The buffer on Symhony is stimulated using 1.3 µm continous wave laser beam. The tester was used to power-up the device and to drive the inputs of the delay chain as well as measuring the propagation delay. A pulse stream which has a 2 MHz frequency is applied to the DUT and supply voltage is kept at 1.2 V. The test flow at tester was used to acquire the outputs.

While a 32x32 pixel area which is shown in Figure 8.14 is scanned with the laser beam, the delay between input and output is measured for each pixel by tester and the delay values are stored in a file with the information of the laser beam position. Then, this file is used to obtain the delay mapping which is shown in Figure 8.15. The red pixels pinpoint a buffer (Buffer1) in the delay chain path.

The input-output delay is measured 5 times and the average of these values is calculated for each pixel. The scanning time per pixel is 300 ms which results an overall acquisiton time of 25 minutes.

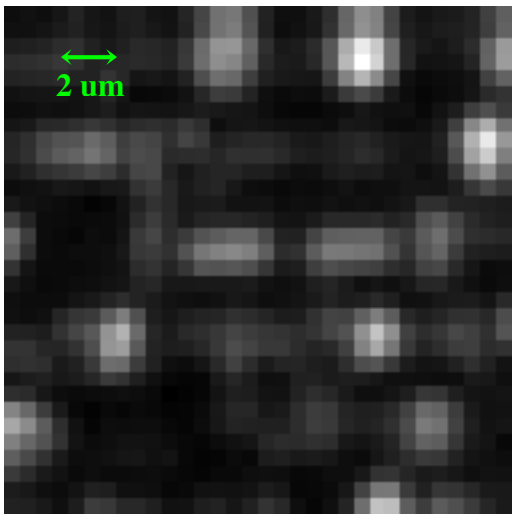


Figure 8.14. The reflected laser micrograph of the scanned area (32x32 pixel)

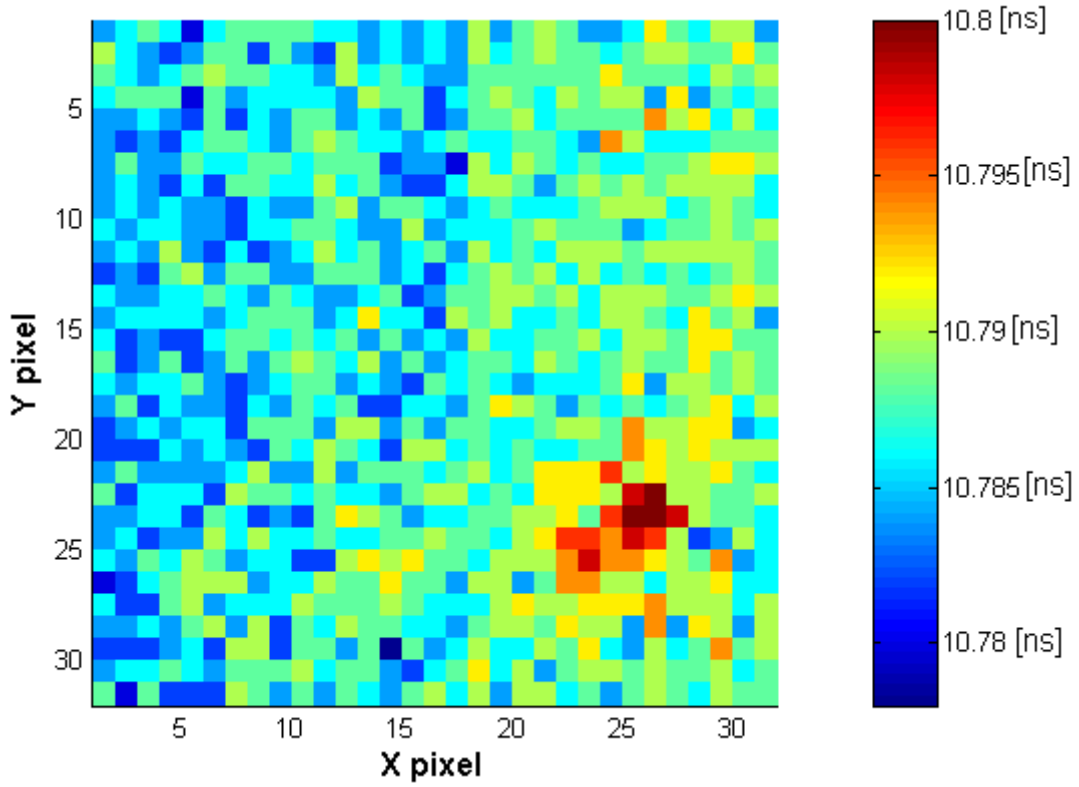


Figure 8.15. Propagation delay map that is obtained by tester under 1.3 µm CW laser beam irradiation (Laser power is 20.6 mW)

The results prove that thermal laser stimulation on Buffer1 causes a slow down in the delay chain at nominal supply voltage and the amount of delay introduced by laser beam is around 15 ps.

Instead of measuring the delay once at each pixel and scanning the selected area 5 times, the delay is measured 5 times at each pixel and then the laser moves to the next pixel, so that we overcome the problems that arise from mechanical drift and poor timing resolution.

It is obvious from the delay variation map that the overall delay gets larger through the end of the experiment. Because, the constant switching of the transistors causes heat build up in the device and as a result the delay increases approximately 4 ps. Therefore, the device is kept running for 25 minutes before starting the experiment.

8.2.2. Time to Amplitude Converter Measurements

The input and the output of the delay chain are fed into START and STOP terminals of the TAC, respectively. As a result; a stream of pulses is produced at the output of the TAC, whose amplitude is proportional to the time delay. The output of the TAC is connected to the DALS kit of the PHEMOS 1000, which integrates this signal. The reflected laser light is used to get the scanned area image and later the grey image is overlaid on the reflected light micrograph. The experiment is repeated for various supply voltages. Different scan speeds and integration numbers are applied for each supply voltage.

The parameters of the experiment are given below:

- 50x lens

- 32x32 pixel area scan
- 20.6 mW laser power
- 2 MHz input frequency of the delay chain
- 277 KHz TAC output frequency

The results are shown in Figure 8.16. The red spots indicate a slow down mechanism whereas the green spots a speed up.

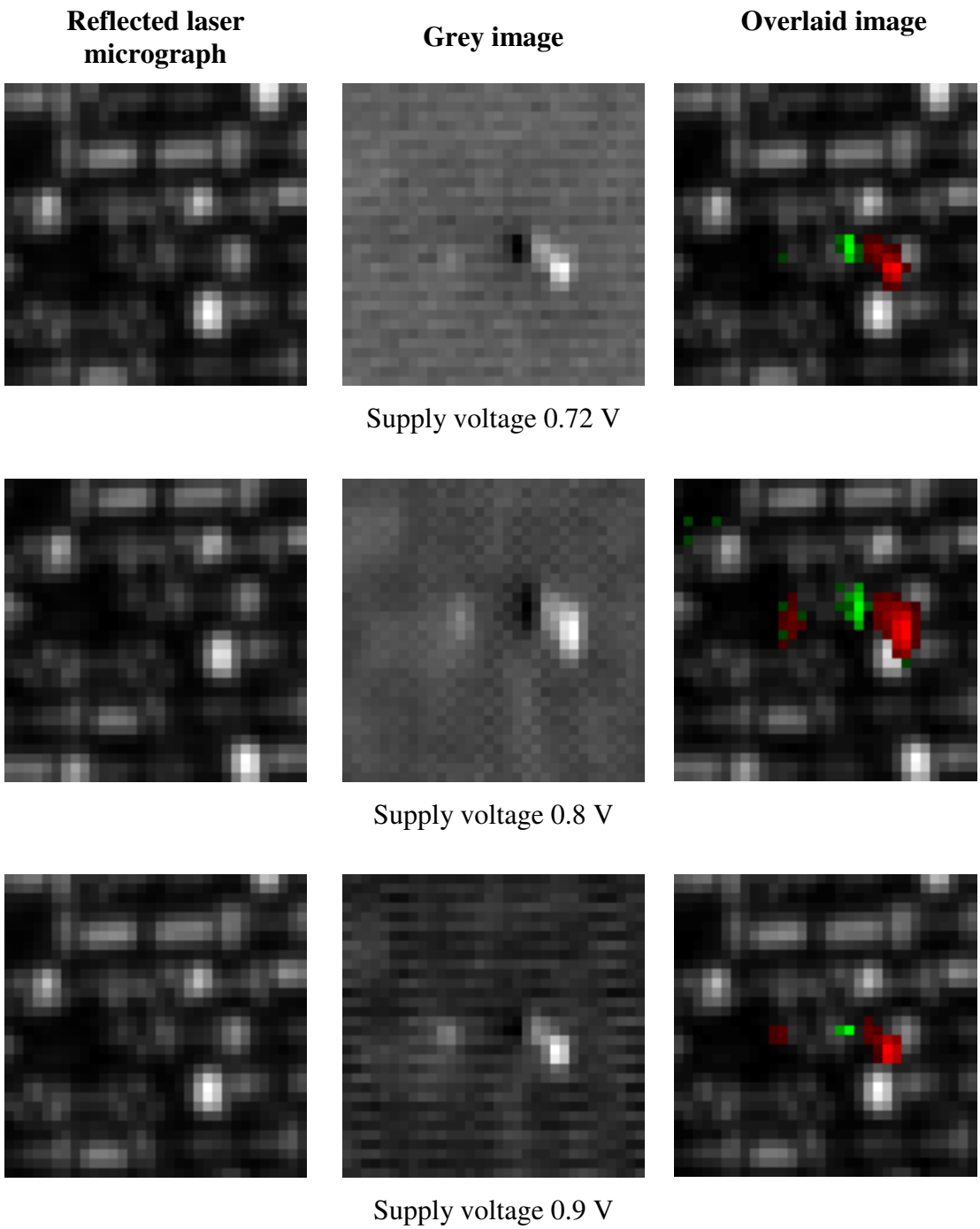


Figure 8.16. Laser stimulation of the selected buffer obtained by TAC measurement unit for various supply voltages

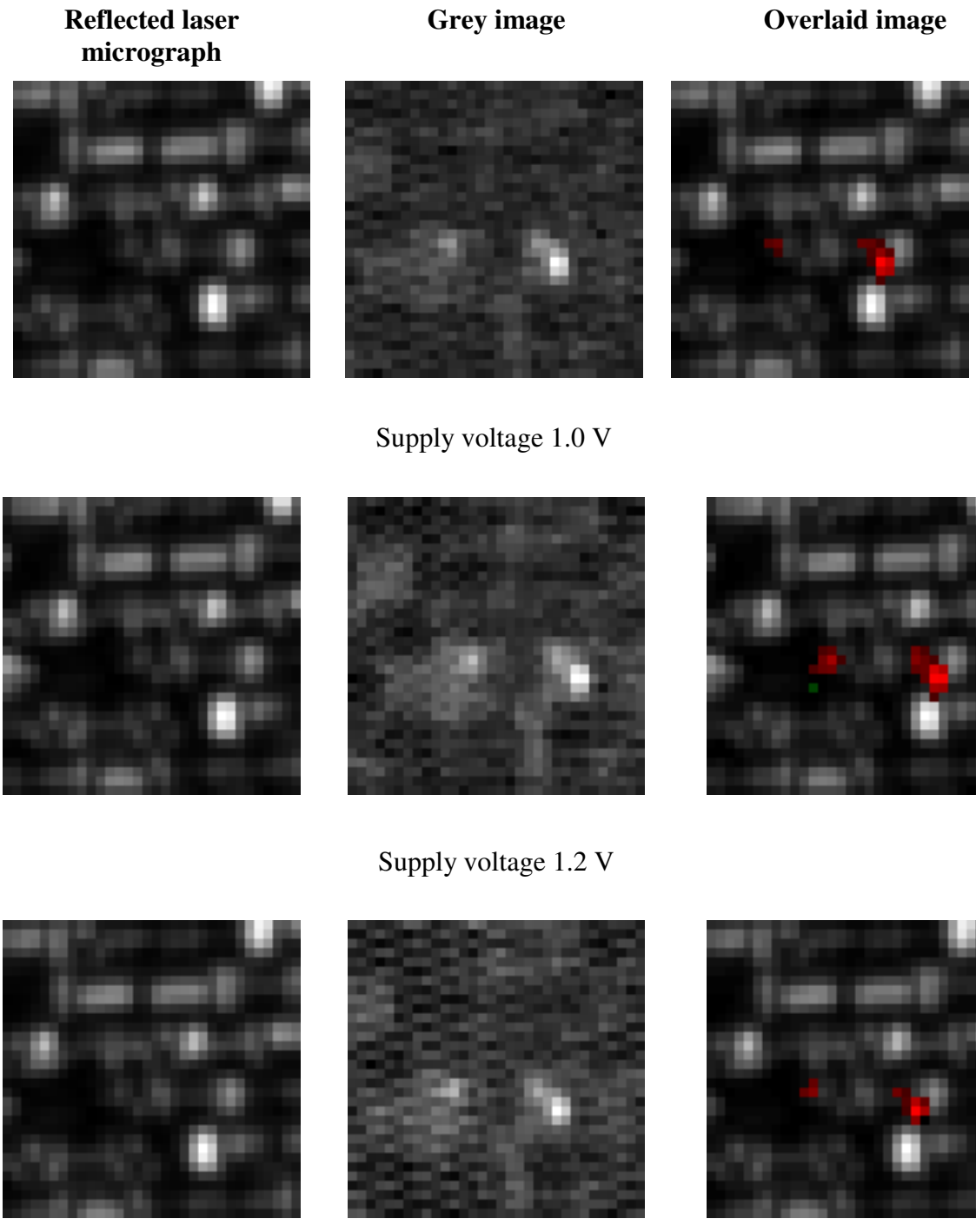


Figure 8.16. (continued) Laser stimulation of the selected buffer obtained by TAC measurement unit for various supply voltages

When the device is biased at 1.2 V, thermal laser stimulation on the buffer causes a delay increase. However, a green spot is observed on the nMOS-network when the supply voltage is lower than 0.9 V. We assume that it is either caused by heating of a transistor operating at a low gate voltage or it is a result of seebeck effect. As it is explained in section 6.1.1, when the transistor operates at a low gate voltage, the dominant effect is the threshold reduction due to

carrier density increase for higher temperatures. Threshold reduction increases drain current and higher drain current decreases the propagation delay.

The other possible reason is the seebeck voltage which is dominant for the transistors operating at low supply voltages. In case of laser stimulation of an ohmic resistor, seebeck effect is normally visible only when the device is biased at zero or close to zero. Because the resistivity change at higher supply voltages masks the seebeck effect. Nevertheless, the seebeck effect can be observed at supply voltages close to the nominal supply voltage, if the thermoelectric junction is on a gate conductor of a transistor. When the laser hits the gate of the transistor, seebeck voltage is generated. Because of the low current on the gate, the effect caused by resistivity change can be neglected.

8.2.3. Lock-in Amplifier Assisted TAC Measurements

As explained previously, the lock-in amplifier can be used together with the TAC to get a higher sensitivity. The output of the TAC is connected to the input of the Lock-in Amplifier. The reference signal is taken from the TAC, which has the same frequency with its output signal. As a result, the lock-in amplifier produces a DC signal whose voltage level is proportional to the amplitude of the TAC. The same 32x32 pixel area is illuminated with 1.3 μm CW laser beam. The results are shown in Figure 8.17.

The experiment parameters are as follows:

- 12s scanning time per frame
- 30 frames are obtained and integrated for the grey image
- 6 min. overall acquisition time
- 50x lens
- 32x32 pixel area scan
- 20.6 mW laser power

Reflected laser micrograph **Grey image** **Overlaid image**

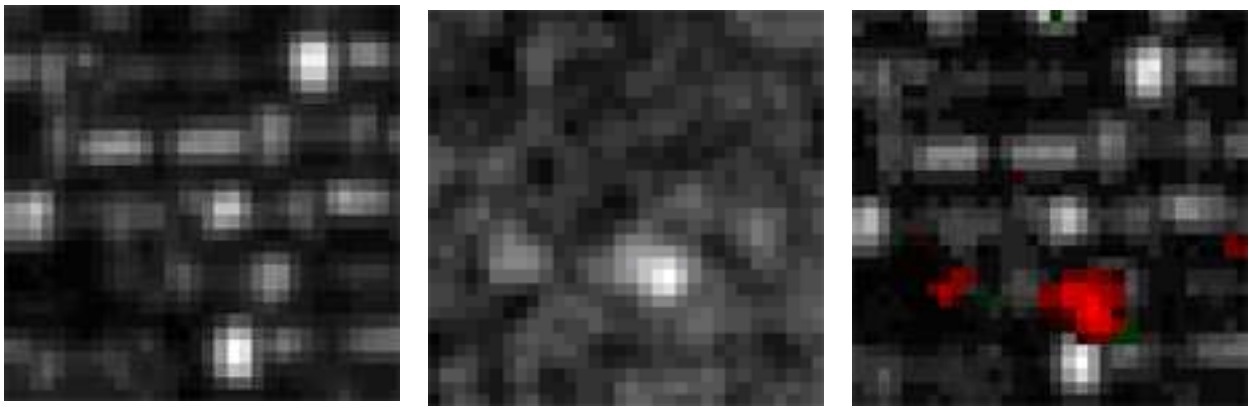


Figure 8.17. Laser stimulation on Buffer1 and Buffer2 obtained by TAC and Lock-in Amplifier (Supply voltage is 1.2 V and laser power is 20.6 mW)

As a result, with the lock-in assisted TAC configuration we get a stronger spot in shorter time. The overall acquisition time for TAC measurements is 7.18 minutes whereas it is 6

minutes for lock-in assisted TAC measurements.

8.2.4. Oscilloscope Measurements

The 6 GHz LeCroy oscilloscope is used for this experiment. The laser beam is positioned on Buffer1 and the variations in the delay time between the input and the output are measured. Only the red colored areas are analyzed. This experiment is repeated for various power supplies and the delay values are measured more than once for each bias condition. The results are shown in Figure 8.18. There is noise accompanied with the measured delay shift. Some of the reasons can be the deviations in the position of the laser beam or the signal jitter occurring at the output of the delay chain. It was observed that the delay shift caused by the laser perturbation increases with decreasing supply voltages and the curve follows a parabolic function.

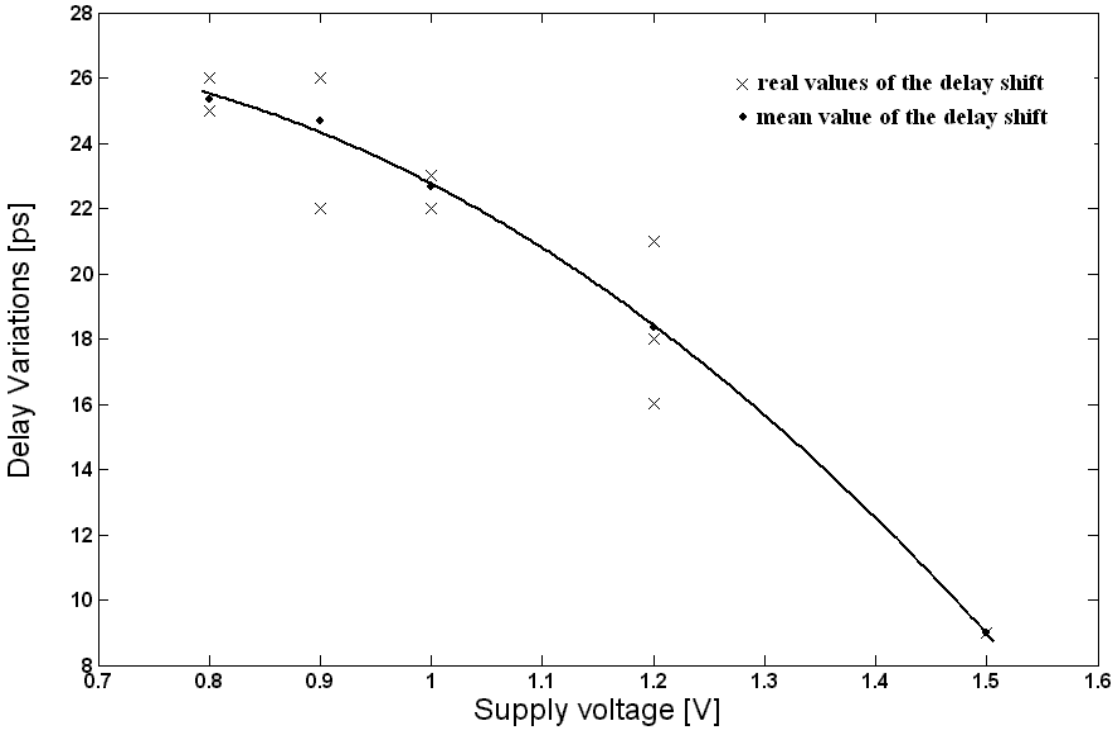


Figure 8.18. Delay shift observed on the Buffer1 for different supply voltages obtained by oscilloscope

8.3. Photoelectric Laser Stimulation

The devices can be stimulated with 1064 nm laser in two ways: either a CW laser or a pulsed laser can be utilized. We have both continuous and pulsed laser diodes installed in our system. There are two modes of operation for the pulsing method: pulse mode and impulse mode. Since impulse mode has too high power, there is the risk to damage the devices. Therefore, we use pulse mode for our experiments.

As explained above in TLS part, different measurement units are employed for the experiments.

8.3.1. Automatic Test Equipment Measurements

The Automatic Test Equipment is used to measure the delay variations versus the location

of the scanning laser microscope. The tester has an overall timing accuracy of 300 ps, which makes it harder to localize sensitive areas.

Moreover, the experiments with the pulsed laser have some difficulties. The maximum repetition rate of the pulsed laser is 50 KHz for the pulse mode and 10 KHz for impulse mode. So, the input pattern frequency can be maximum 50KHz, which slows down the acquisition time. Propagation delay measurement test function initiates a functional test and checks for a pass/fail transition. If none is found, the relevant timing is shifted again and the functional test runs again. By repeating this sequence, the test function performs a search. The vector pattern is sent to the DUT during each functional test. This process slows down the acquisition time even more.

If the acquisition time is too big, the drift of the microscope may cause severe consequences, such as the result image will be shifted towards vertical direction and the localization will be misleading. The microscope moves 2 pixels (1 μm) in every 5 minutes in vertical direction. One experiment takes around 25 minutes, which means that the reflected light image will be shifted around 10 pixels at the end of the experiment compared to the beginning.

Therefore, a Time to Amplitude Converter, which has a better resolution, is employed to extract laser sensitive areas.

8.3.2. Time to Amplitude Converter Measurements

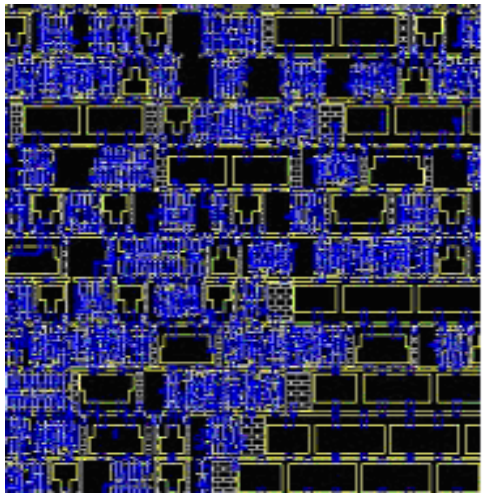
TAC is employed in order to analyze the delay variations of the delay chains. The result for 64x64 pixel size area laser stimulation is shown in Figure 8.19.

The experiment parameters are:

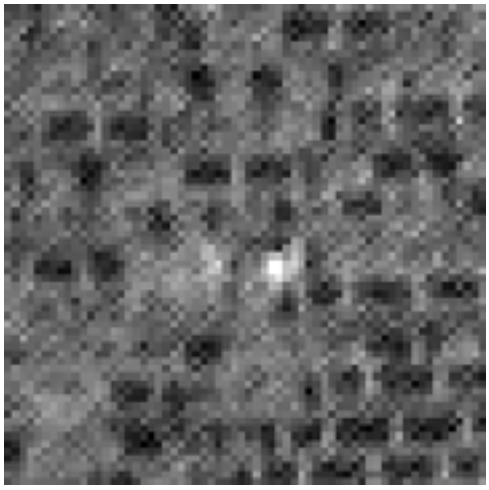
- 1.2 V of supply voltage
- 2 MHz input frequency
- 5.79 s/frame scanning time
- 300 frames were taken and integrated
- 277 KHz TAC output frequency
- 50x lens
- 32x32 pixel area scan
- 24 mW of laser power

The results for CW and pulsed laser stimulation on 32x32 pixel size area are shown in Figure 8.20 and Figure 8.22, respectively. The experiment is repeated for various supply voltages. Different scan speeds and integration numbers are applied for each bias voltage. The experiment parameters for CW laser are:

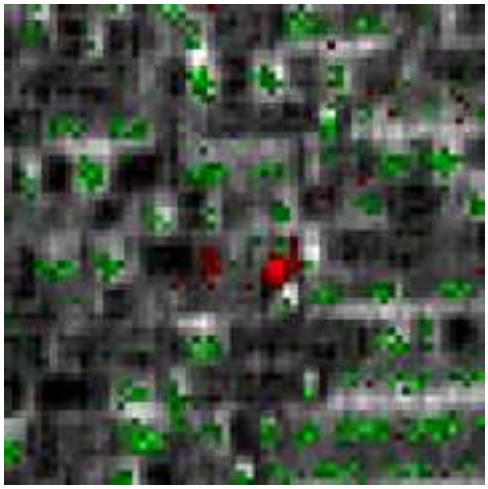
- 277 KHz TAC output frequency
- 50x lens
- 2.5 MHz input frequency
- 24 mW of laser power



Layout of the scanned area



Reflected laser micrograph



Overlaid image

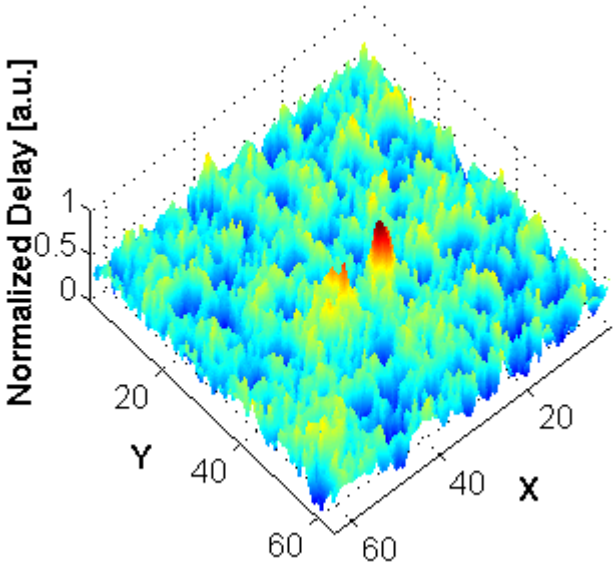


Figure 8.19. Delay Variations induced by CW laser illumination on a 64x64 pixel size area obtained by TAC measurement unit, the DUT biased at 1.2 V

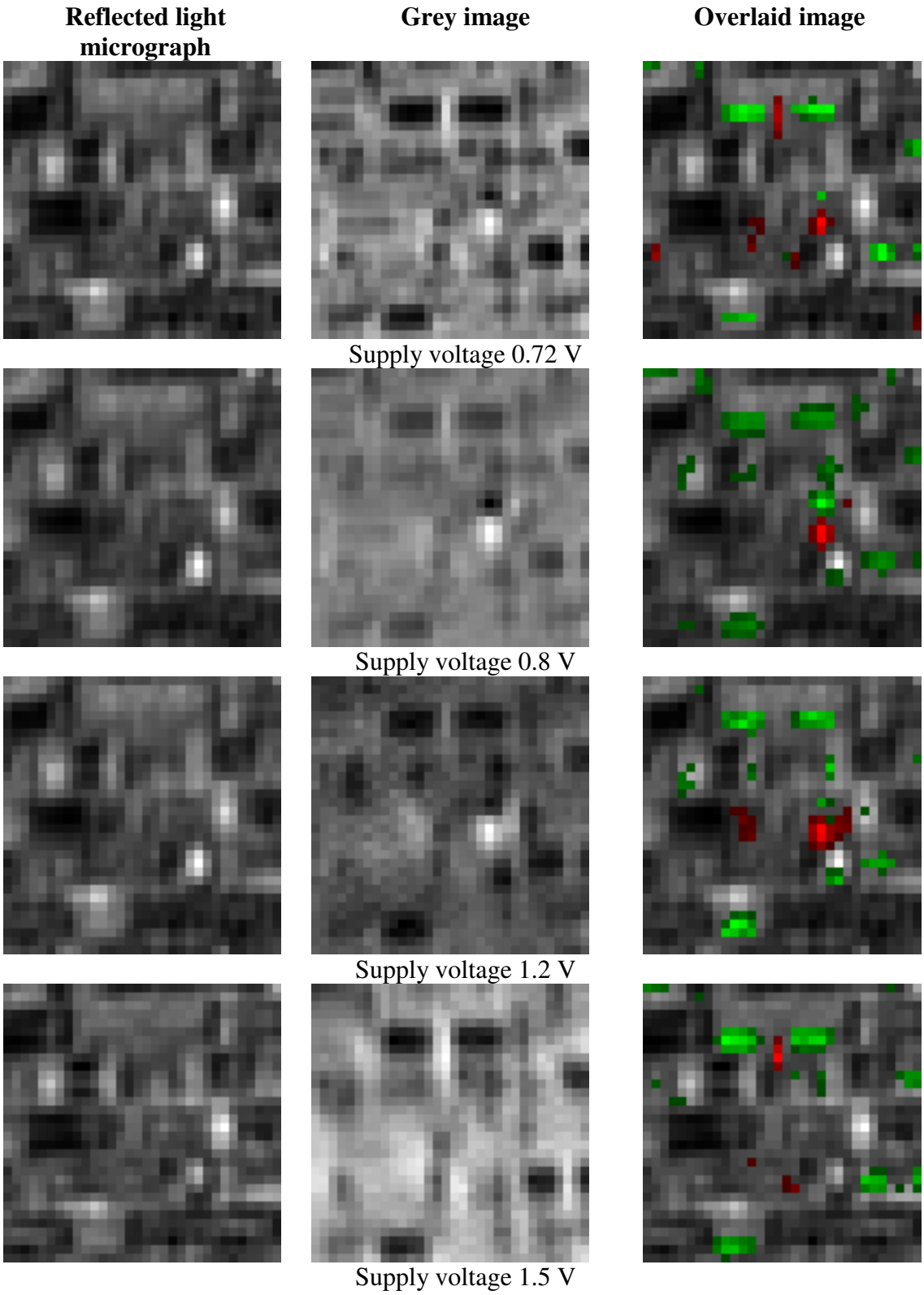


Figure 8.20. Delay Variations induced by CW laser illumination obtained by TAC measurement unit for different supply voltages

Since the TAC output frequency is lower for pulsed laser compared to CW laser, more frames are integrated. The parameters are:

- 2.5 MHz input frequency

- 50 KHz TAC output frequency
- 50x lens
- 32x32 pixel area scan
- 200 ns laser pulse width
- 24 mW of laser power

The laser pulse signal in respect to the input and the output signals are shown in Figure 8.21. The laser is triggered before the input and the output signal rising edges and it switches off before the falling edges.

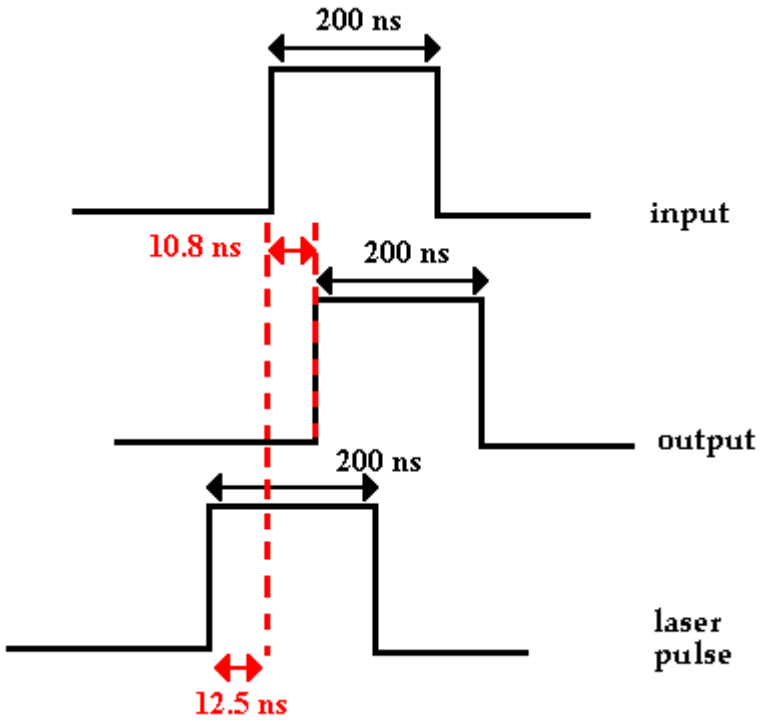


Figure 8.21. The laser pulse and the input and the output signals

Laser stimulation on Buffer1 appears as a black (green) spot on the nMOS transistors and as a white (red) spot on the pMOS transistors for continuous and pulsed laser as shown in Figure 8.20 and 8.22. It means that laser stimulation causes speed up for the nMOS transistors and slow down for the pMOS transistors in the operation speed of the inverter. The generated electron-hole pairs in the drain-well p-n junctions of the transistors change the drain current and therefore perturb the switching time of the delay chain. The theory of the photoelectric laser stimulation effects on the switching time of the inverters is explained in more detail in Chapter 6.

The intensity of the spots decreases with the increasing supply voltage, requiring slower scanning times or more integration.

The analysis of the scanned area showed that CW laser perturbs the speed of the delay chain even when it hits the filling structures, which are p-implanted structures (see Figure 8.19). Those structures are situated inside the same n-well as transistors and they are connected to the ground line. Therefore, they form a p-n junction. The generated electron-hole

pairs in the space charge region of these p-n junctions cause a photocurrent generation, which perturbs the switching time of the buffer. Laser incident on these structures decreases the propagation delay time.

The stimulation on the filling structures is not desired. The analysis with the pulsed laser shows that short laser pulses can suppress this effect. In pulsed operation, the laser can be switched on just on the edges of the input and the output signals and thus minimizing the secondary effects. For the supply voltages higher than 1.2 V, the pulsed laser stimulation on the filling structures is not detectable anymore. The only visible spots are the buffers and the inverters. Moreover, in case of 1.064µm CW laser beam and 1.5V supply voltage, there is no stimulation detected on the Buffer1. However, pulsed laser is able to localize it.

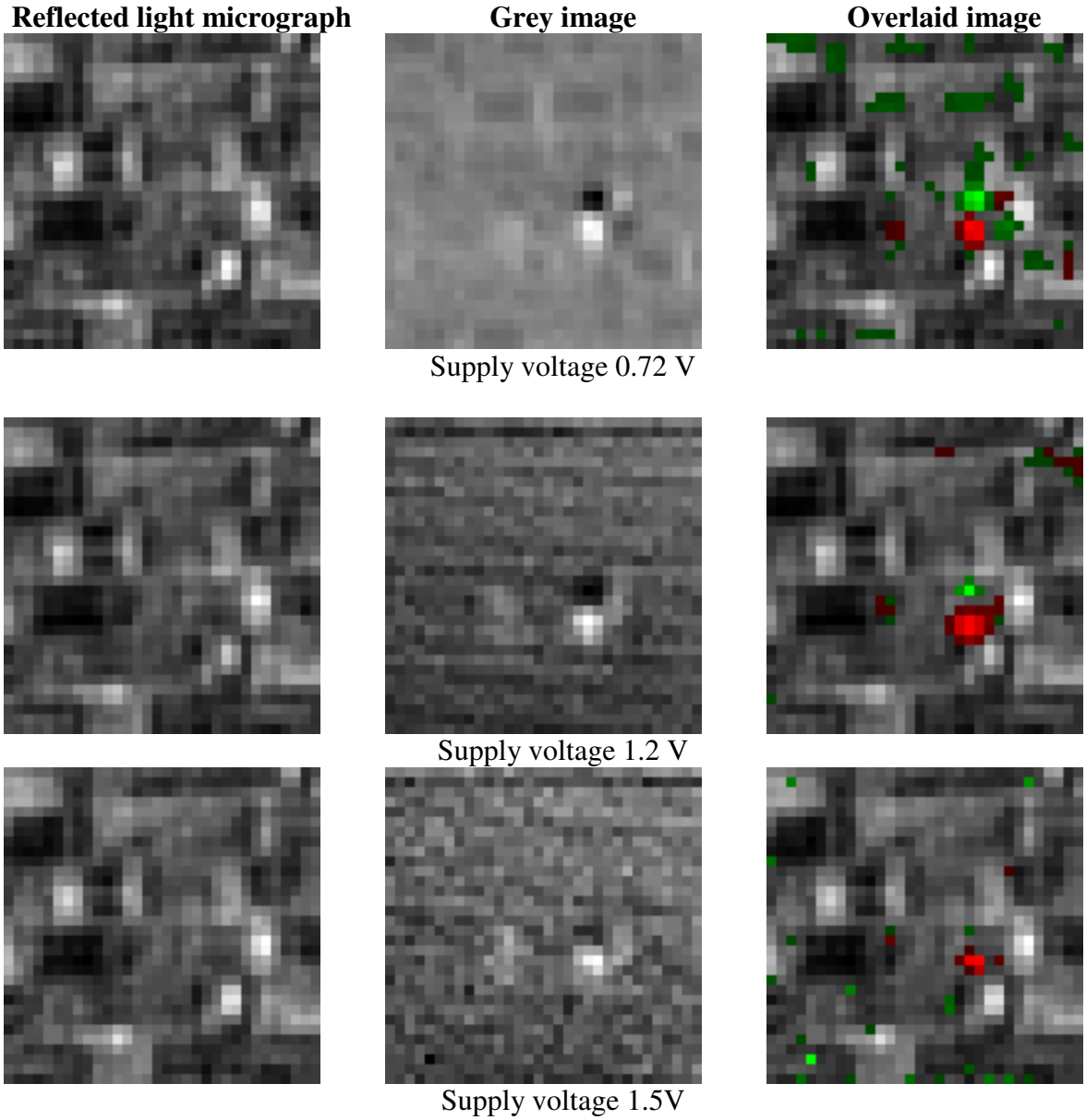


Figure 8.22. Delay Variations induced by pulsed laser illumination obtained by TAC measurement unit for different supply voltages

8.3.3. Lock-in Measurements

The results obtained from Lock-in measurements are seen in Figure 8.23. Red pixels indicates speed up mechanism whereas green slow down.

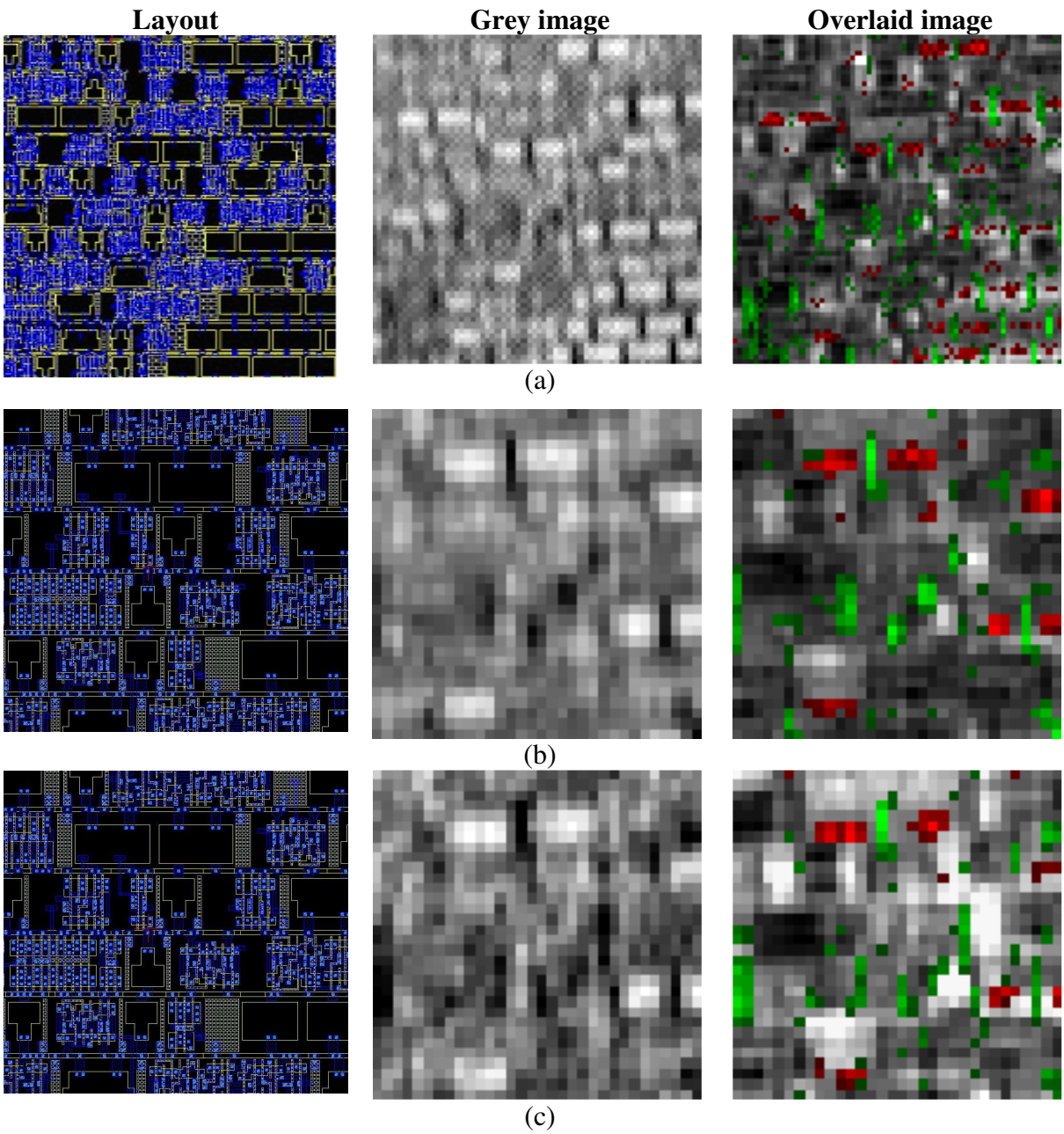


Figure 8.23. The results obtained by lock-in amplifier set-up for (a) 64x64 pixel size area and 100 MHz input frequency, (b) 32x32 pixel size area and 100 MHz input frequency (c) 32x32 pixel size area and 41 MHz input frequency

The experimental parameters are:

- Input signal frequency is 100 MHz (Figure 8.23a and 8.23b) and 41 MHz (Figure 8.23c)
- Scanning time is 10s/frame for 32x32 pixel size area and 20 s/frame for 64x64 pixel size area
- Integration number is 13

- Overall acquisition time 2.16 minutes for 32x32 pixel size area and 4.33 minutes for 64x64 pixel size area
- 0.8 V supply voltage

It is hard to interpret the results acquired with the lock-in amplifier. Because this method is sensitive not only to the delay variations at the rising edges, but also to the falling edges. Nevertheless, it is possible to localize Buffer1 and Buffer2. But the speed-up spots on the transistors are not visible except p-implanted structures.

One of the most important parameter is the input frequency. Raising the frequency boosts the signal quality. Because, keeping the scan speed at a constant value, processed data per pixel will be greater for a high frequency signal than for a low frequency signal. However, at very high frequencies, the reflections are more probable. It is not always practical to prepare a high frequency environment experimental set-up. Therefore, we carried out the same experiments for relatively high and low frequencies for comparison. The spots acquired with 41 MHz input frequency are weaker than the ones obtained with the 100 MHz frequency. (Figure 8.23 b and c)

8.3.4. Oscilloscope Measurements

The oscilloscope measurements are carried out on Buffer1. The laser beam is positioned either on nMOS or pMOS transistors and the delay times are measured by oscilloscope. Later, the laser off delay time is subtracted from laser on delay values to obtain the shift introduced by laser. The measurements are repeated for different supply voltages. Time delays are measured for each bias point more than once. Later, these values are averaged to get the mean value. Both CW and pulsed laser is used. A laser pulse width of 200 ns is applied in pulse mode.

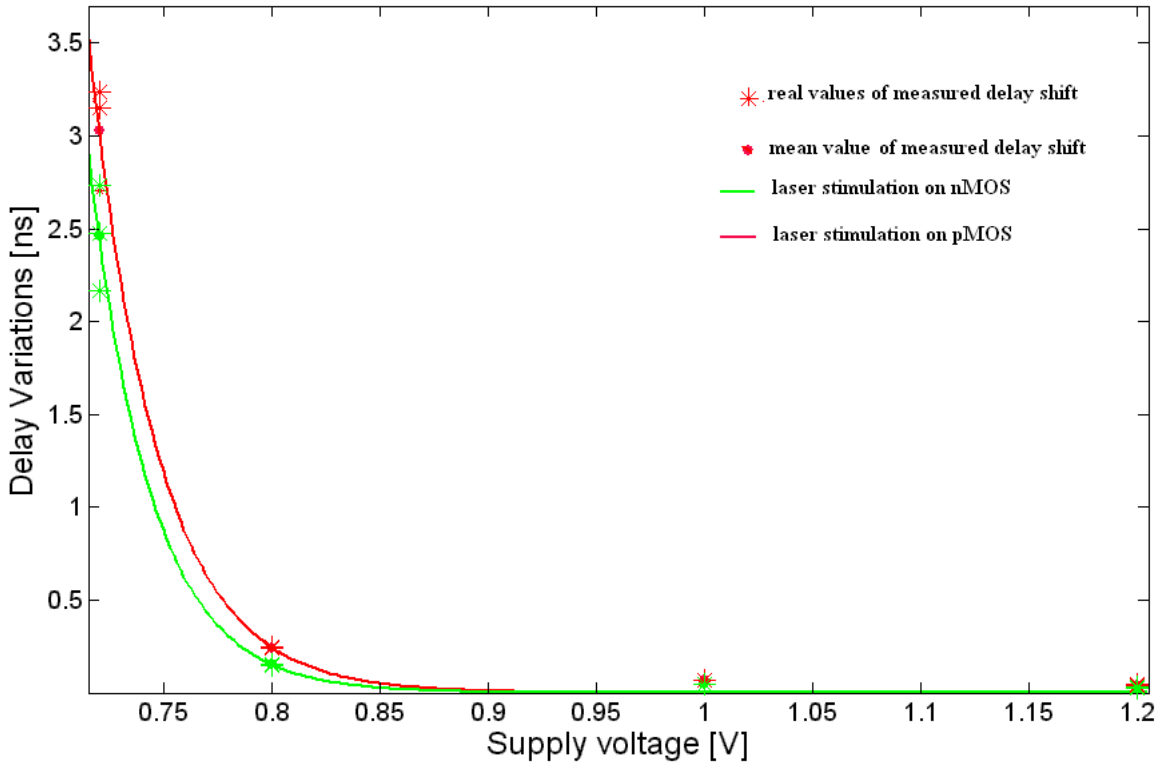


Figure 8.24. Delay variations on Buffer1 versus supply voltage for CW 1064 nm laser obtained by oscilloscope

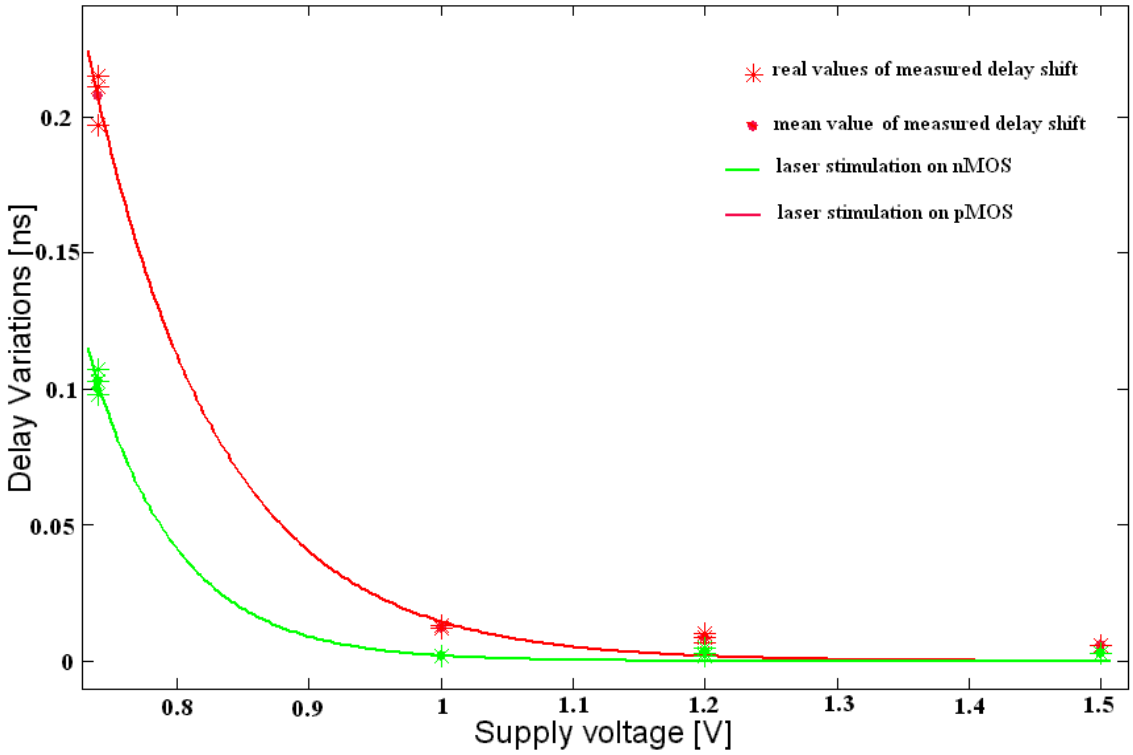


Figure 8.25. Delay variations on Buffer1 versus supply voltage for pulsed 1064 nm laser obtained by oscilloscope

The measured delay shift values and a curve, which fits to the points, are plotted in a graph. These graphs can be seen in Figure 8.24 for CW laser and in Figure 8.25 for pulsed laser.

It was observed that the delay shift caused by the CW laser is greater than the pulsed laser delay shift, which is understandable. Because, the total energy per pixel introduced by the laser diode is greater in continuous mode than in pulse mode.

The laser beam induced signals are stronger at lower supply voltages. Because, while the injected photocurrent introduced by the laser stays the same (because it is proportional to the laser power and the laser power is kept constant), the current level of the MOS transistors are reduced at lower supply voltages.

Table 8.2 summarizes the delay shift variations caused by well charging and the single transistor stimulation for 0.8 V and 1.2 V supply voltages. We biased the device at 1.2 V and 1064 nm CW laser beam is positioned on a region where no stimulation spots were observed by the TAC laser scanning experiments (in other words, away from delay chain). In this case, the induced delay variation in respect to laser off condition is around 23 ps, which can be explain by the well charging phenomenon. The continuous illumination creates a lot of electron-hole pairs in the n-well – p-epi junction. Those carriers created in bulk will be dissociated by the well space charge region in a wide area, so the well gets charged up over the whole scanning time. This well charge delays the switching of the delay chain. But, the local measurements on the transistors may be still detectable if the right voltage, temperature etc. parameters are selected. The delay time is measured while the laser is hitting the nMOS and pMOS transistors in the Buffer1 and well charging delay time is subtracted from over all delay time. It leads to a time delay difference of around -2 ps and 5.4 ps for nMOS and pMOS, respectively.

In Figure 8.24 and 8.25, the measured delay variations comprise the delay caused by n-well charging and the single transistor effects. Therefore, the delay variations are positive, pointing a slow down mechanism.

Table 8.2. The delay variations caused by well charging, laser stimulation on nMOS and pMOS transistors for CW laser

VDD	n-well charging	nMOS	pMOS	filling structures
0.8 V	133 ps	-36 ps	31 ps	-24 ps
1.2 V	23 ps	-2 ps	5,4 ps	-2,7 ps

8.4. Comparison of the Techniques

Experiments on the delay chains show that the intensity of the delay variation signals occur in a range of strength. The same laser power may cause a strong signal on a spot or a signal buried in noise on another spot. Weak signals require much integration to identify. To be able to distinguish a spot from the background, the signal has to be strong enough to rise significantly above the background noise level.

The mostly used parameter for the identification of an image quality is Signal-to-Noise Ratio (SNR).

The formula that is used to calculate SNR can be expressed as:

$$SNR = \frac{S - B}{\sigma_B} \tag{8.7}$$

S is the maximum intensity in the spot, *B* is the mean of the background intensity and σ_B is the standard deviation of the background pixels. The standard deviation can be calculated as follows:

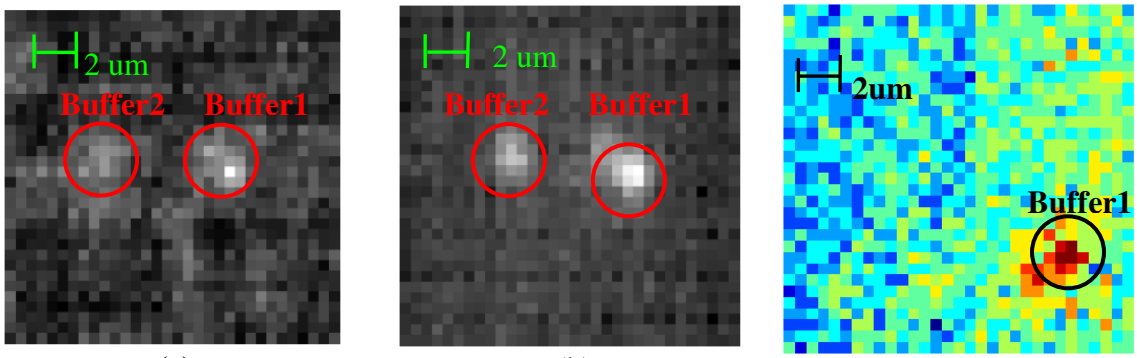
$$\sigma_B = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{8.8}$$

where $\bar{x}$ is the arithmetic mean of the x_i, intensity level values. The intensity of each pixel is obtained with the aid of PHEMOS 1000 software. The intensities are just qualitative values, which are calculated by the image acquisition card of the system, and it demonstrates the strength of the related pixel.

The results of 1.3 μm laser stimulation on Buffer1 obtained by Tester, TAC and Lock-in amplifier assisted TAC configurations (shown in Figure 8.26) are compared in the means of SNR and given in Table 8.4. The experiment parameters for these results are given in Table 8.3. In terms of SNR values, the TAC and Lock-in Amplifier assisted TAC configuration gives better results. The spots obtained with TAC are stronger compared to the tester set-up, despite of the shorter acquisition times. Tester approach is not able to localize Buffer2.

Table 8.3. The experiment parameters for three different experimental set-up configurations

	TAC	TAC + Lock-in Amp.	Tester
Scan speed [*s/ pixel*]	1.4	1.4	307.2
Integration number	100	100	5
Overall acquisition time [*min.*]	2.33	2.33	25



(a) (b) (c)

Figure 8.26. The 32x32 pixel gray images obtained by using (a) TAC (b) TAC and Lock-in Amplifier (c) Tester (VDD is 1.2 V and the laser power is 20.6 mW)

Table 8.4. SNR values for two buffers

	TAC	Lock-in amp. assisted TAC	Tester
Buffer 1	9.82	12.30	6.16
Buffer 2	5.02	8.48	-

Some of the results obtained by oscilloscope are summarized in Table 8.5. The laser beam is positioned on the Buffer1, where a slow down mechanism is observed. The delays are measured five times and the average is taken.

Table 8.5. The delay shift induced by laser stimulation on the buffer studied for different supply voltages and different laser types

	1.0 V	1.2 V	1.5 V
1.3 µm	22.66 ps	18.33 ps	9 ps
1.064 µm (CW)	64.66 ps	39.87 ps	-
1.064 µm (Pulsed)	12.33 ps	8.66 ps	6 ps

In all three cases, the induced delay shift increases with decreasing supply voltage. Therefore, the intensity of the spots drops off with rising supply voltage, which requires slower scanning time or more frames to integrate. CW 1.064 µm laser has a bigger effect than 1.064 µm pulsed laser. Because, the laser beam energy per amount of time introduced by the CW laser is higher than the pulsed laser. The delay shift induced by TLS is very hard to detect and it doesn't vary much with the supply voltage. Because, thermal effects scale with the device current. However, 1.064 µm laser induced delay shift varies a lot with changing supply

voltage. Because, while the transistor current levels get lower for small supply voltages, the injected photocurrent introduced by the laser beam does not show a big change for different bias voltages. Injected photo-current is a strong function of the laser power.

1.064µm laser beam stimulates also some filling (or dummy) structures as well as the buffers and the inverters. Those structures are p-implanted active areas and they are situated in an n-well forming a p-n junction. Laser illumination on these p-n junctions at the filling structures speeds up the delay chain if the n-well charging is not taken into account. The pulsed laser stimulation on the filling structures is minimized due to short laser pulses.

CW laser operation is not very appropriate to stimulate a precisely localized delay variation in a circuit region with large coherent well areas. The continuous illumination creates a lot of electron-hole pairs in bulk Si where the beam is off focal depth and a spread of the carrier cloud is in the range of the diffusion length (altogether easily in 100µm range), not really part of the focused laser beam but emerging in parallel. The focused beam area creates a very dense profile of carriers within the wells that are extracted by the space charge regions of the active devices only within laser spot size. So, the photocurrent in the stimulated device is occurring only when the laser is hitting their active area. But those carriers created in bulk will be extracted from the well space charge region in a wide area, so the well gets charged up over the whole scanning time. This well charge may cause a much larger delay effect to the delay chain than the single device stimulation, but we observed that local measurements on the transistors are still effective unlike stated in [RE03]. They had difficulty in separating out the laser effects for nMOS and pMOS devices due to the n-well charging.

In pulsed operation, the laser can be switched on just when the effective stimulation is required, minimizing the additional widely spread signal to a negligible quantity and assuring resolution to the laser spot size. Pulsed laser is able to suppress the secondary effect caused by the well charging and caused by filling structures.

Photoelectric and thermal laser stimulation techniques can be applied to the devices to find the speed limiting logics. Any noise (e.g. the stimulation at filling structures) may be misleading. Therefore, pulsed laser is more appropriate than CW in order to isolate the performance limiting circuits and it is a great opportunity for failure analysis laboratories.

8.5. Conclusion

Laser stimulation investigations are carried out on a delay chain. Different laser wavelengths are used and various set-up configurations are built. Mainly, 50x objective is used for the experiments and the feature size of device under test is 90 nm.

It is shown that thermal and photoelectric laser stimulation can be utilized to localize the critical paths, to isolate the design anomalies or to localize the soft defects in advanced flip-chip packaged complex ICs. Near-infrared laser stimulation effects on modern day CMOS FET devices are analyzed by achieving picosecond scale timing adjustments.

The laser stimulation investigation on delay chains is a scalable timing analysis technique for defect free ICs. But it is not systematic enough for quantitative analysis. Therefore, we introduce a new laser-based methodology to investigate the timing sensitivity analysis on scan chains by means of soft fault injection into internal nodes of scan flip-flops.

9. Fault Injection into Internal Nodes of Scan Cells

The technique uses a laser incident from the backside to inject soft faults into internal nodes of the master-slave scan flip-flop in consequence of localized photocurrent. The laser-induced photocurrent within the space charge region of reversed biased junctions is drained by the terminals of the illuminated MOS transistor adding extra drain current and therefore perturbating the normal electrical conditions of the MOS transistors. Depending on the illuminated type of the transistors (n- or p-type), the generated extra current increases or decreases the voltages at some critical nodes in the circuitry temporarily, resulting injection of a logic '0' or '1' into the master or the slave stage of the scan flip-flop. The laser pulse is externally triggered and can easily be shifted to various time slots in reference to clock and scan pattern. This feature of the laser diode allows triggering the laser pulse on the rising or the falling edge of the clock. Therefore, it is possible to choose the stage of the flip-flop in which the fault injection should occur. As it will be discussed later in this chapter in more detail, triggering the laser pulse at the rising edge flips a bit in the master stage whereas triggering at the falling edge causes a bit flip in the slave stage.

9.1. Experimental Set-up

This investigation uses the 1064 nm laser diode. Illumination by 1064 nm laser beam induces photocurrent which is the root cause of a bit flip in the scan chain output. The laser diode is dynamically triggered by the tester, very convenient to adjust to the internal signal sequence. This feature gives us freedom to shift the laser pulse in time along the vector pattern, which in turn allows detailed timing investigations. Due to many metal layers in the front side, front side illumination is avoided and DUT is exposed from the backside.

The experimental set-up is shown in Figure 9.1. Principally, we use PHEMOS 1000 as a laser stimulation system for our experiments. It has laser scanning synchronization system synchronizing the laser pulse trigger with the input stimulus of the DUT. 200 mW power laser is used for the perturbation of the devices. The laser can operate in CW and pulsed mode. The features of the laser diode specified by Hamamatsu are 50 ns minimum pulse width and 50 KHz maximum repetition rate. But, we also used 10 ns laser pulse length for our investigations. The pulse width and the frequency can be tuned by an external pulse generator, tester in our case, and shifted to the appropriate time slots through test pattern. The shifting procedure is accompolished by setting the delay of the laser pulse trigger in the timing definition of the tester.

The laser pulse is triggered by tester and it is synchronized with the scan-in and scan clock signal. The output of the scan chain is fed into DALS kit if the scanning mode is used. The DUT is driven by Agilent 83000 tester. DALS employs an analog integrator to integrate the output signal of the device per pixel. At each pixel the signal from DALS kit is evaluated and converted into gray scale data by image acquisition card of PHEMOS 1000 to obtain the gray image. The output of the analog integrator is not accesable. If there is a missing pulse or frequency, duty cycle and amplitude decrease within a pixel, the ouput of the analog integrator will decrease resulting a dark pixel in the grey image. Therefore, DALS kit is able to detect any electirical change in the signal.

The reflected laser light from the device is also collected by a detector in order to create a micrograph and these two images are overlaid for the localization. Moreover, the laser beam is positioned permanently on a single point and the output signal is analyzed at a LeCroy 6 GHz bandwidth oscilloscope for more detailed investigations.

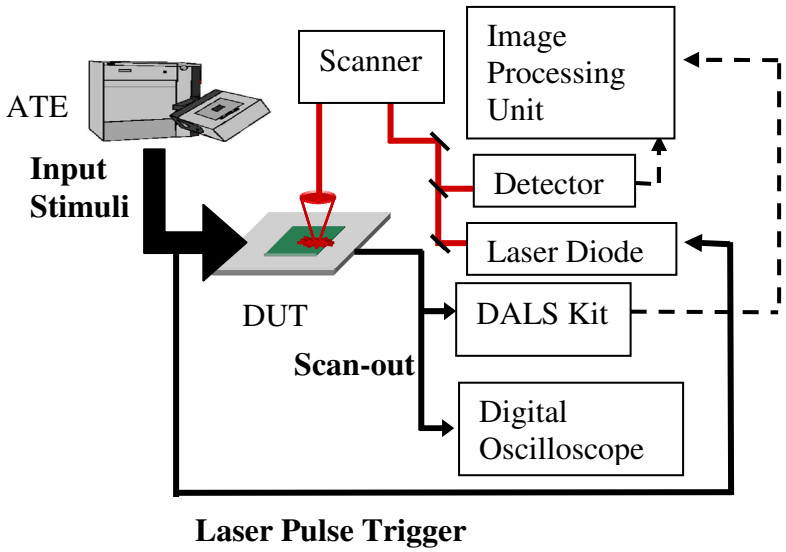


Figure 9.1. Experimental set-up

One should find a compromise between scan speed, scan area size and the strength of the spot. If the scan speed is too fast and/or area size is too large, the fast image acquisition will be achieved, but the signal may vanish in the background noise. For the experiments, a 128x128 pixel area is scanned with a scanning time of 33s so that strong spots are obtained with a fast acquisition time. 50x lens is used for all the measurements. Because, 50x lens has the highest numerical aperture among all the other lenses. Therefore, the calculated spot size is the minimum for 50x lens. One can get higher magnification with 100x, but the maximum laser power and the image quality will be poor.

9.2. Experimental Results

Pulsed and CW modes are used for the experiments and the DUT is illuminated from the backside.

The shmoo plot of the device is seen in Figure 9.2. It graphically displays the operating condition of the device with respect to period and supply voltage. Since the device functionality is weaker close to pass/fail border, the supply voltage is lowered to enhance the laser effect. An illustration of the principal scan-out response to the laser can be seen in Figure 9.3. Under illumination of CW laser, the scan chain behaves as if there is a stuck-at fault in an internal node of the exposed scan flip-flop since the output does not switch under illumination. In case of pulsed laser, the output is modulated for that clock cycle, which the laser pulse is triggered. The effects of both lasers are discussed in more detail below.

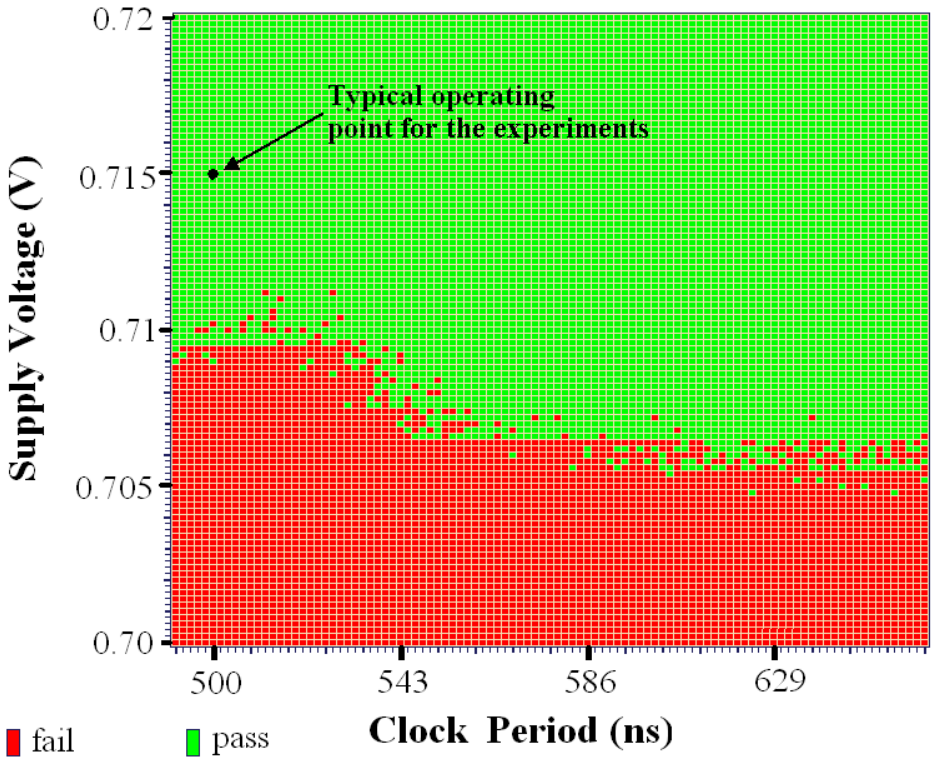


Figure 9.2. The shmoo plot of the DUT

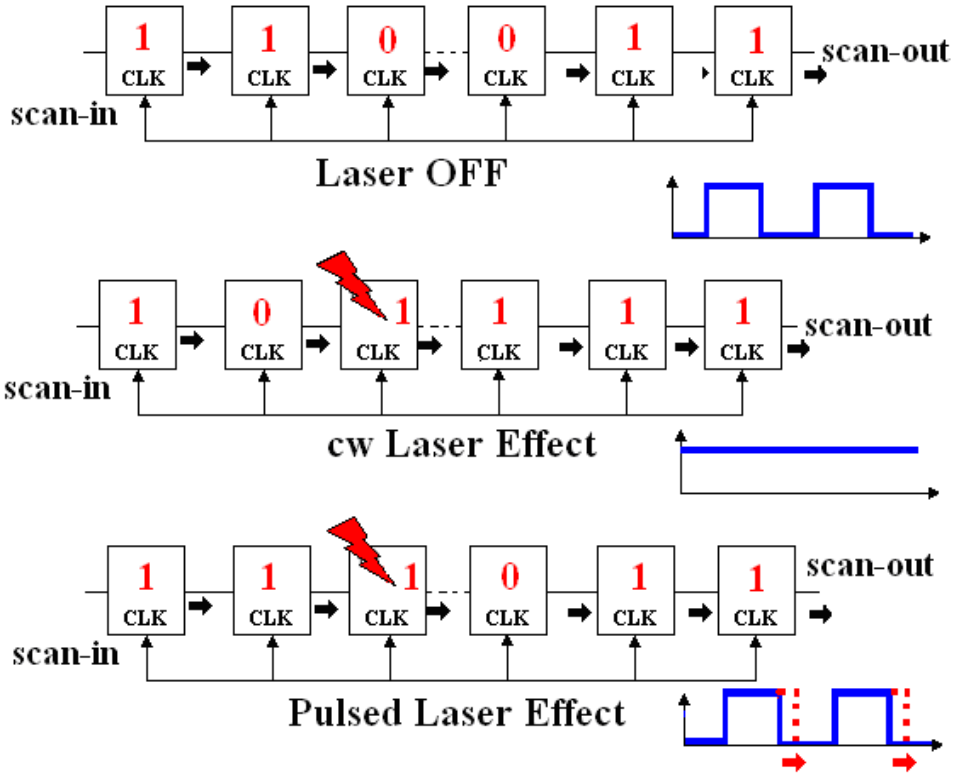


Figure 9.3. The effect of laser stimulation on scan output stimulus when the input stimuli is {110011…}

9.2.1. Continuous-wave Laser Stimulation

For these measurements, the CW mode of the laser is utilized. A scan flush pattern, {001100…}, is applied to the DUT and it is shifted in with a clock frequency of 2 MHz. For the testing of scan flip-flops, the scan flush pattern generally is used. Since it has all the logic transitions, all of the cases can be tested.

Scan shift test is not an at-speed testing. It just tests the scan chain integrity. Moreover, scan shift testing at high frequencies cause voltage drop at output signal. Because, all the scan flip-flops switch at the same time. Therefore, we picked 2 MHz clock frequency for our investigations.

The gray and superimposed images are seen in Figure 9.4. The superimposed image is derived by overlaying gray image on the reflected laser light micrograph. The superimposed image is overlaid on the device layout and we observed that each group of 3 spots indicates one scan flip-flop. One of the spots is strong and the other two are weak in intensity.

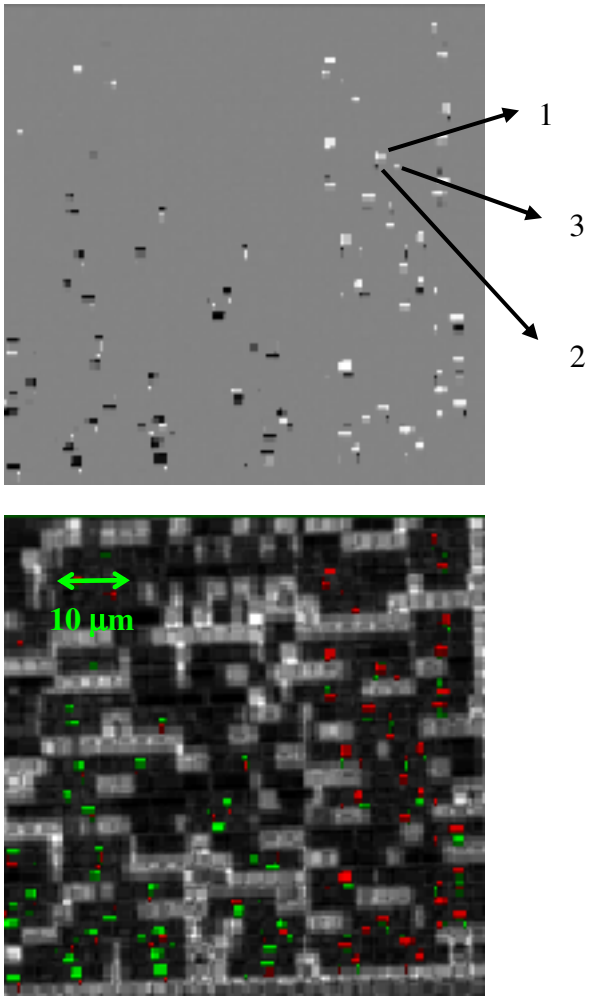


Figure 9.4. The gray and the superimposed images for CW laser stimulation. (Clock frequency is 2 MHz, the scanning speed is 33 s/frame, integration number is 10 and scanned area is 128x128 pixel size, full laser power)

The colors of the spots in the right side of the image are inverted in respect to the left side. For example, the strong white spot on the right side is black on the left side. Because, there is an inverter between the flip-flops on the right side and the ones on the left side of the

image.

The laser beam is positioned on white and black spots in single spot mode and it was observed that when white (red) spots are exposed, the scan-out stays at logic high while the black (green) spots cause the scan out to be logic '0'. The output does not switch as long as there is a laser exposition. It means that positioning the laser on a white spot causes a stuck-at-one like fault, while black spots behave like a stuck-at-zero fault.

If there is a stuck at fault in a circuitry, one way to locate the problem is to do a reset operation. After reset, the first scan shift pattern is applied. In this case, the reset pattern should come out. If there is a defect, the faulty pattern will generally be shifted out indicating the location of the stuck-at fault. However, one can only localize one type of stuck-at faults, stuck-at-one or stuck-at-zero per reset pattern with this technique. For example, if the reset sets the scan flip-flops to '1', then it would not be possible to localize stuck-at-one faults. In some circuits, the reset operation is not available.

The CW laser stimulation technique presented here allows locating both types of faults without any need of reset operation. If there were a stuck-at fault in the middle of the chain, scanning the device with laser beam would not change the output when the flip-flops before the defect are illuminated. Soft fault injection would be observable only if the flip-flops after the break are exposed which points out the location of the defect.

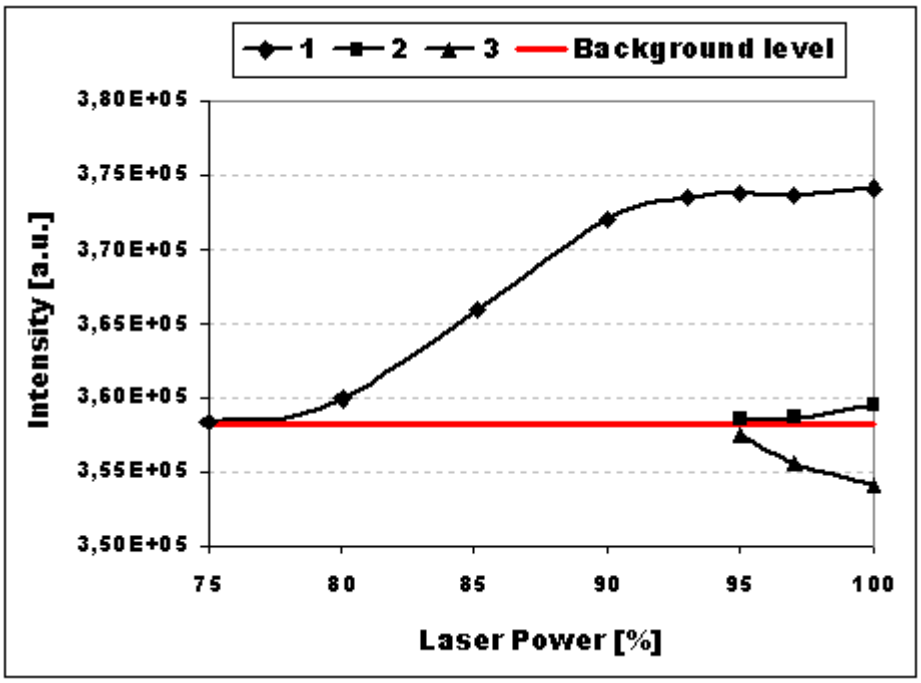


Figure 9.5. Intensity-Laser Power diagram for three different spots

The intensity-laser power diagram for the three spots, which are marked in Figure 9.4, is seen in Figure 9.5. The intensity depends on the scan speed and the numbers of the frames that are integrated. In this case, 10 frames were taken and the scanning speed was 33 s/frame for a 128x128 pixel area. The image acquisition card uses the DALS output signal for the calculation of the spot intensities. The intensity is a qualitative measure and it just demonstrates the strength of the spots. The graph shows that the strongest stimulation occurs when the laser hits the 1st site whereby the intensity for the 2nd and the 3rd spots are weak. There is a minimum power required to cause a fault injection and it is around 80% of full

laser power for the 1st spot and it saturates at 90% of laser power. The required minimum power is around 95% for 2nd and 3rd spots. At this point, we can assume that the minimum required laser energy to cause an upset depends on the logic and the layout design of the circuitry. This will be investigated later.

9.2.2. Analysis of the Sensitive Sites

The superimposed image is overlaid on the layout of the circuitry and the sensitive structures in a scan flip-flop to the soft fault injection are acquired. These sensitive spots are numbered in Figure 9.4 and correlated to the schematic section of the sensitive circuitry of the master and slave stage in Figure 9.6 and 9.7. The first and the second spots are related to the reset and clock-controlled inverter and the transfer gate in the master stage while the third one is in the clock-controlled inverter and transfer gate of the slave part.

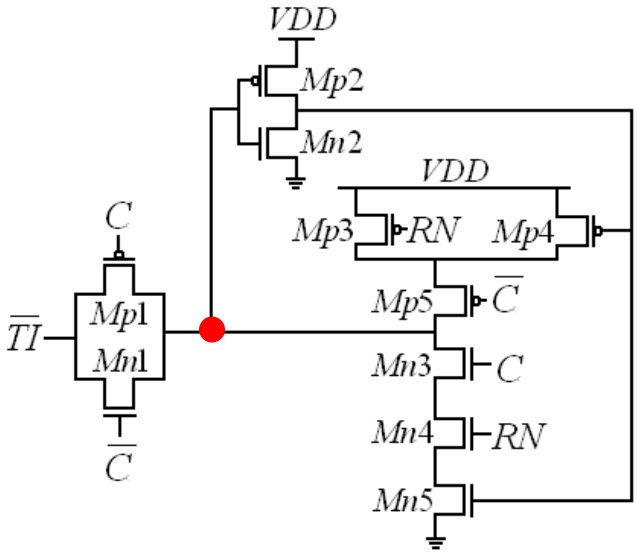


Figure 9.6. The schematic section of the sensitive structures in the master stage

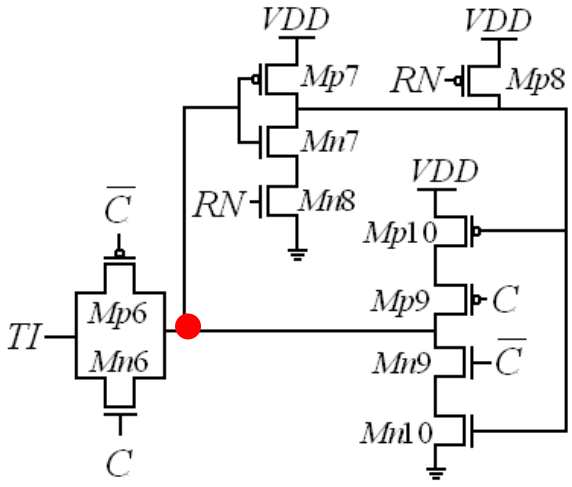


Figure 9.7. The schematic section of the sensitive structures in the slave stage

First, the master stage is discussed. The node, which is pointed, changes its state from high to low or from low to high depending on the illuminated site. If the sensitive node is logic low, there will be an upset when the pMOS-network in the clock and reset controlled

inverter is exposed and it will cause 1st spot to appear. If the sensitive node is logic high, it will toggle to logic low once the laser beam hits the nMOS-network in the clock and reset controlled inverter, which will make the 2nd spot to become visible. Mn1, Mn3, Mn4, Mn5 transistors are situated in the same active area as well as Mp1, Mp3, Mp4, Mp5.

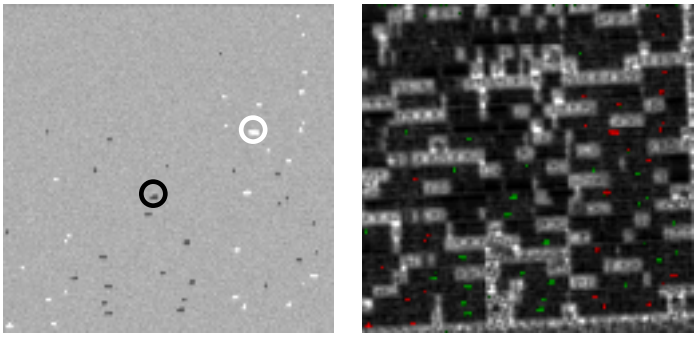
The schematic section of the sensitive structure in the slave stage is seen in Figure 9.7. When the laser hits the drain of the Mn9 and Mn6, the sensitive node flips from logic '1' to '0' which in turn causes the 3rd spot to become visible.

9.2.3. Pulsed Laser Stimulation

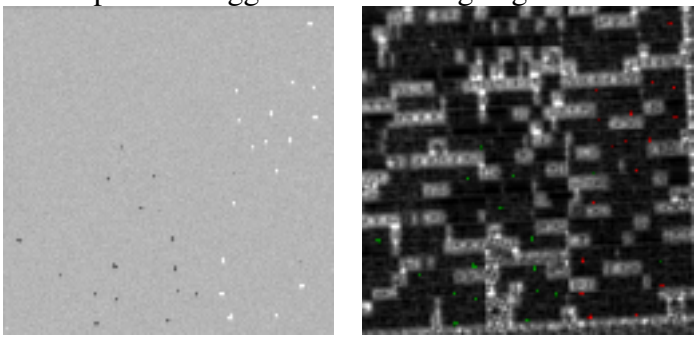
In this case, the laser pulse is triggered by tester with a pulse width of 50 ns and a signal frequency of 48 kHz. A scan shift pattern, which has a 48 kHz signal frequency, is applied to DUT and it is shifted in with 2 MHz scan clock signal.

At first, the laser pulse is triggered at the rising edge of the clock and a 128x128 pixel area is scanned pixelwise. As a result, the gray and the superimposed images are acquired which are shown in Figure 9.8. The investigations showed that pulsing the laser at the rising edge of the clock causes stimulation in the master stage. In this case, the 1st and the 2nd spots become visible. When the laser pulse is triggered at the falling edge of the clock, the n-network in the slave stage causes a bit flip from '1' to '0' at the sensitive node which causes the 3rd spot to appear.

Now, with the laser beam position at the red (white) spot (marked in Figure 9.8), the scan out signal is investigated with an oscilloscope. The scan output signals are shown in Figure 9.9. This shows how the pulsed laser beam inserts '1' into the output stream of the scan out and therefore increases the duty cycle. When the laser is positioned on the green (black) spot, an injection of a '0' occurs.

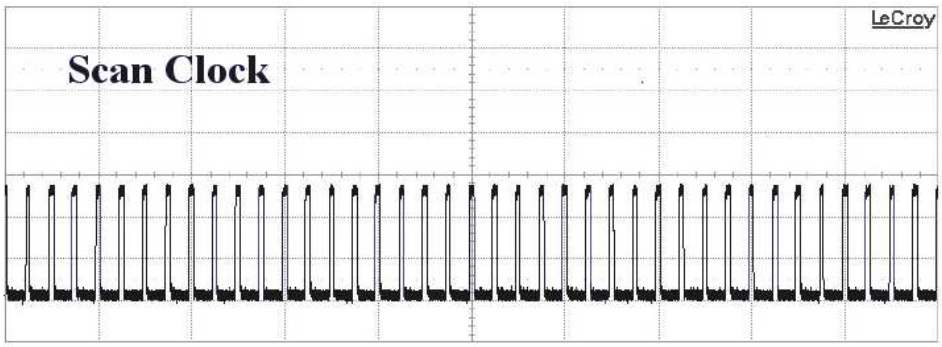


Laser pulse is triggered at the rising edge of the clock

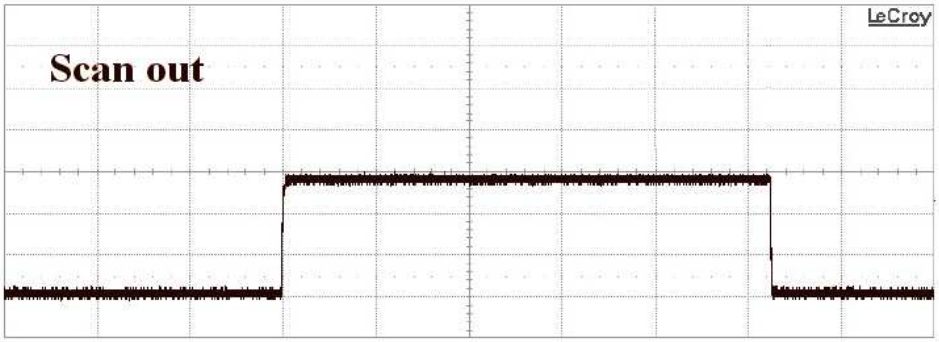


Laser pulse is triggered at the falling edge of the clock

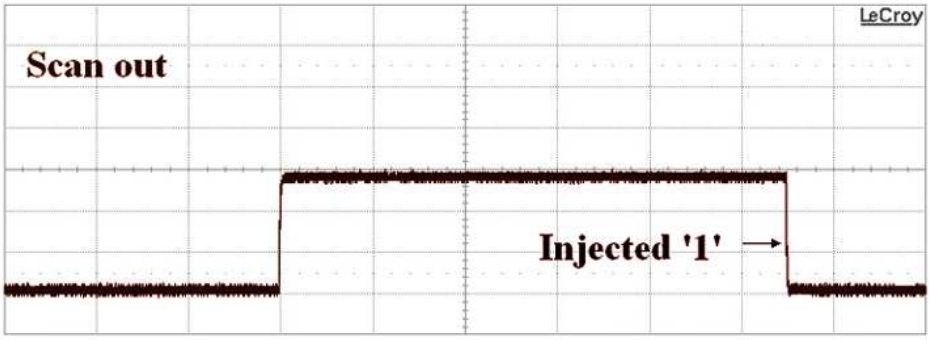
Figure 9.8. The gray and the superimposed images for pulsed laser triggered at the rising and the falling edge of the clock (Clock frequency is 2 MHz, the scanning speed is 33 s/frame, integration number is 10 and scanned area is 128x128 pixel size, full laser power, laser pulse repetition rate is 48 kHz, laser pulse width is 50 ns)



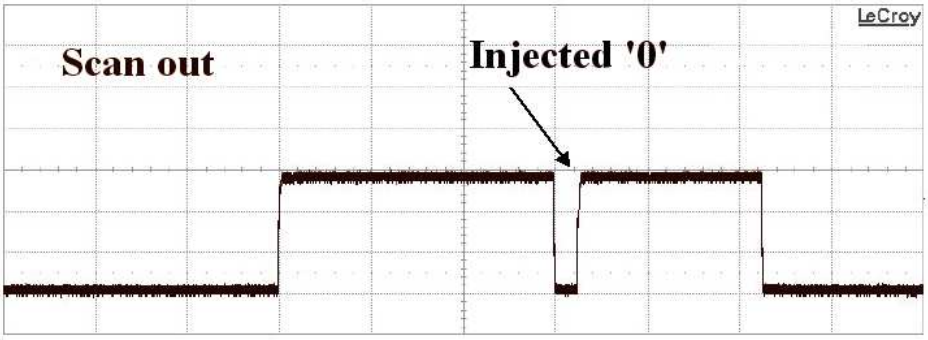
Scan Clock



Scan-out signal for Laser OFF



Scan-out signal when the laser is on the red spot



Scan-out signal when the laser is on the green spot

Figure 9.9. Signals measured at oscilloscope (Clock frequency is 2 MHz, the scanning speed is 33 s/frame, integration number is 10 and scanned area is 128x128 pixel size, full laser power, laser pulse repetition rate is 48 kHz, laser pulse width is 50 ns)

These results prove that pulsed laser stimulation is able to localize the defected scan flip-flop and the defected stage within the scan flip-flop, which is not possible with ATPG (Automatic Test Pattern Generation) tools.

For further investigations, a stream that consists of only logic ones, {1111…} is shifted into the scan chain and a laser pulse is introduced to the system with a repetition rate of 48 KHz and a pulse width of 50 ns. It is triggered at the rising edge of the clock. In the second case, we applied {0000…} to the scan input. The overlaid images are seen in Figure 9.10. The white spots are visible for the first image whereas there are only black spots for the second case. It should be emphasized that the applied input to the scan chain comes out inverted.

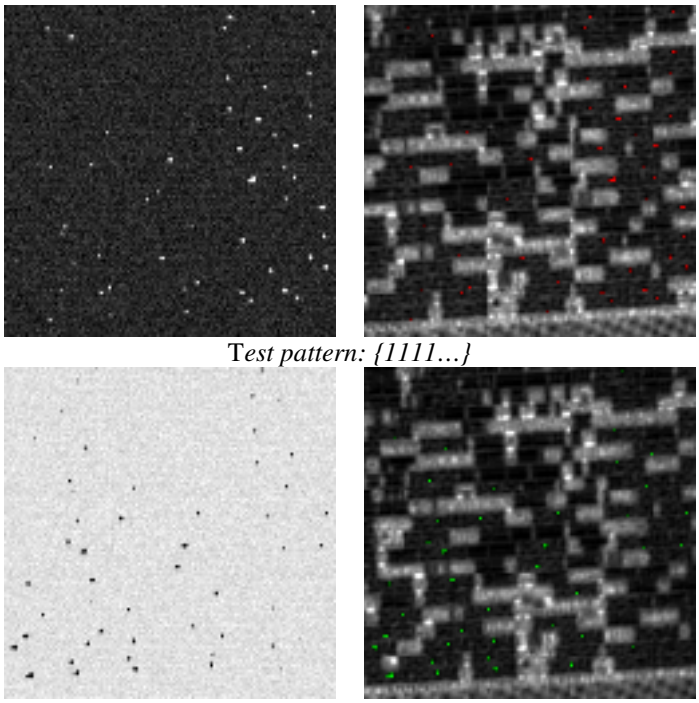


Figure 9.10. The overlaid images for pulsed laser for the test patterns {1111…} and {0000…} (Clock frequency is 2 MHz, the scanning speed is 33 s/frame, integration number is 10 and scanned area is 128x128 pixel size, full laser power, laser pulse repetition rate is 48 kHz, laser pulse width is 50 ns)

In order to probe the scan-out signal, we positioned the laser beam on the white and black spots in single spot mode and analyzed the output signal at the oscilloscope. As a result, we observed a modulated scan-out signal due to the laser stimulation. The oscilloscope images are seen in Figure 9.11. Illumination of the white spots injects '1' to the output stream whereas black spots inject '0' for one cycle at which the laser pulse is triggered.

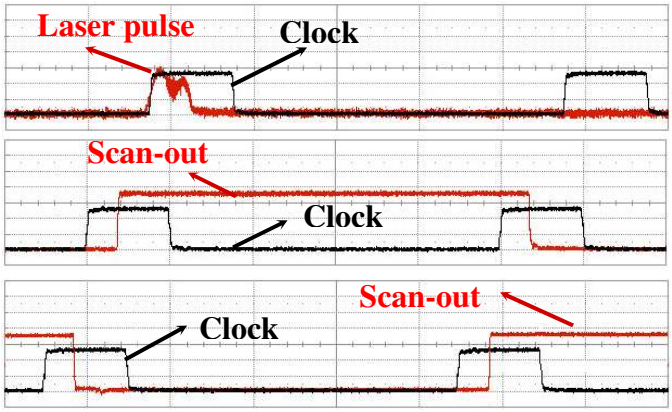


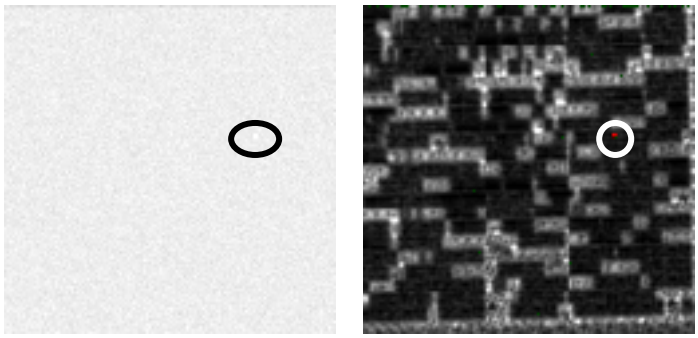
Figure 9.11. Laser pulse, scan clock and laser induced scan-out pulse signals (100 ns/div) (Clock frequency is 2 MHz, the scanning speed is 33 s/frame, integration number is 10 and scanned area is 128x128 pixel size, full laser power, laser pulse repetition rate is 48 kHz, laser pulse width is 50 ns)

9.2.4. Timing Sensitivity

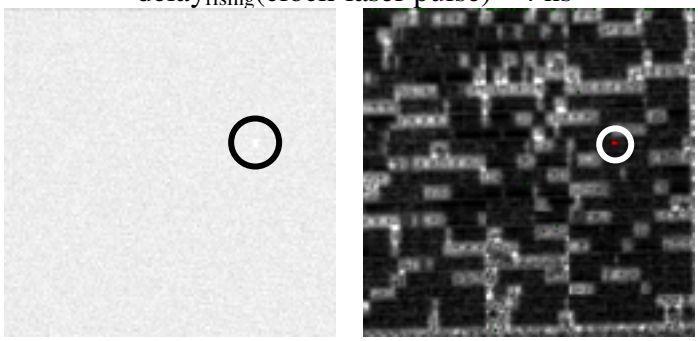
The laser pulse is shifted in time to find out the time interval with maximum stimulation. Consequently, it was observed that there is a correlation between the intensity of the laser induced signal and the clock to laser pulse delay. The gray and the superimposed images are shown in Figure 9.12 for different clock to laser pulse delays. When the delay increases, the number of the spots and their intensities decrease. Therefore, the maximum effect occurs when the time delay between the edges of the laser pulse and the scan clock is small. The clock and the laser pulse signals are shown in Figure 9.13 at which the best injection occurs for 50 ns and 10 ns laser pulses.

In Figure 9.14, it is illustrated that the stimulated activity is correlated to the edge of the clock signal pattern. The time window width, in which the upset could occur, varies depending on the laser power and the pulse width. The time window for 10 ns pulse width and 65% power laser pulse is narrower than 10 ns laser pulse with 100% power. Since the laser power is higher for 50 ns width laser pulse, the time window gets broader. The delay diagram of clock to laser pulse (Figure 9.14) demonstrates that the two investigated logical streams result in a critical timing information in 18 ns range (for 50 ns pulse width), much smaller than the stimulation pulse width itself. One reason may be that the stimulated activity is correlated to the edge of the signal pattern. So the time resolution for application in failure analysis correlates with the sharpness of the pattern edges. One another reason to the limiting factor could be also the shape of the laser pulse.

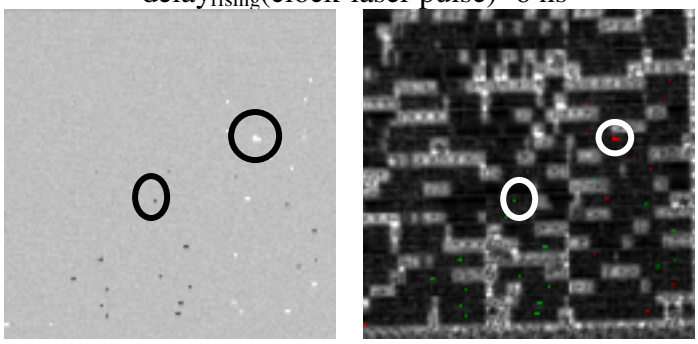
As a result, pulsed laser stimulation is able to define the most critical point in time of a switching activity at which the stimulation occurs. Tuning the laser power down to the stimulation limit and the use of shorter pulses narrow down the effective time window.



$delay_{rising}$(clock-laser pulse)=-4 ns



$delay_{rising}$(clock-laser pulse)=6 ns



$delay_{rising}$(clock-laser pulse)=0 ns

Figure 9.12. Gray and superimposed images for pulsed laser

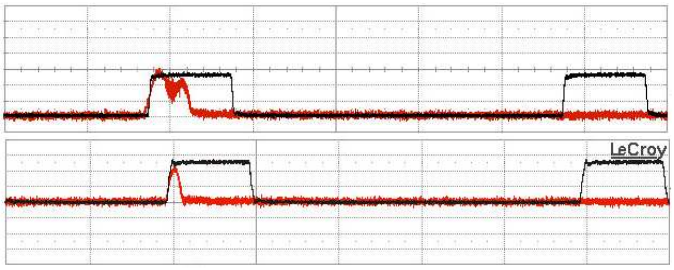


Figure 9.13. The clock and the laser pulse signals at which the strongest injection occurs for 50 ns and 10 ns laser pulse width (100 ns/div.)

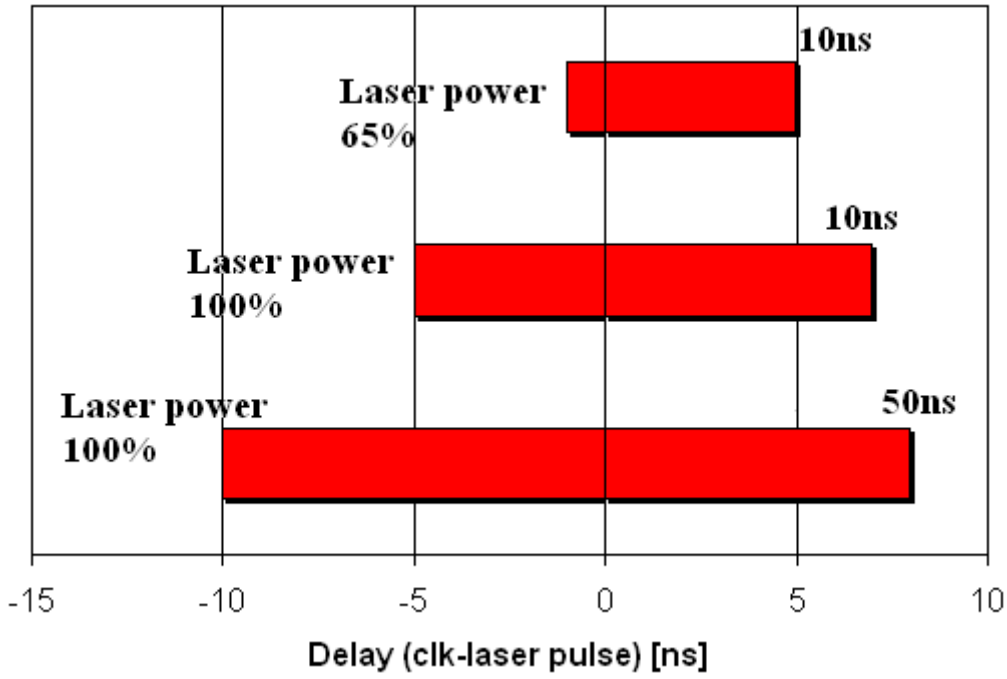


Figure 9.14. The laser pulse width-delay $_{\text{clock-laser pulse}}$ diagram for different laser powers

9.2.5. The Stimulation Dependence on Supply Voltage

In order to evaluate the laser stimulation intensity dependence on supply voltage, the DUT is scanned with a scan speed of 33 s and 20 frames are integrated for various supply voltages. The intensities of the white and black spots which are marked in Figure 9.12, are calculated and an intensity-supply voltage diagram (Figure 9.15) is obtained. The intensities of the spots get weaker as the supply voltage increases. This is due to lower operating currents at low supply voltages. But, the injected current is more or less constant. Thus, the perturbation caused by injected photocurrent is stronger for lower supply voltages. It shows that stimulating the device at 0.8 V is also possible, which is a much more stable operating condition regarding the jitter. Signal jitter was observed for the working points lower than 0.715 V.

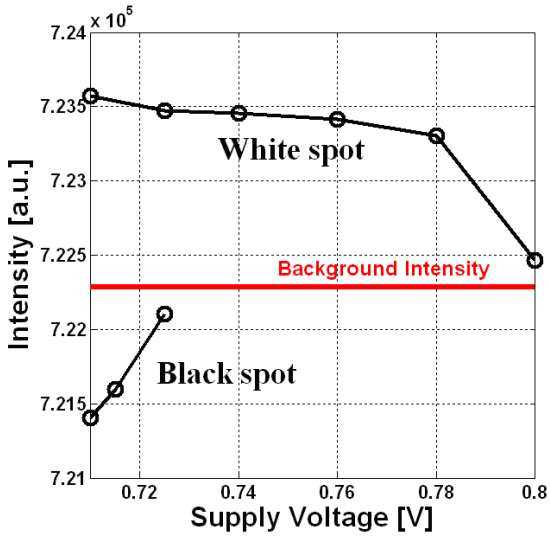


Figure 9.15. The intensity versus supply voltage diagram for white and black spots

10. Simulations

The simulations were done in order to verify the experimental results. The aim is to reproduce the photoelectric effect caused by laser illumination in software level. The simulations are done by using Titan circuit simulator and 90 nm technology transistor model parameters (supplied by Infineon AG, Munich) were applied. The output characteristics of single nMOS and pMOS transistors are simulated and shown in Figure 10.1 and Figure 10.2, respectively. They fit with the experiment results well for high V_{GS} voltages. But there is some discrepancy for low V_{GS} levels as shown in Table 10.1. It would be very time consuming to further optimize the parameter set for fitting the whole voltage range. This optimization is not expected to change the results considerably and therefore exceeds the scope of this work.

Table 10.1. The I_{DS} current values obtained by experiments and simulations for various V_{GS} voltages at nominal V_{DS} voltage

V_{GS} [V]	I_{DS} [µA] nMOS		I_{DS} [µA] pMOS	
	Experiment	**Simulation**	**Experiment**	**Simulation**
0.4	1.6	4	0.48	1
0.6	16	23	4	7
0.8	41	46	13	16
1.0	67	69	24	27
1.2	92	92	36	39

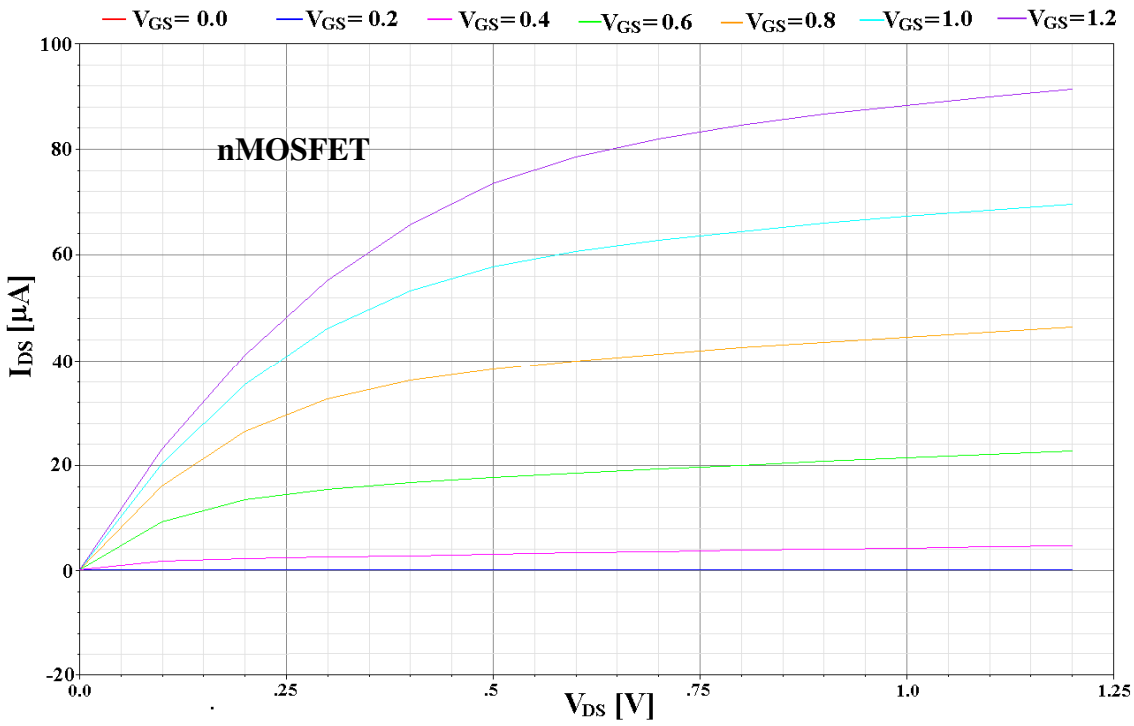


Figure 10.1. Simulated 0.12x0.08 µm size n-channel MOSFET output characteristics

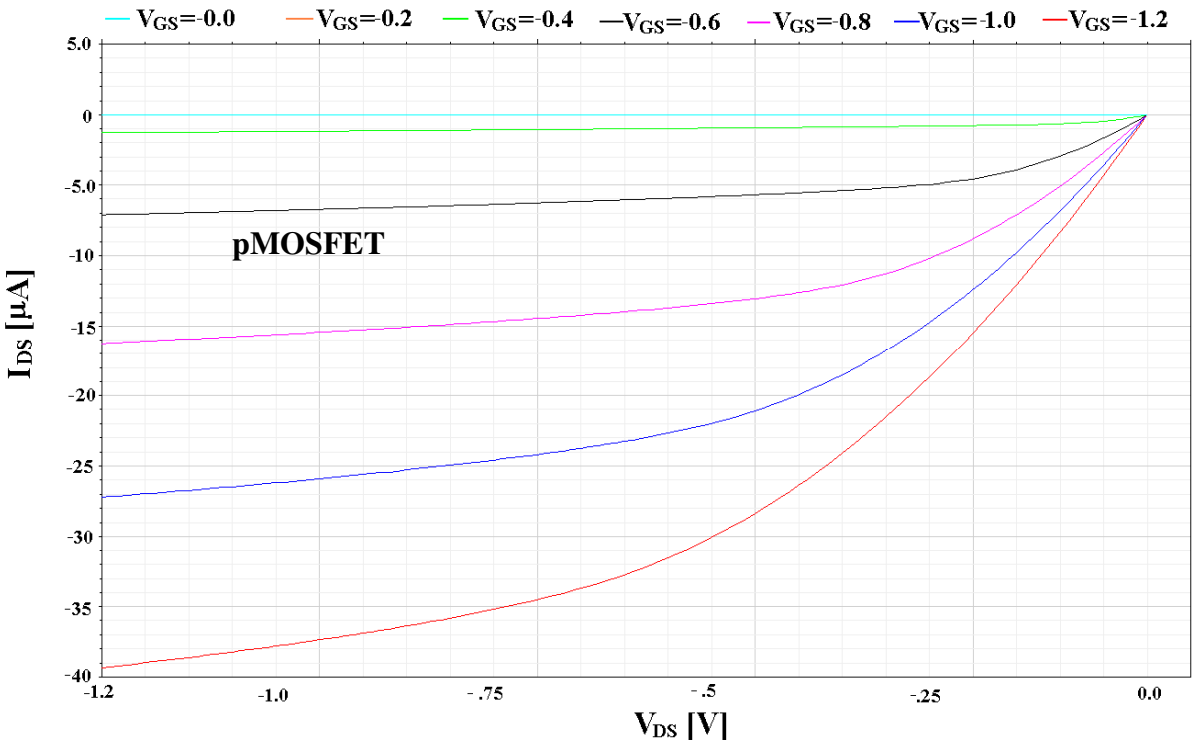


Figure 10.2. Simulated 0.12x0.08 µm size p-channel MOSFET output characteristics

10.1. PLS Effect on the Switching Speed of a Buffer

The effect of photoelectric laser stimulation on the switching speed of the Buffer1 (Figure 10.3) is simulated. The same transistor sizes are chosen to match the experimental structures. A current source bypassing the supply voltage and the drain terminal is used in order to model the photoelectric effect on a pMOS transistor whereas the current source between the drain terminal and the ground models the effect on nMOS transistor. A CMOS inverter is connected to the output of Buffer1 and the simulation results are measured at the output of this CMOS inverter.

A constant current source (20 µA) is connected at first parallel to the pMOS transistor and then to the nMOS transistor in the output inverter. When the current source value is chosen to be 20 µA, the simulation results fit well to the experimental results. A pulse switching between 0 V and 1.2 V with a period of 200 ps is applied to the input and simulated. The period is chosen to be short, in order to show the effects for the rising and falling edges in the same graph. The output signal is shown in Figure 10.4. Apparently, there is a delay or advance in the output signal depending on the illuminated type of the transistor and the edge. The output falling edge is delayed when the nMOSFET is illuminated and it is advanced when the pMOSFET is illuminated. The opposite behavior is observed at the rising edge of the output.

Besides, the duty cycle is modified by the stimulation. In the case of the nMOS stimulation, a duty cycle increase is observed whereas a reduction in the duty cycle occurs when the pMOSFET is illuminated.

The delay variations caused by nMOS and pMOS perturbations are given in Table 10.2 for different supply voltages and for a current source of 20 µA. These results are plotted for

rising and falling edge and shown in Figure 10.5 and Figure 10.6, respectively. The delay variations decrease with the increasing supply voltage as presented in experimental results previously. The simulation results match to the experimental results (Table 8.2) if the n-well charging is not taken into account. N-well charging effect is discussed in detail in Section 8.4.

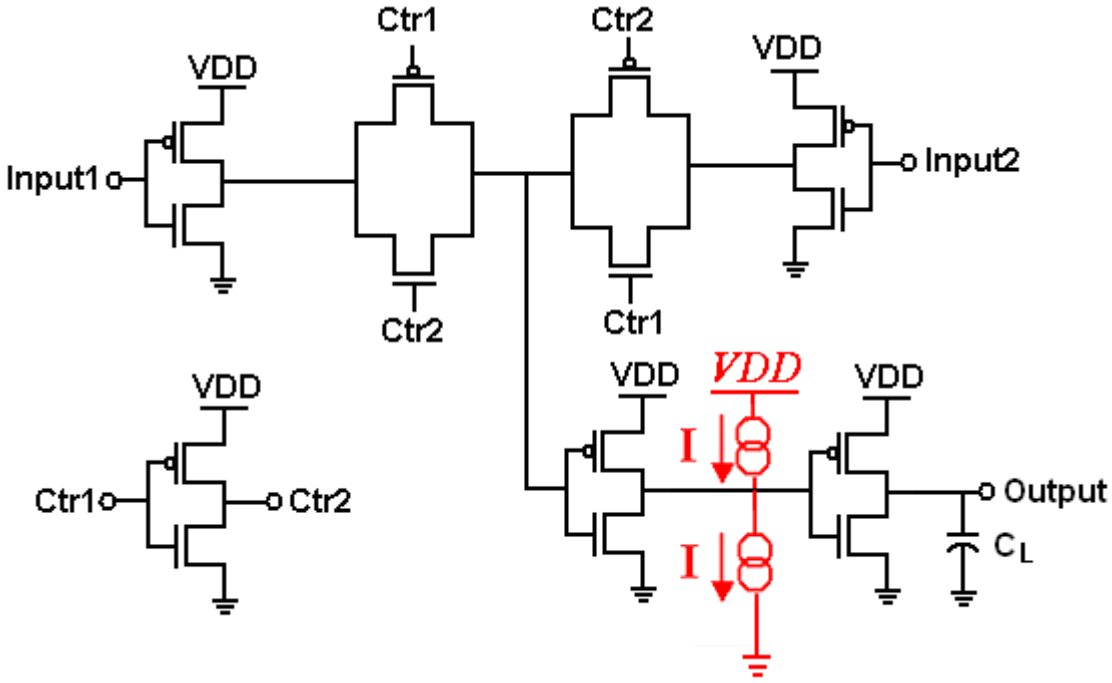


Figure 10.3. The schematic of the simulated buffer (Buffer1)

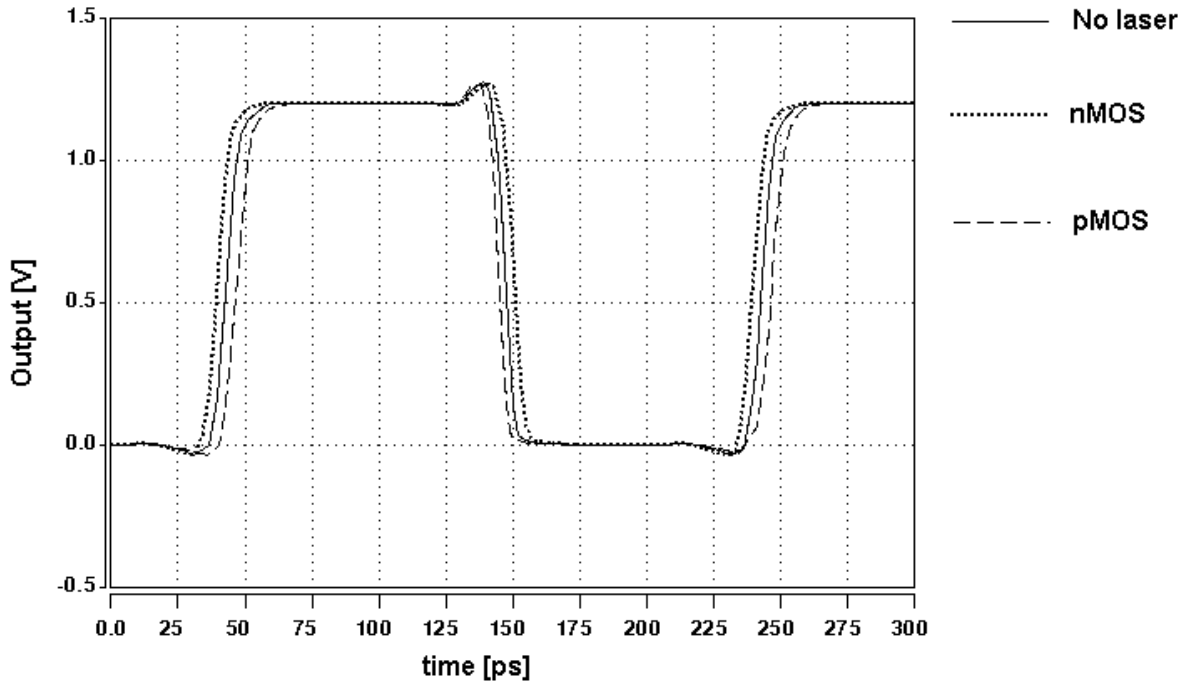


Figure 10.4. The delay shifts caused by current sources connected parallel to nMOSFET and pMOSFET of Buffer1 (a CMOS inverter is connected serial to the Buffer1 output)

Table 10.2. Induced delay shifts as a function of the edges and the type of the transistor for various supply voltages, obtained by simulations (I_{source}=20 µA)

	Delay shift at the rising edge [ps]		Delay shift at the falling edge [ps]	
Supply Voltage [V]	**nMOS**	**pMOS**	**nMOS**	**pMOS**
0.8	40,652	-35,927	-23,426	57,296
1.2	1,8	-5,38	-2,417	2,846
1.5	0,641	-0.555	-0,989	0,913

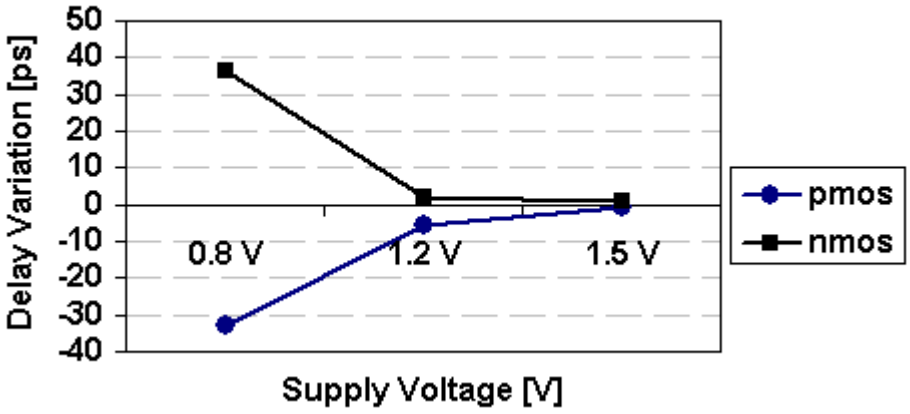


Figure 10.5. The simulated delay variation diagram for rising edge and for various supply voltages (I_{source}=20 µA)

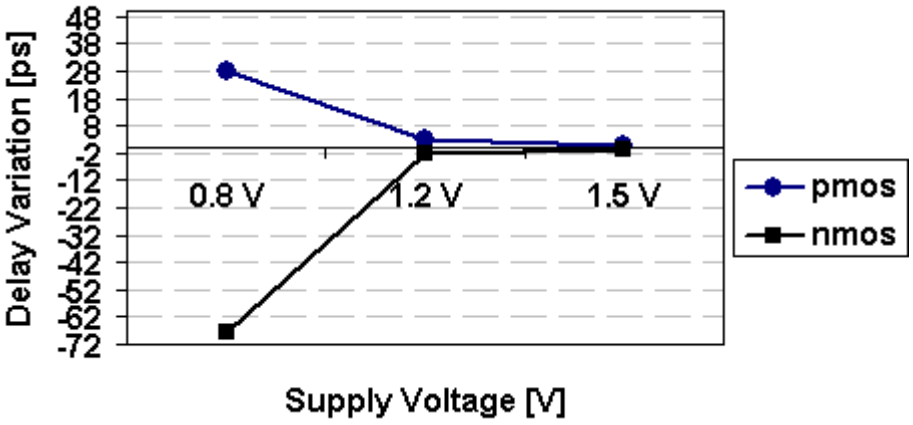


Figure 10.6. The simulated delay variation diagram for falling edge and for various supply voltages (Isource=20 µA)

According to [RE03], the laser illumination on pMOSFETs changes the threshold voltage through n-well charging. Because generated charge carriers alter the body voltage. This effect is simulated by applying body bias voltage source different from supply voltage on Buffer1. Various body bias voltages are applied to all of the pMOS transistors in the Buffer1 when the circuit operates at nominal supply voltage. The maximum delay shift caused by back gate is around 4 ps as shown in Figure 10.7. The simulated effect is smaller than the experimental results. Because, the n-well charging affects not only one buffer, but many buffers in the chain, which are sharing the same n-well. But, we simulated only one buffer.

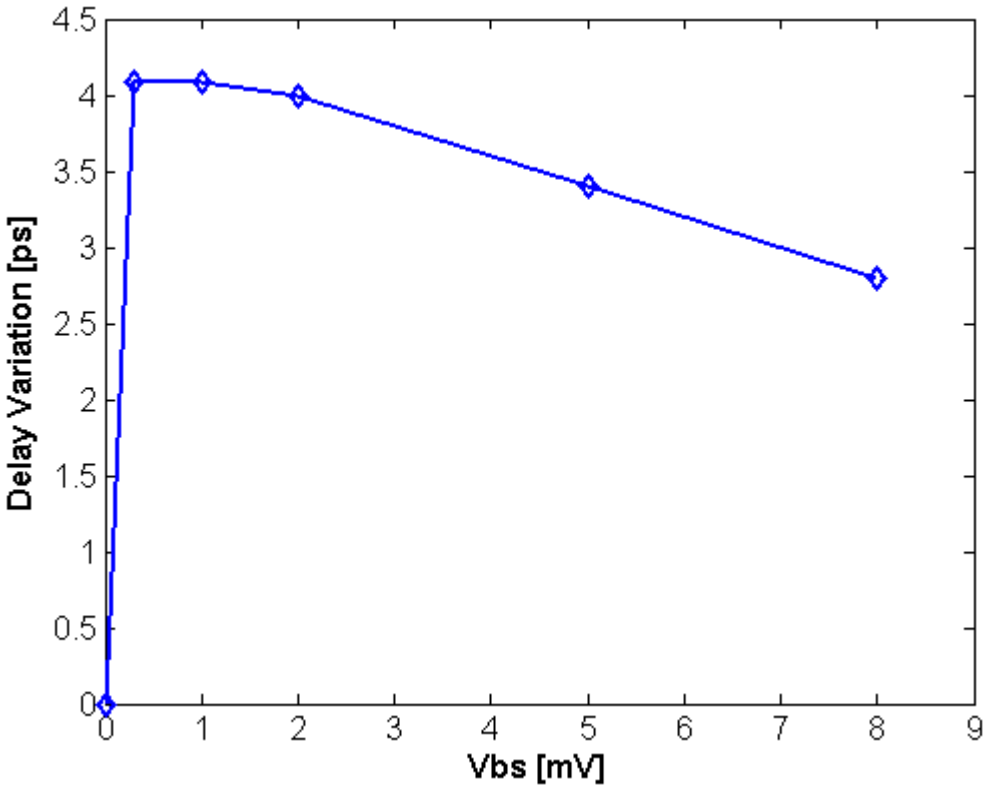


Figure 10.7. The simulated delay variation diagram as a function of V_{BS} voltage for rising edge (VDD is 1.2 V)

10.2. PLS Effect on the Scan Chains as Fault Injection

The soft fault injection phenomenon is studied in terms of timing analysis. The scan flip-flop, which is used for experiments, is simulated to investigate the PLS soft fault injection mechanism on scan chains. Illuminating the transistors changes the drain current, which perturbs the node voltages via inducing or reducing charge at the internal parasitic capacitances. Because of the feedback loop present in memory elements, the modified voltages are preserved and conveyed to the following stages.

As it is discussed in the Section 6.1.2, the photoelectric effect can be modeled by a current source bypassing drain and source terminals if the resistance of the well is neglected (Figure 6.15). For the simplicity of the circuit simulations, we use independent current sources, which is exponential in time. This model was enough to reproduce the results in software level (see Figure 6.11 and 6.13). The transport of the excess carriers under a laser pulse illumination is an exponential equation, which describes the carrier diffusion (explained in Section 5.10).

The following equations describe the current through the source as a function of time:

$$I_{out} = I_1 \qquad 0 \le t \le TDR \tag{10.1}$$

$$I_{out} = I_1 + (I_2 - I_1) \cdot \left(1 - e^{-\frac{t-TDR}{TR}} \right) \qquad TDR < t \le TDF \tag{10.2}$$

$$I_{out} = I_1 + (I_2 - I_1) \cdot \left(e^{-\frac{t-TDF}{TF}} - e^{-\frac{t-TDR}{TR}} \right) \qquad t < TDF \tag{10.3}$$

where I_1 is the initial value, I_2 is the pulse value, *TDR* is the rise delay time, *TR* is the rise time constant, *TDF* is the fall delay time and *TF* is the fall time constant.

According to the experiments, the best injection occurs when the delay between the rising or falling edges of the clock pulse and laser pulse is zero. The exponential current source pulse is shifted in negative and positive direction as shown in Figure 10.8 to obtain the best timing window for injection.

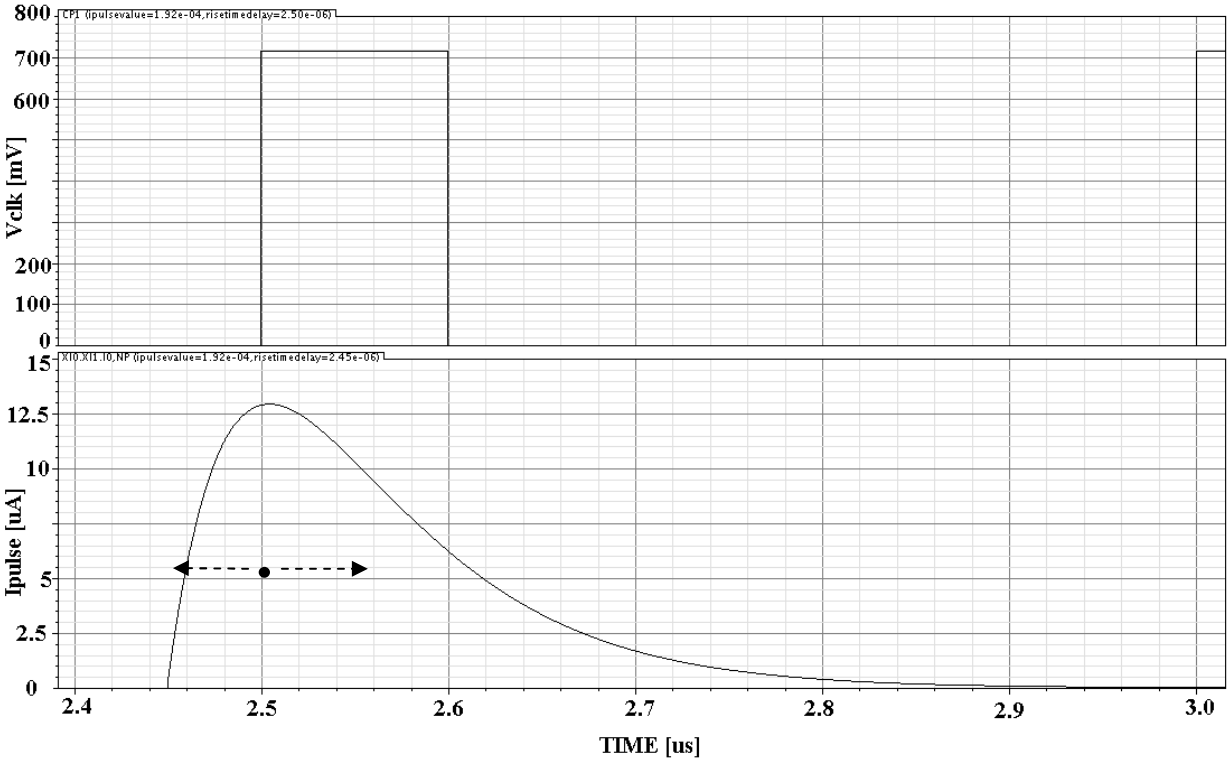


Figure 10.8. Simulated clock pulse and exponential current pulse source signals (TR=50 ns, TF=100 ns, TDR=2.45 μs and TDF= 2.51 μs)

The logic and the transistor level circuit diagram of the scan flip-flop are shown in Figure 10.9 and Figure 10.10, respectively. There are three sensitive parts in the scan cell to the fault injection, two of them are in the master stage and one is in the slave stage. The experiments prove that the strongest stimulation happens when the pMOS network in the master stage is exposed. Therefore, we will discuss first the strongest spot. A current source is connected between VDD and the drain electrode of the Mp5 transistor, which is a candidate for fault injection as shown in Figure 10.10.

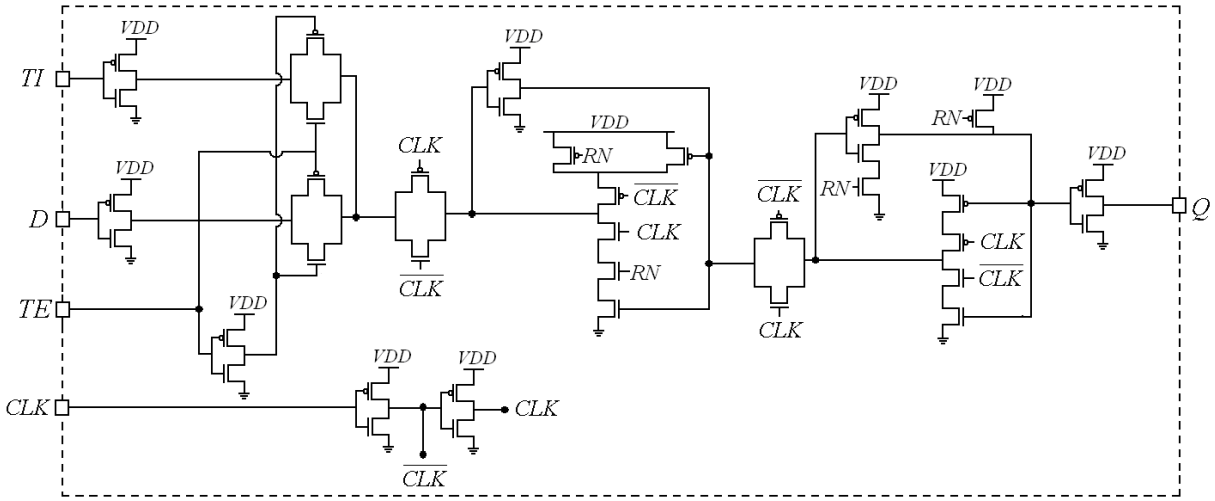
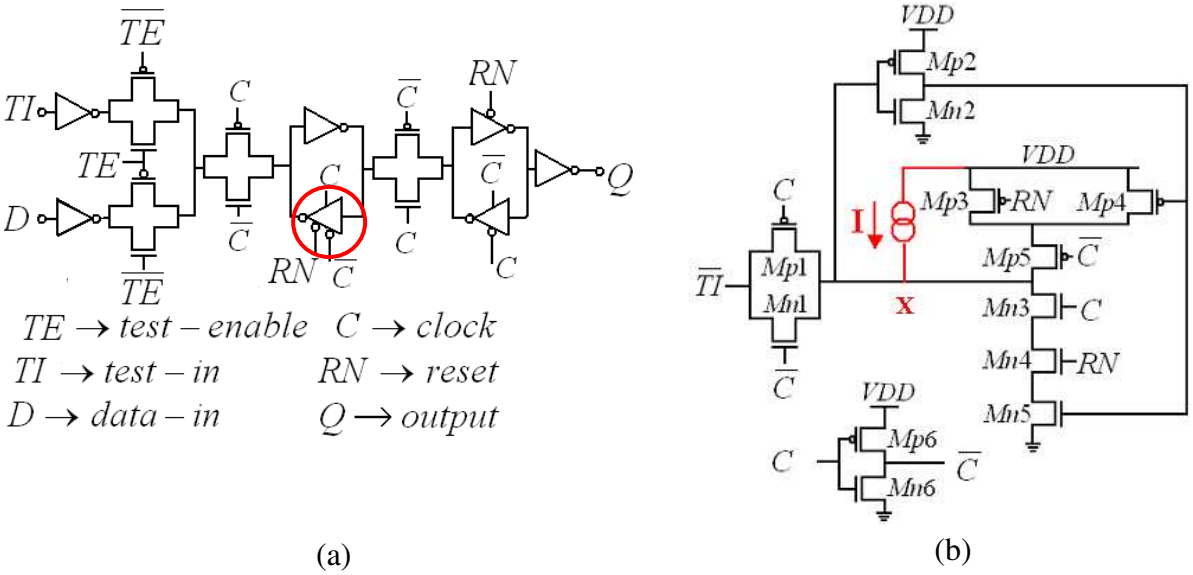


Figure 10.9. The circuit schematic of the scan flip-flop in transistor level



(a) (b)

Figure 10.10. (a)The circuit schematic of the scan flip-flop in logic level (b) The master stage of the scan flip-flop with connected current source modeling the PLS effect

The clock pulse, laser pulse, input and the output of the scan cell and the *Vx* signals are shown in Figure 10.11 and Figure 10.12 for fault injection and no fault injection cases, respectively.

There are two critical input voltage points for an inverter, V_{IL} and V_{IH}. Any input voltage level between the lowest available voltage in the system and V_{IL} is interpreted as logic "0" while any input voltage level between the highest available voltage in the system and V_{IH} is interpreted as logic "1" (see Appendix A).

The current source connected to node X, increases the node voltage. If the voltage exceeds the V_{IH} of the (Mp2, Mn2) pair inverter, it will be interpreted as logic high and will be amplified by the feedback configuration of the flip-flop. The modified logic value will be conveyed to the second stage and finally to the output causing a bit-flip from logic high to logic low at the scan cell output. *Vx* may exceed the supply voltage value, which may result in damage in the circuitry.

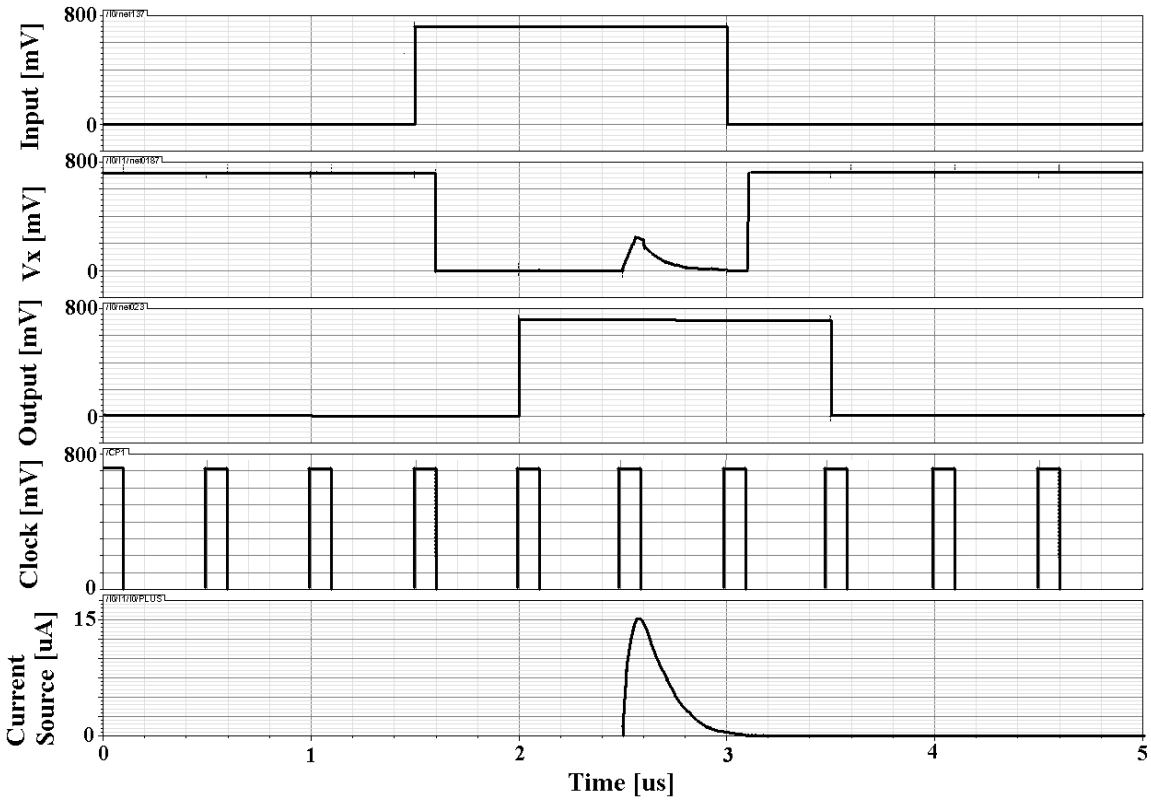


Figure 10.11. Input, Vx, output, scan clock and current source signals simulated in time for no injection case

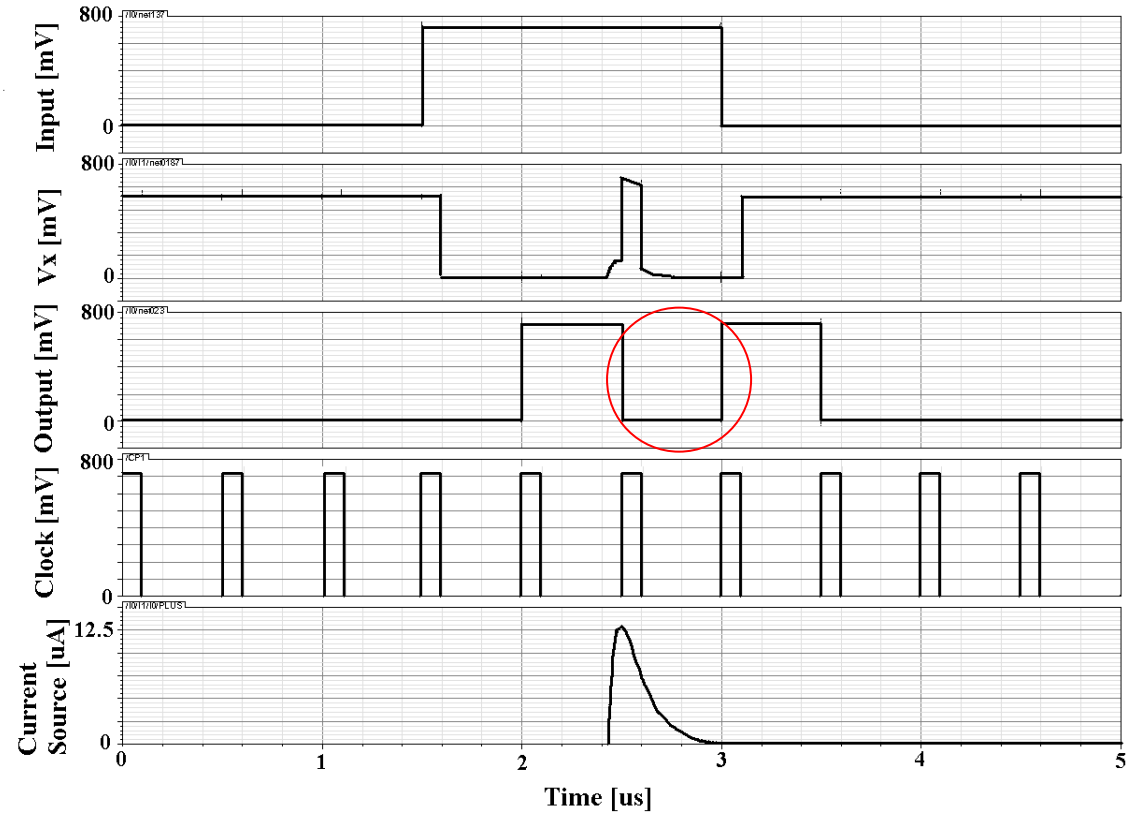


Figure 10.12. Input, Vx, output, scan clock and current source signals simulated in time for the injection case

Mp1, Mp3, Mp4, Mp5 transistors are situated in the same active area, thus very close to each other and contributing to the induced net current. Exposing the transistor Mp3 and Mp4 has a similar effect like Mp5.

The second sensitive logic in the master stage is the nMOS network in the feedback loop inverter. If the Mn3, Mn4, Mn5 transistors are illuminated (shown in Figure 10.13), the net induced current will be from the drain terminal of the exposed transistor to the ground which charges off the load capacitance at node X and decreases the node voltage. The stimulation at the nMOS network is not as strong as the stimulation on the pMOS network. Because, the induced drain current on parallel transistors, Mp3 and Mp4 sums up and flow over Mp5. On the other hand, there is no such a parallel topology in nMOS network. The transistors are connected serially. Therefore, the total collected current at pMOS network is higher than on nMOS transistors. Besides, as discussed in section 6.1.2, the induced current on pMOS transistors is larger than on nMOS transistors.

Similarly, exposing the nMOS network in the slave stage drops off the node voltage at Y (Figure 10.14). Due to serial connection topology between the transistors, the effect is weak like the nMOS stimulation in the master stage. The fault injection intensity at nMOS network is 89% weaker than the stimulation in the pMOS network.

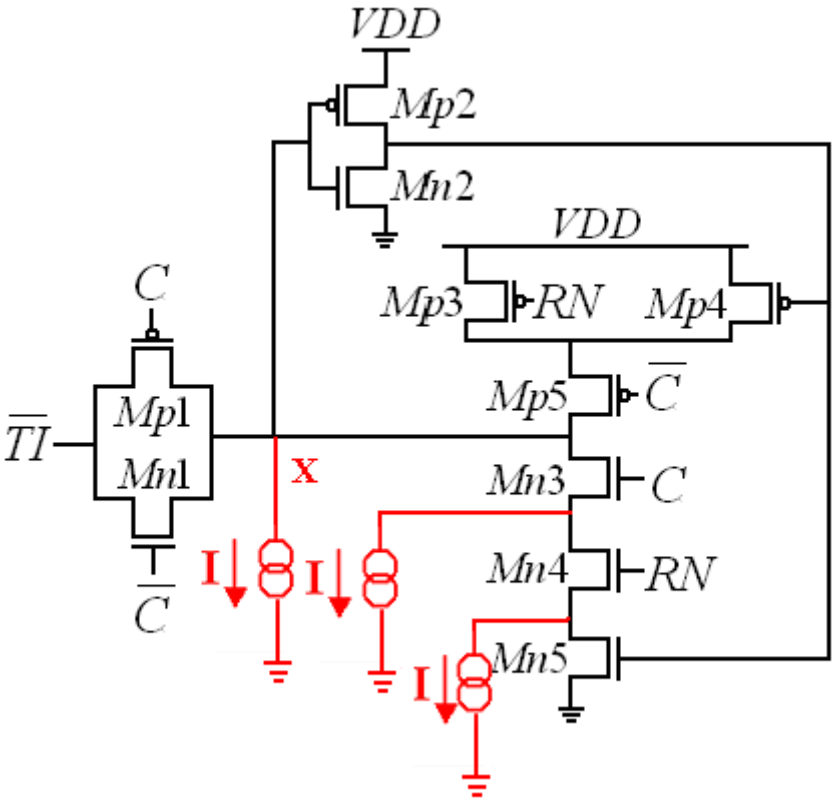


Figure 10.13. The master stage of the scan flip-flop with the current sources modeling the laser stimulation on nMOS-network

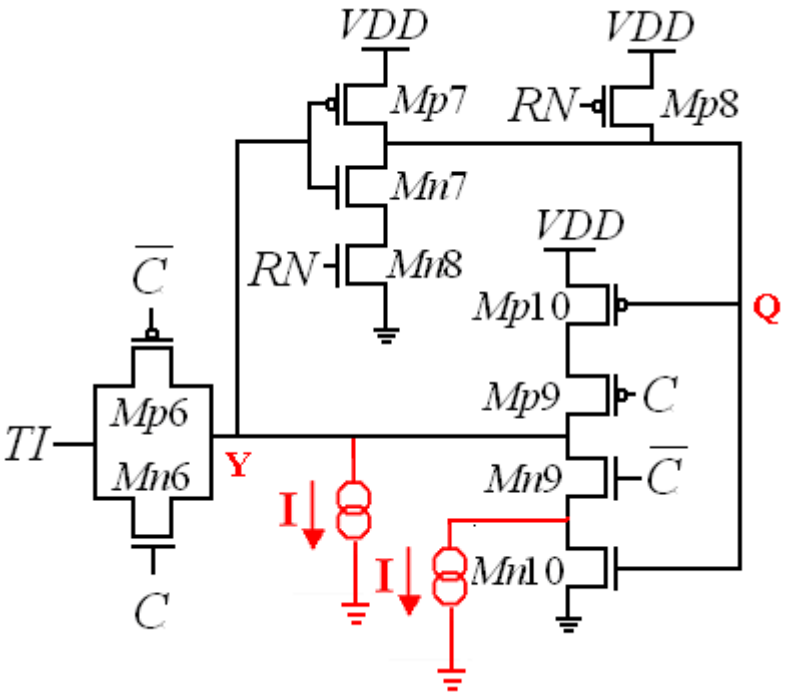


Figure 10.14. The slave stage of the scan flip-flop with the current sources modeling the laser stimulation on nMOS-network

There is no fault injection caused by pMOS network in the slave stage (Figure 10.15). Mp6, Mp7, Mp8, Mp9 and Mp10 are in the same active volume and the generated charge carriers in this active volume will give a rise to the drain currents of all these pMOS transistors. While the additional drain current at Mp9, Mp10 is increasing the voltage at node Y, photocurrents induced at Mp7 and Mp8 terminals work against to that and increases the voltage at node Q pulling down the voltage at Y.

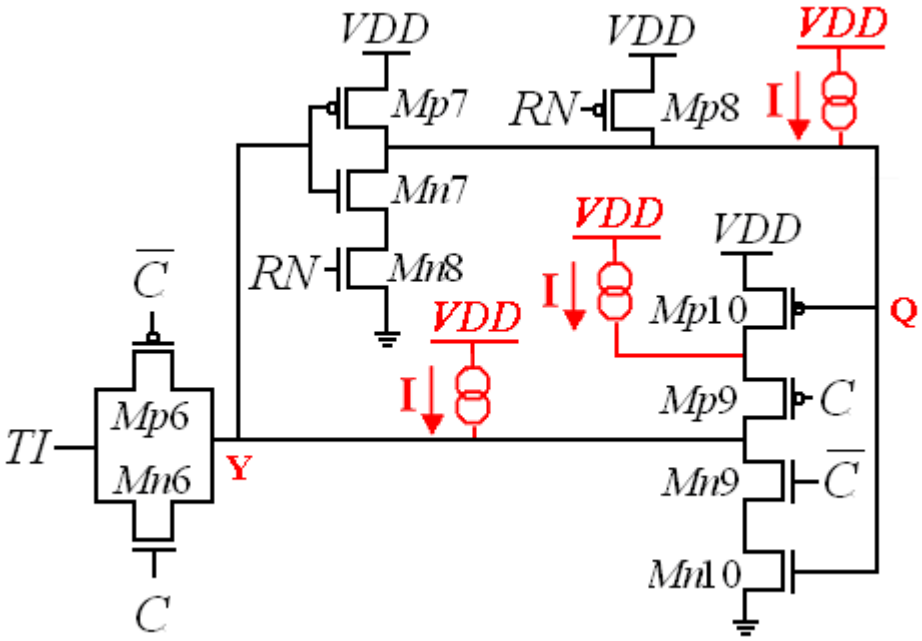


Figure 10.15. The slave stage of the scan flip-flop with the current sources modeling the laser stimulation on pMOS-network

10.2.1. Timing Sensitivity

We observed that the best injection occurs when the time difference between the current peak and the clock pulse is zero. Moreover, it will be proven that the time window, in which the fault injection occurs, also depends on the maximum peak current value, rising/falling time constants and the pulse width of the current source.

An exponential current source which has 50 ns rise time constant and 100 ns falling time constant was applied; simulated and timing mapping is derived. The exponential pulse is shifted in negative and positive direction by adjusting TDR and TDF parameters. The acquired current source-delay diagram is shown in Figure 10.16. Green boxes represent soft fault injection case whereas red boxes indicate no fault injection. The required current for fault injection is around 12.4 μA when the clock-to-laser pulse delay is zero and it gets greater for larger delays (13.4 μA for -30 ns delay). We can conclude that the required current for the fault injection drops off with the decreasing clock-to-laser pulse delay, which was also investigated experimentally (section 9.2.4).

The reason why the fault injection depends strongly on the delay can be explained by studying noise margins of the clock buffer and the inverters. The noise margins are explained in more detail in Appendix A. We will discuss the case that is shown in Figure 10.10. The current source is connected between VDD and drain terminal of the Mp5. There are two different mechanisms happening in the circuit considering the clock voltage. Let's assume that the clock voltage is increasing from 0 V to V_{IH} voltage level of the clock buffer.

If the clock voltage is lower than V_{IH} voltage level of the clock buffer, Mn1 in the transfer gate works in linear region and Mp1 is in cut-off region while the clock signal is still interpreted as logic "0". In this case, the charge carriers, induced in the active volume of the illuminated transistors (the current supplied by the current source in the simulation) are drained via the transfer gate. When the current increases more and more, these transistors start to work in saturation region, which results in high V_{DS} voltage on the transfer gate transistors. As a result, the voltage increases at node X. If it increases up to V_{IH} value of Mp2, Mn2 pair inverter, it will be interpreted as logic high causing a bit flip.

When the clock voltage exceeds V_{IH} of the clock buffer, the clock voltage will be interpreted as logic high and the transfer gate becomes OFF. The current does not flow over transfer gate anymore, but nMOS transistors (Mn3, Mn4, Mn5) drain the current. Increasing the current source value raises the voltage at node X. When the current exceeds a certain value, Vx voltage approaches to V_{IL} voltage level of (Mp2, Mn2) pair inverter, which causes node Y to be logic "0". Thus, Mn5 goes OFF and the pMOS transistors (Mp3 and Mp4) enter from cut-off to linear region. The voltage at node X jumps to 1V and the current starts flowing from node X to VDD via transistors Mp4 and Mp5.

As a result, the required current to cause a bit flip is pretty low when the clock voltage is lower than V_{IH} value of the clock buffer due to two different current drain loop mechanisms. Therefore, fault injection is more likely at the rising/falling edges of the clock compared to logic high state of the clock where the voltage is definitely over V_{IH}.

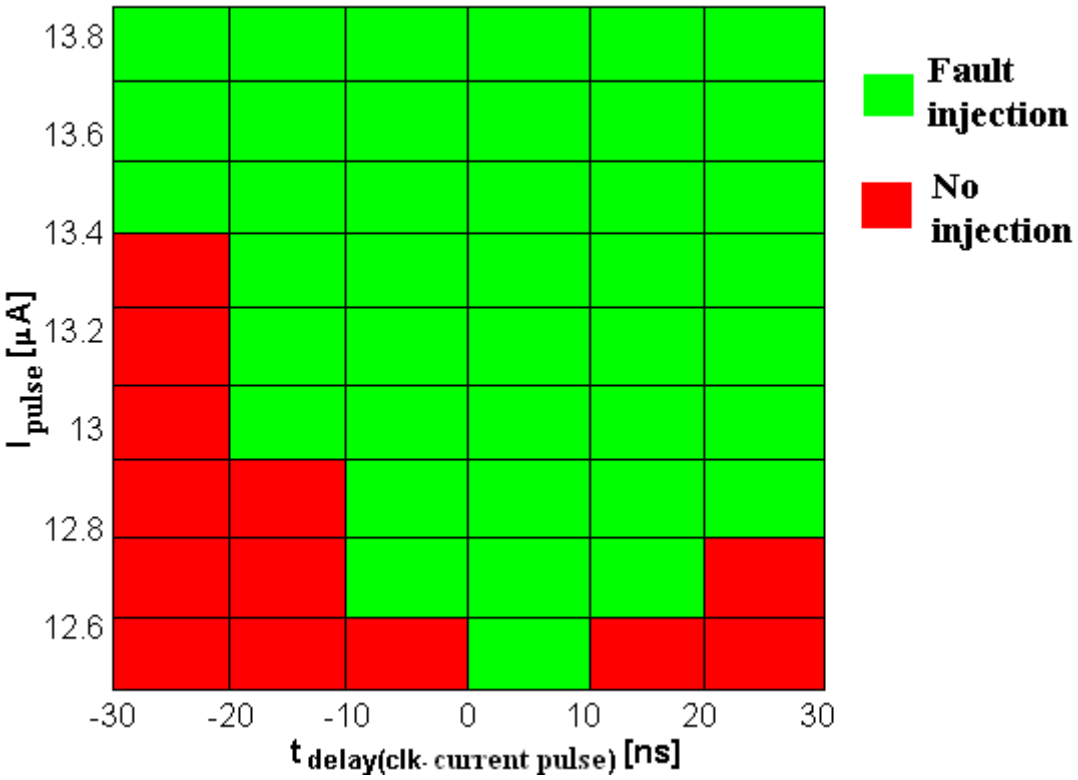


Figure 10.16. Clock to current pulse delay-current pulse diagram

The peak current should provide minimum required current level for fault injection. On the other hand, the timing window narrows down, as the maximum peak current gets lower (lower laser power in terms of experiments). Knowing that the minimum required current for fault injection is 12.5 µA, two different current peak values are applied to the circuitry, 15 µA and 12.6 µA (shown in Figure 10.17). The area under the current curve until the minimum required current level determines the width of the timing window. The effect of the current peak value on the timing window is shown in Figure 10.18. The time window width for a current peak value of 12.5 µA is around 12 ns whereas it is 26 ns for 12.6 µA current source amplitude.

A lower rising/falling time constant will cause the edges to be less steep, leading to a larger current source pulse width and thus a larger timing window. When a laser incident is present on the active volume of the devices, excess carriers are generated. The excess carrier concentration at a given time and point depends on the diffusion coefficient, electric field, mobility and carrier lifetime. Since the laser beam is absorbed by different layers that have different doping concentrations, it is hard to predict the carrier life time. It also depends on the capture rate for electrons and holes at recombination centers. Therefore, two different rising and falling time constants are applied and compared as shown in Figure 10.19. At first, an exponential current source with 40 ns rising time constant and 50 ns falling time constants is applied to the circuitry and the timing window is evaluated and found to be 14 ns. When both time constants are increased 10 ns, the time window width rises up to 29 ns.

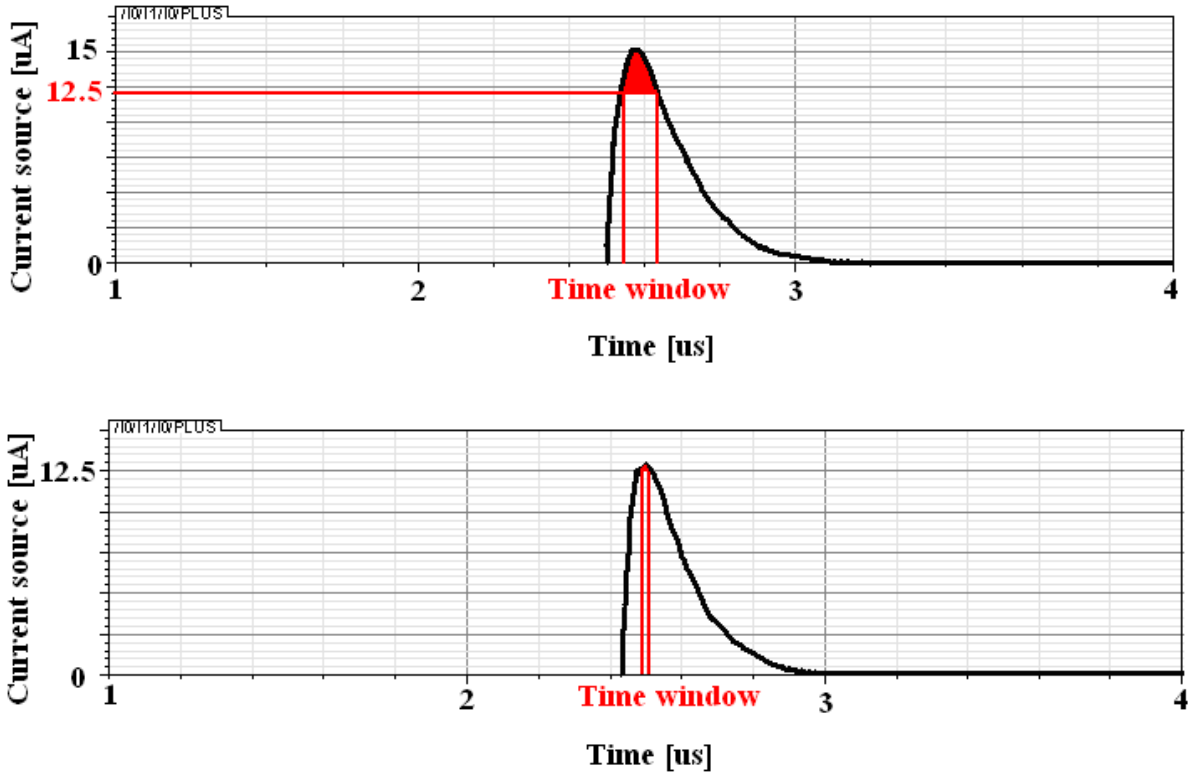


Figure 10.17. The exponential current sources with two different peak values, 12.6 µA and 15 µA, showing the difference between the timing window widths

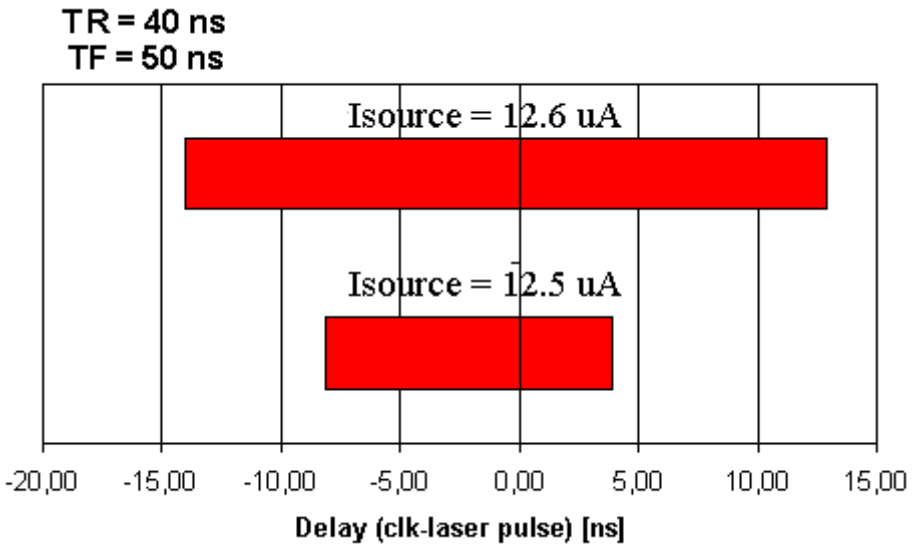


Figure 10.18. Timing windows, which the fault injection happens for different current source peak values

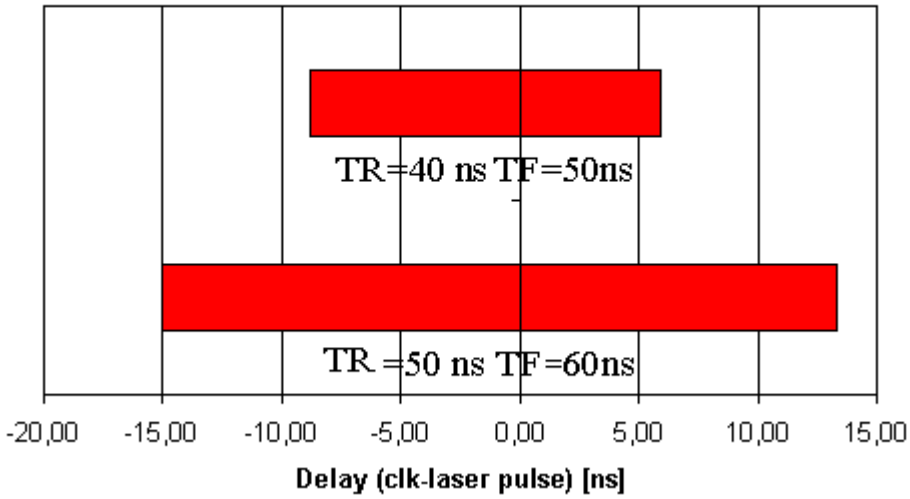


Figure 10.19. Timing windows, which the fault injection happens for different rising and falling time constants

Simulation is a great opportunity to understand the consequences of the photoelectric effect at circuit level. The timing sensitivity of the scan flip-flop to the photoelectric laser stimulation is reproduced at software level. Investigation of all the dependencies might help to improve the experiments. It is shown that the timing resolution can get smaller by adjusting the parameters.

Some experimental parameters can be varied, like laser power or laser pulse width in order to tune the timing sensitivity window of the laser stimulation. Practically, the time resolution can be improved if the laser power is tuned until the minimum laser power required for fault injection or the laser pulse width can be decreased. But they have some limitations. For example, it is not possible to adjust the rising and the falling time of the laser pulse. But simulations give a wide range of freedom in parameter selection. The timing window gets even smaller if the TDR and TDF of the current source are decreased as shown in Figure 10.19.

11. Future Prospects

It is assumed that dynamic analysis of ICs will keep on progressing especially due to the increasing clock rates. Operation in higher frequency ranges requires a high-level quality delay testing capability.

Smaller feature sizes lead to reduction in node capacitance and supply voltage. As a result, the logic edge transition times have dropped off down to 10-50 ps. These improvements enable the use of laser stimulation techniques for timing analysis. Because then, the induced delay variations are comparable with the gate delays. Therefore, dynamic laser stimulation techniques applied from the backside of the devices will gain even more importance in the future, since it is able to isolate the performance limiting circuitry and is possible to locate the soft defects which cause timing failures.

Scan chain design ICs are widely used for testing in failure analysis laboratories. Therefore, timing investigations of scan design circuits are very important. We have shown with this work that PLS technique injects soft faults into internal nodes of scan flip-flops. It is also shown that laser stimulation may delay or advance the switching speed of the devices. These two effects might be combined by applying a transition delay vector pattern instead of a scan shift vector pattern to the scan chain structures.

As it is explained in Chapter 3, a transition delay pattern is able to find the large delay failures. Laser stimulation method might be used to cause a larger delay at a logic gate in a defect free IC.

A scan chain structure is illustrated in Figure 11.1. Laser stimulation positioned on the combinational logic may delay the signal at node X as shown in Figure 11.2. If the induced delay is large enough, the signal at node X will switch from logic low to logic high after the capture clock rising edge. And the scan flip-flop connected to the node X will capture the previous value (logic low instead of logic high). If this failure is transferred to one of the observable outputs, the performance limiting circuitry will be localized.

This technique would improve the timing characterization of the devices one step further. On the other hand, it has some difficulties. Because the induced delay variations are in the picosecond ranges. Therefore, the scan clock frequency should be very high which is hard to apply in a conventional failure analysis laboratory. The only way might be the usage of internal phase locked loop circuitry, which allows high frequency operation in normal mode.

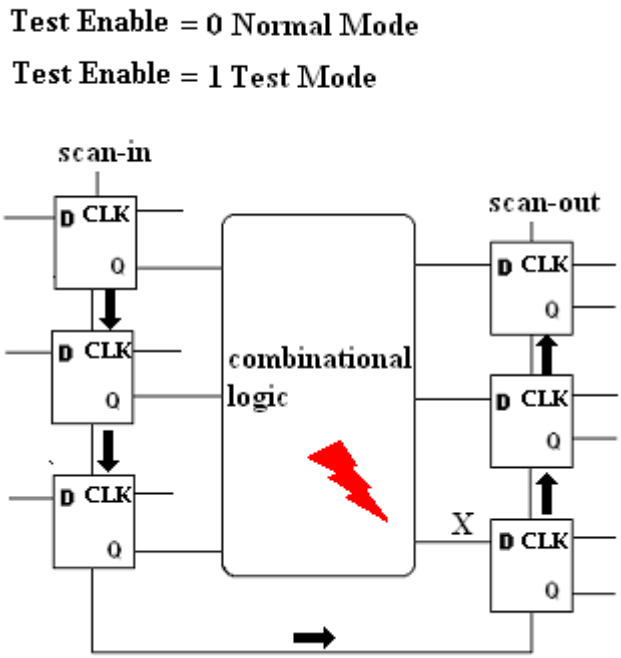


Figure 11.1. The illustration of a scan chain

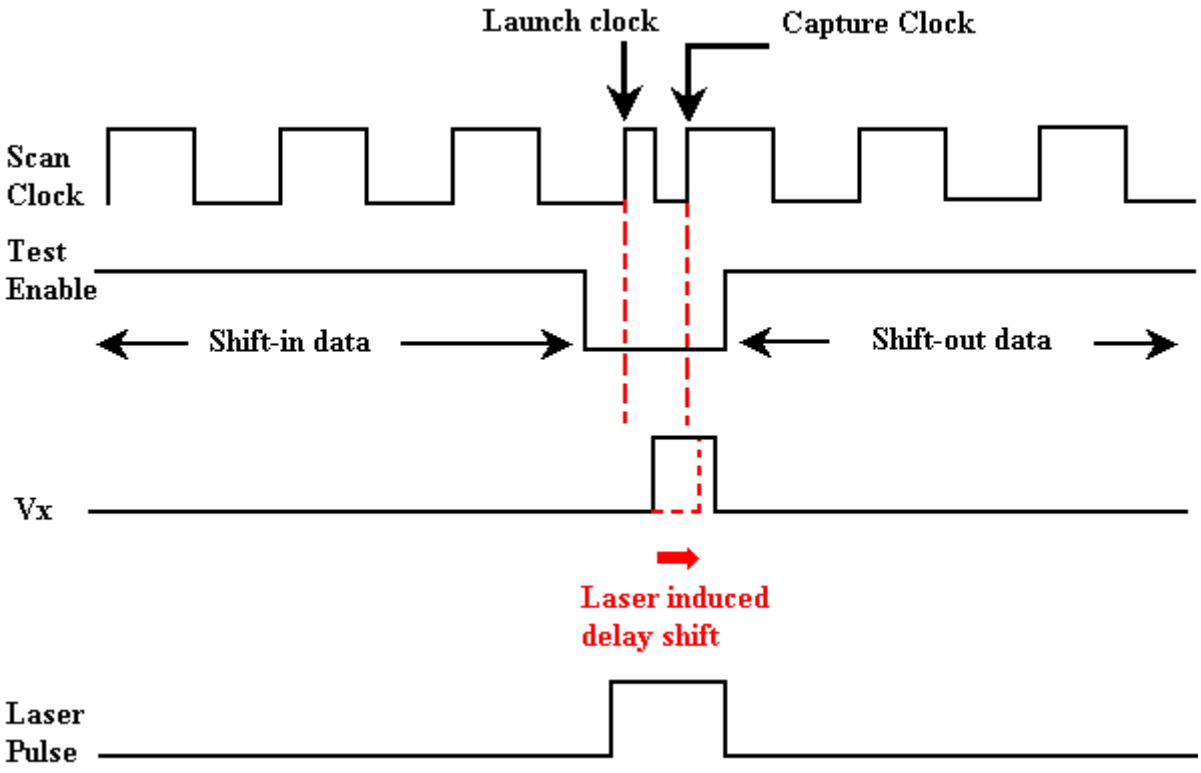


Figure 11.2. The illustration of the signals for a transition delay vector pattern

12. Conclusion

Dynamic laser Stimulation techniques have been applied to the delay chain and scan chain structures on a “real” test IC. The DUT is provided by Infineon AG and the minimum feature size is 90 nm. It was possible to choose between two laser diodes with different laser wavlength: 1300 nm and 1064 nm. 1064 nm laser is applied both in pulsed and CW mode.

The experiments on delay chains prove that laser stimulation techniques perturb the switching time of the devices in the ranges of picoseconds. Various experimental setups were built in order to obtain the best timing accuracy with the lowest possible acquisition time. The tester was used to drive DUT and different measurement units were employed, such as tester, TAC, lock-in amplifier and high bandwidth oscilloscope.

It was shown that a 1300 nm wavelength laser beam changes the electrical characteristics of the device via localized heating on interconnects and highly doped areas, which perturbs the switching speed of the devices. If the transistor is illuminated, operation in high gate voltages of a transistor slows down the switching speed whereas low gate voltages speeds up. Stimulation by TLS technique is very weak, but free of secondary effects.

Alternatively, 1064 nm wavelength laser beam injects localized photocurrent into the Si, which influences the propagation delay of the devices which has a bigger effect than TLS. It may advance or delay the switching time depending on the type of the transistor (n or p) and the edge of the output (rising or falling). Since, illumination generates charge carriers an all pn junctions, this technique is not free of secondary effects. Some dummy structures, which are in the same n-well with the transistors, might also change the switching speed of the device via n-well charging. On the other hand, it was shown that the use of pulsed laser minimizes the secondary effects from nearby structures.

Continuous-wave and pulsed photoelectric laser stimulation on scan chains causes soft fault injection in logical nodes of the scan flip-flops, which results in the modulation of the scan-out. Using the opportunity of shifting the laser pulse with respect to clock, pulsed laser stimulation was successfully utilized for time-resolved fault injection. The laser pulse is triggered on the falling and the rising edge of the clock. It is shown that triggering the laser pulse on the rising edge of the clock and exposing the p-MOS network in the inverter of a master stage injects a logic ‘1’ while illumination of n-MOS network causes logic ‘0’ injection. Hitting the n-MOS network with the laser and pulsing the laser on the falling edge of the clock causes a bit flip from ‘1’ to ‘0’ in the slave stage. It is shown that pulsed laser stimulation can be used for reverse-engineering applications in golden devices, for example to localize a desired scan flip-flop in the chain.

Besides, pulsed laser stimulation methodology on scan chain ICs identifies the most sensitive signal condition of a clock driven circuitry to the laser stimulation. This information gives a better insight into the physics of the laser stimulation impact on the operation of a clock driven circuitry. It is presented that the time resolution of the effective stimulation correlates with the sharpness of the signal pattern edges and the laser pulse as well as laser power for failure analysis applications. Using a laser diode pulsed by external triggering provides a very convenient opportunity to do a time mapping of stimulation sensitivity, which is a significant improvement for failure analysis.

It was also shown that a current source connected to the drain of the stimulated transistor models the additional drain photocurrent and it changes the switching speed of the transistors or causes an upset in the scan flip-flop. Moreover, timing sensitivity of the signals on the laser stimulation is reproduced in software level.

Bibliography

[ABF90] M. Abramovici, M. A. Breuer and A. D. Friedman, Digital systems testing and testable design, revised printing, AT & T, 1990.

[Agi00] Agilent 83000 F330 user training manual, Agilent Technologies GmbH, 2000.

[BA00] M. Bushnell and V. Agrawal, Essentials of electronic testing for digital memory and mixed-signal VLSI circuits, Kluwer Academic Publishers, 2000.

[BBG05] S. K. Brahma, C. Boit, A. Glowacki, Seebeck effect detection on biased devices without OBIRCH distortion using FET readout, Microelectronics Reliability, vol. 44, issue 9, pages 1487-1492, 2005.

[Ble86] M. Bleicher, Halbleiter-Optoelektronik, Hüthig Verlag Heidelberg, 1986.

[BBE02] M.R.Bruce, V.J.Bruce, D.H.Eppes, J.Wilcox, E.I. Cole Jr., P.Tangyunyong, C.F.Hawkins, Soft defect localization on ICs, Proceedings of the 28th International Symposium for Testing and Failure Analysis (ISTFA), Phoenix, Arizona, pages 21-27, 2002.

[BGB04] C. Boit, A. Glowacki, S. K. Brahma, K. Wirth, Thermal laser stimulation of active devices in silicon - a quantitative FET parameter investigation, Proceedings of 42nd IEEE International Reliability Physics Symposium, Phoenix, AZ, pages 357-360, 2004.

[Boi04] C. Boit, Fundamentals of photon Emission (PEM) in silicon- electroluminescence for analysis of electronic circuit and device functionality, Microelectronic Failure Analysis Desk Reference, Fifth Edition, ASM International, pages 356-368, 2004.

[BWK87] S. Buchner, D. Wilson, K. Kang, D. Gill, J. Mazer and W. Raburn, Laser simulation of single event upsets, IEEE Transactions on Nuclear Science, vol. 34, pages 1227-1233, 1987.

[CC78] C. Cremer, T. Cremer, Considerations on a laser scanning microscope with high resolution and depth of field, M1CROSCOPICA ACTA, Vol. 81, No.1, pages 31-44, 1978.

[Chi00] W. K. Chim, Semiconductor Device and Failure Analysis: Using Photon Emission Microscopy, John Wiley & Sons, 2000.

[CL05] I. Chun, C. Lim, ES-debugger : the flexible embedded system debugger based on JTAG technology, The 7th International Conference on Advanced Communication Technology,Volume 2, pages 900-903, 2005.

[Col04] E. I. Cole Jr., Beam based defect localization methods, Microelectronic Failure Analysis Desk Reference, Fifth Edition, ASM International, pages 406-417, 2004.

[Cro99] A. L. Crouch, Design for test for digital IC's and embedded core systems, Prentice Hall, 1999.

[Cro05] A. L. Crouch, Debugging and diagnosing scan chains, Magazine article from Electronic Device Failure Analysis, Vol. 7, Issue 1, 2005.

[CTH01] E. I. Cole Jr., P. Tanyunyong, C.F. Hawkins, M.R. Bruce, V.J. Bruce, R.M. Ring, W. Chong, Resistive interconnect localization, Proceedings of the 27th International Symposium for Testing and Failure Analysis (ISTFA), Santa Clara, California, pages 43-50, 2001.

[FD07] T. J. Fellers, M. W. Davidson, Introduction to confocal microscopy, Olympus

Fluoview Resource Center, 2007.

[GBS07] A. Glowacki, S. K. Brahma, H. Suzuki and C. Boit, Systematic characterization of integrated circuit standard components as stimulated by scanning laser beam, IEEE-TDMR, vol. 7, no. 1, pages 31-49, 2007.

[Ham] Emission microscope system-PHEMOS 1000, Instruction Manual, Hamamatsu Photonics.

[Ham03] Characteristics and use of FFT-CCD area image sensor, Technical notes, Hamamatsu Photonics, 2003.

[HS51] J. R. Haynes and W. Shockley, The mobility and life of injected holes and electrons in germanium, Phys. Rev., vol. 81, pages 835-843, 1951.

[LGB08] P. Laskowski, A. Glowacki, C. Boit, Detectability of dynamic photon emission in static Si CCD for signal path determination in integrated circuits, Microelectronics Reliability, vol. 48, issue 8, pages 1295-1299, 2008.

[LPB01] Dean Lewis, Vincent Pouget, Felix Beaudoin, Philippe Perdu, Herve Lapuyade, Pascal Fouillat, and André Touboul, Backside Laser Testing of ICs for SET Sensitivity Evaluation, IEEE Transactions On Nuclear Science, Vol. 48, No. 6, pages 2193-2201, 2001.

[MFS98] W. Moreno, F. Falquez, N. Saini, Fault tolerant design validation through laser fault injection, Proceedings of IEEE International Caracas Conference on devices, circuits and systems, pages 132-137, 1998.

[Nea03] D. A. Neamen, Semiconductor physics and devices: basic principles, McGraw-Hill, 2003.

[Paw06] J.B. Pawley, Handbook of Biological Confocal Microscopy, Springer, 2006.

[PLF04] V. Pouget, D. Lewis, P. Fouillat, Time-resolved scanning of integrated circuits with a pulsed laser: application to transient fault injection in an ADC, IEEE Transactions on Instrument and Measurement, vol. 53, pages 1227-1231, 2004.

[Rab96] J. M. Rabaey, Digital integrated circuits: a design perspective, Prentice Hall, 1996.

[RE03] J.A. Rowlette and T.M. Eiles, Critical timing analysis in microprocessors using near-ir laser assisted device alteration (lada), Proceedings of the IEEE International Test Conference (ITC), Vol.1, pages 264-73, 2003.

[SB87] R. A. Soref and B. R. Bennett, Electrooptical effects in silicon, IEEE Journal of Quantum Electronics, QE-23, 1987.

[SDB05] K. Sanchez, R. Desplats, F. Beaudoin, P. Perdu, S. Dudit, G. Woods, D. Lewis, Dynamic laser delay variation mapping (DVM) implementations and applications, Proceedings of the 31st International Symposium for Testing and Failure Analysis (ISTFA), Santa Clara, California, pages 106-114, 2005.

[Shu90] M. Shur, Physics of Semiconductor Devices, Prentice Hall, 1990.

[Sie86] A. E. Siegman, Lasers, University Science Books, 1986.

[Sig00] Signal Recovery (Perkin Elmer Instruments) technical note, What is a lock-in amplifier?, 2000.

[SPB07] K. Sanchez, P. Perdu, F. Beaudoin, Phase variation mapping, a dynamic laser stimulation technique with picosecond timing resolution, Proceedings of the 45th IEEE International Reliability Physiscs Symposium, Phoenix, Arizona, pages 534-41, 2007.

[Sze98] S. M. Sze, Modern Semiconductor Device Physics, John Wiley & Sons, 1998.

[Tsi99] Y. Tsividis, Operation and modeling of the MOS transitor, The McGraw-Hill Companies, 1999.

[WC65] P. H. Wendland and M. Chester, Electric field effects on indirect optical transitions in silicon, Phys. Rev., 140(4A):1384-1390, 1965.

Further Dynamic Laser Stimulation Related Publications of the Author

[KBB08] T. Kiyan, C. Brillert, C. Boit, Timing analysis of scan design integrated circuits using stimulation by an infrared diode laser in externally triggered pulsing condition, Microelectronics Reliability, Volume 48, Issues 8-9, pages 1327-1332, 2008.

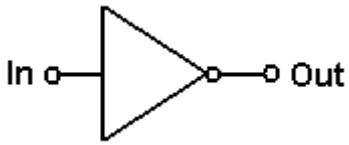
[KBB08/2] T. Kiyan, C. Brillert, C. Boit, Timing Sensitivity Analysis of Logical Nodes in Scan Design Integrated Circuits by Pulsed Diode Laser Stimulation, Proceedings of ASM/EDFAS 34th International Symposium for Testing and Failure Analysis (ISTFA), Portland OR, USA, pages 180-187, 2008.

[B+08] Boit et al., Physical IC debug - backside approach and nanoscale challenge, Advances in Radio Science (ARS), Kleinheubacher Berichte, Volume 6, pages 265-272, 2008.

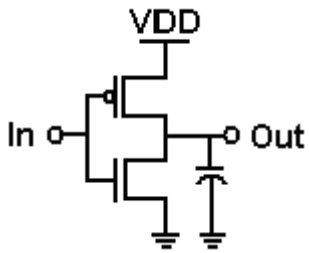
Appendix A

The Noise Margin of CMOS Inverters

The inverter is the most fundemental structure in digital logic design. The logic symbol, circuit diagram and the truth table of a CMOS inverter is seen in Figure A.1.



(a)



(b)

In	Out
0	1
1	0

(c)

Figure A.1. (a) The logic symbol, (b) circuit diagram and (c) the truth table of a CMOS inverter

The voltage transfer characteristic (VTC) of a CMOS inverter is seen in Figure A.2. As it is seen, the VTC of the inverter exhibits a very narrow transisiton zone. There are two critical voltage points on this curve, where the gain of the inverter becomes equal to –1. (In case of a buffer, the gain is 1). Both of these points has a significant importance determining the noise margins of the inverter.

The definitions of critical voltages are given below.

V_{OH} : Maximum output voltage when the output level is logic “1”

V_{OL} : Minimum output voltage when the output level is logic “0”

V_{IL} : Maximum input voltage which can be interpreted as logic “0”

V_{IH}: Minimum input voltage which can be interpreted as logic “1”

The shape of the VTC determines the noise immunity properties of an inverter. Therefore the voltage levels defined above are the most important criteria in the means of an inverter design.

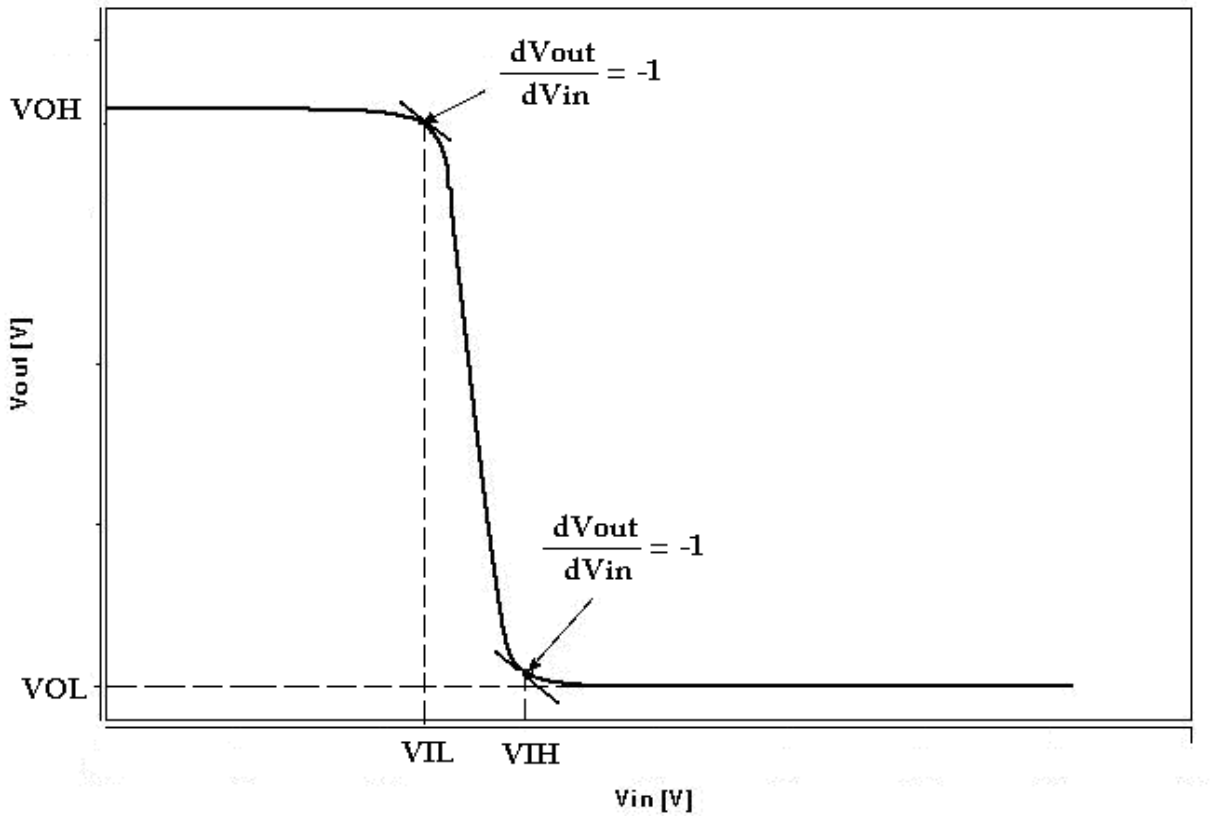


Figure A.2. Typical voltage transfer characteristic of a CMOS inverter

Appendix B

Propagation Delay of an Inverter

The propagation delay formula of an inverter can be derived as follows which can be found in the literature [Rab96]. An inverter can be simplified with a switch model as seen in Figure B.1.

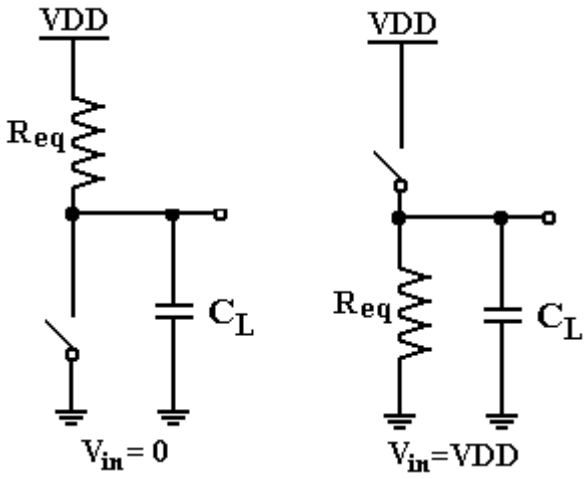


Figure B.1. Switch model of CMOS inverter

The propagation delay is the time required for the output to reach 50% of its final output level. Therefore, VDD/2 is an important output voltage level considering the propagation delay. Assuming that the transistor stays in saturation during the whole transition, we can write

$$R_{eq} = \frac{1}{-V_{DD}/2}\int_{VDD}^{VDD/2}\frac{V}{I_{DSAT}(1+\lambda V)}dV = \frac{3}{4}\frac{V_{DD}}{I_{DSAT}}(1-\frac{7}{9}\lambda V_{DD}) \quad \text{B.1}$$

where I_{DSAT} is the saturation current and λ is the channel length modulation.

The operation of this simple RC network is an exponential function if a step input is applied. It can be expressed as

$$V_{out}(t) = (1-e^{-t/R\times C})V \quad \text{B.2}$$

The time to reach to VDD/2 is computed as

$$t = \ln(2)R_{eq}\times C_L = 0.69R_{eq}\times C_L \quad \text{B.3}$$

Combining Equations B.1 and B.3 and ignoring the channel length modulation, we get

$$t = 0.69\frac{3}{4}\frac{C_L V_{DD}}{I_{DSAT}} = 0.52\frac{C_L V_{DD}}{I_{DSAT}} \quad \text{B.4}$$

Acknowledgment

This research project would not have been possible without the support of many people.

First and foremost, I would like to express my gratitude to my supervisor, Prof. Dr.-Ing. Christian Boit, who was abundantly helpful and offered invaluable assistance, support and guidance. His encouragement, patience, especially enthusiasm and optimism cleared dark clouds and gave me motivation whenever I needed.

Deepest gratitude is also to Dr. Christof Brillert from Infineon, Munich without whose knowledge and assistance this study would not have been successful. Thank you for fruitful discussions!

I am deeply grateful to my ex-supervisor and the referee of my dissertation, Prof. Dr. Atilla Ataman for encouraging me to start my PhD in TU-Berlin, for his reporting and his interest in this work.

My thanks to Uwe Voss for introducing me to the tester and to repair it immediately whenever it gets broken, Werner Eschenberg for his help in computer related issues and Rene Hartmann for bonding my devices.

All employees of Hamamatsu Photonics, Herrsching for their endless support and immediate reaction when PHEMOS or laser diode gets broken over and over again.

All employees of Infineon Technologies AG, who were dicing and bonding devices, were very supportive. I wish to thank especially Christian Burmer and Siegfried Görlich for making this cooperation possible.

The generous financial support from TU-Berlin, especially Harald Ermel from Foreign Relations Office, is greatly appreciated. They granted me the scholarship. Without their support, my ambition to study abroad could hardly be realized.

I wish to thank Andreas Eckert and Marlies Mahnkopf for dicing and packaging my devices.

My thanks to all of my undergraduate professors who brought me up to this level.

I offer my regards and blessings to all of my colleagues who supported me in any respect during the completion of my work. I truly thank Peter for his friendship, academical or personal advices over years and sticking by my side, even when I was irritable and depressed. I am very happy to have such a lovable friend by my side. Rudolf was a friend to me more than just being a colleague. I wish to thank him for opening the doors of his WG, introducing me to his friends, and helping me to settle down into a completely new environment. I will cherish our WG dinners and the time that we all had together in Kreuzkölln. My thanks to Pon for sharing very warm friendship in such a short time and having deep conversations accompanied with his delicious Taiwanese food, Piotr for teaching me to say ‘NO’ when it comes to PHEMOS scheduling and the nice accompany for dinner after work and before my dance classes, Titu for picking me up at Flughafen Tegel on that sunny day-my first day in Berlin, Ulrike for useful advices about PhD, jobs and life, Arek for constructive discussions, Philipp, Mahyar and Oana for picking me up for lunch everyday, and coffee sessions after, Dr. Uwe Kerst for his assistance in laboratory related issues, my ex-colleagues Nihan, Burcu, Revna for encouraging me when I first set off for Berlin from Istanbul and not forgetting their best-friend over years.

I have made many friends along the way. They have helped me, one way or another, in my struggle to complete a PhD. Many thanks to Ebru, Mahmut, Ruzin, Ivana, Berrin, Mine,

Serdar, Ines, Shireen, Özge, Robab, Senem, Seher, my ex-flatmates Julia, Nicolas and Anna who turned a shared-flat into our home. They have listened my hard-to-solve PhD problems and watched my sour face over years. They have given me strength in struggling with the problems. I cherish every each second of Berliner days and nights that I had with my beautiful friends.

Last but not least, I would like to express my love and gratitude to my beloved extended family. It is, however, not possible to list them all here. My parents provided me with a loving home, one where an academic mind was celebrated, and my brother, Sertan's sharp sense of humour has been a gift in my life. I wish to thank them for their understanding, encouragement and unconditional love, through the duration of my PhD.

Tuba KIYAN