Picosecond imaging circuit analysis

by J. C. Tsang
J. A. Kash
D. P. Vallett

A newly developed optical method for noninvasively measuring the switching activity of operating CMOS integrated circuit chips is described. The method, denoted as picosecond imaging circuit analysis (PICA) can be used to characterize the gate-level performance of such chips and identify the locations and nature of their operational faults. The principles underlying PICA and examples of its use are discussed.

1. Introduction

We have recently developed a noninvasive, passive optical method for measuring electrical switching in CMOS integrated circuits [1, 2]. The technique, PICA, is based on the detection and analysis of the weak light pulses which are emitted by field-effect transistors (FETs) during switching.[1] The need for PICA has been driven by the exponential increase over the last quarter century in the density, complexity, speed, and capabilities of semiconductor integrated circuits. The most recent wave of progress has imposed new physical constraints on electrical measurements that make previously used techniques obsolete. The technology and science behind PICA rely on the dramatic improvement in measurement methods for temporally resolving extremely faint optical signals on picosecond time scales, as well as our rapidly expanding knowledge of ultrafast processes in solid-state physics. Combining these advances has allowed us to develop PICA for the quantitative characterization of switching activity in silicon integrated circuits. The PICA technique should satisfy many of the critical needs of the integrated circuit (IC) industry for electrical and physical failure analysis and characterization in the foreseeable future.

In this paper, we first discuss the changes in silicon IC technology that render existing methods of measuring electrical waveforms in chips obsolete. We then describe the physical mechanism underlying PICA, i.e., the processes that produce picosecond pulses of light whenever FET gates are switching in ICs. We show how the techniques developed over the last quarter century to temporally resolve weak optical signals are well suited for the detection of the pulsed optical emission from working ICs. We close our discussion of PICA with examples of the types of information about normal and abnormal circuit operation that can be obtained today. We briefly describe the use of laser probe techniques as an alternative for acquiring information similar to that provided by PICA.

2. Trends and needs for test and back-side analysis in silicon integrated circuits

The quarter century of progress in IC technology captured by the phrase "Moore's law" has been characterized by substantial decreases in the size and spacing of active and passive structures such as transistors and their interconnecting wires, accompanied by increases in chip complexity, size, and speed of operation [3]. In recent years, these advances have also necessitated an increase in the number of metal layers connecting the devices within the chip, and chip inputs and

[1] C. F. Hawkins, J. M. Soden, E. I. Cole, and E. S. Snyder, Proceedings of the International Symposium for Test and Failure Analysis (ISTFA '90, USA), 1990, p. 55.

0018-8646/00/$5.00 © 2000 IBM

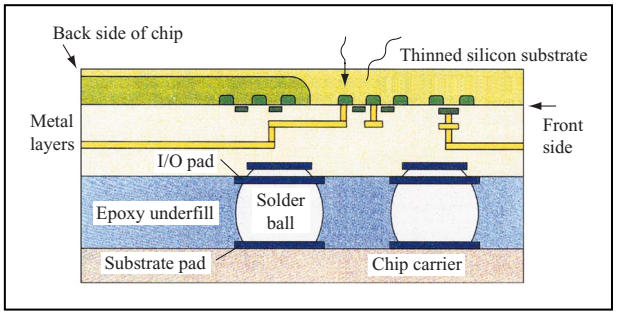


Figure 1

Schematic cross section of a modern silicon integrated circuit emphasizing the BEOL aspects of its structure: its layers of insulated metallization and its solder balls (for flip-chip packaging). The drawing is not to scale. From [5], with permission.

outputs, or I/Os, that are needed to connect the different chips in a system [4] and the use of "flip-chip" packaging methods. Dense "metal fill" patterns inserted in all areas not populated with signal wires are commonly used to optimize chemical–mechanical polishing processes. These layers of metal wiring are often referred to as the "back end of the line," or BEOL, portion of the IC fabrication process, as illustrated in **Figure 1** [5]. This dense BEOL, especially when used in combination with flip-chip packaging, has created a situation in which the transistors in an IC and the signal-carrying metal lines are no longer physically accessible to external probes from the front side, without the destructive disassembly of the chip.

It is often necessary for IC designers and manufacturers to measure the voltages and/or currents deep inside an operating IC. This need can arise, for example, when a newly manufactured product is not performing up to specifications, or an IC is returned that has failed in the field. At present, such measurements are carried out by means of two techniques. One uses an active or passive microscopic metal probe, which physically contacts a metal line of interest to extract an electrical signal. At present, such probes are capable of contacting lines of the order of 0.5 μm, and smaller probes are being developed using techniques based on the scanning probe microscope. As an alternate to a physical probe, a focused beam of electrons may be used as a probe of the conducting lines [6]. In this case, the sample is placed in a vacuum chamber, so that the electrons can interact directly with a conducting line. The electrons of the beam sense the time-varying electric field around the line. By analyzing the kinetic energy distribution of the resulting secondary electrons, a relative voltage on the metal line is determined. If the electron beam is incident as a short pulse, analysis of the energy of the secondaries gives the voltage during the time of the pulse. Conventional sampling techniques then allow a temporal measurement of a repetitive ac electrical waveform in the line.

The widespread use of a dense BEOL structure as well as flip-chip packaging poses a major challenge to these two techniques for measuring electrical activity in operating integrated circuits. Physical and electron-beam probing each require the metal line of interest to be at either an air–metal or vacuum–metal interface. A capacitively coupled form of electron-beam probing is possible that acquires signals through limited thicknesses of dielectric, but if the metal line is buried deep under other metal lines or thick insulators, or is covered because the chip is flipped over for packaging, neither measurement technique can be used without time-consuming and potentially destructive "deprocessing" of the chip. A focused ion beam (FIB) tool can be used to "drill" to the metal line of interest from the front or back side of the chip [5]. Lines can then be probed directly with the electron beam, or subsequent deposition of insulating material followed by a conductor (usually tungsten) creates a test point that is used with either probing technique. However, FIB is very time-consuming, and access is not universal, since target locations are often obscured by other lines or by transistors. More significantly, FIB alters the electrical properties of transistors in the area of interest. A further drawback is that all of the above probing techniques require signals to be acquired in a serial fashion, one at a time. For complex analysis efforts where little advanced knowledge of a failure or performance problem is available, this style of probing can be prohibitively time-consuming.

Because physical and electron-beam probing are becoming less and less practical, there is a demand for new analytic tools which can "remotely" sense electrical activity in an integrated circuit through the back side (i.e., the substrate side) of a silicon IC. The general subject of how changes in metallization practices and chip packaging present major new challenges to the failure analysis community has recently been reviewed by Vallett and Soden [5]. The 1997 edition of the National Technology Roadmap for Semiconductor Technology [7] describes the quantitative conditions that analytic tools for failure analysis and other types of chip diagnostics must satisfy in the next decade. **Table 1** shows projected changes in device sizes, operating voltages, device densities, operating frequencies, and sensitivities to gate-to-gate delays for the next decade. It also verifies that with the seven to nine layers of metallization and dense BEOL configuration that will be common, and the widespread use of flip-chip packaging, there will be no direct convenient physical

Table 1 Characteristics of integrated circuits between 1997 and 2012, as defined in the National Technology Roadmap for Semiconductor Technology, 1997 Edition [7].

Year of shipment	*Technology generation* (nm)	*Chip clock* (MHz)	*RMS clock jitter* (ps)	*No. of wiring levels*	*Chip size* (mm^2)	*No. of transistors per chip*	*Transistor density per* cm^2	*Voltage range* (V)	*No. of chip I/Os*
1997	200	750	27	6	300	11M	3.7M	1.8–5.5	1450
1999	140	1250	16	6–7	340	21M	6.2M	1.3–3.3	2000
2001	120	1500	13	7	385	40M	10M	1.3–3.3	2400
2003	100	2100	9.5	7	430	76M	18M	0.9–3.3	3000
2006	70	3500	5.7	7–8	540	200M	39M	0.9–2.5	4000
2009	50	6000	3.3	8–9	620	520M	84M	0.6–2.5	5400
2012	35	10000	2	9	750	1.4B	180M	0.6–2.5	7300

access to either the semiconductor devices or the wires carrying the electrical signal.

To clarify our terminology, Figure 1 shows a labeled schematic cross section of a modern flip-chip-packaged IC with multiple layers of metallization. The "front side" of today's chip includes the plane on which the silicon devices are fabricated, and upon which the wiring layers and I/O pads are deposited. The combination of the large number and density of wiring layers and the I/O pads means that the transistors and signal-carrying wire layers close to the device plane are almost completely obscured. In flip-chip packages, the chip is mounted on a carrier, front side down, with solder ball contacts bonding to the carrier so that there is external physical access to only the "back side" of the chip. The ability of light to penetrate a silicon substrate is indicated by the upper oscillating lines.

Because silicon is transparent to light at wavelengths longer than 1 μm when lightly doped, and moderately transparent near 1 μm even when heavily doped, it behaves as a window in the infrared [8, 9]. Thus, for a chip mounted in a flip-chip package, optical means can be used to investigate its transistors. (When observed through the substrate, the transistors are always in view, whereas metal lines may be covered by the transistors or other metal lines.) In **Figures 2(a)–2(c)**, we show the absorption coefficients [9] and the index of refraction of silicon in the near infrared [10]. Because silicon is opaque at wavelengths shorter than about 1 μm, optical imaging of emission through the silicon substrate must be carried out at wavelengths longer than 1 μm. For undoped silicon, at these wavelengths, even a 1-mm-thick specimen is transparent. For both n- and p-type Si, for doping levels below 10^{18} cm^3, there is moderate absorption. However, for Si thicknesses less than 100–200 μm, light at wavelengths longer than 1 μm can be transmitted with reasonable efficiency even in doped Si. While most Si substrates used in the fabrication of ICs are between 0.5 and 1 mm thick, it has been shown that such substrates can be thinned to 50–200 μm without adversely affecting circuit performance. For such wavelengths, diffraction-limited imaging with conventional optics has a resolution limit of about 0.5 μm. The values in Table 1 indicate that optical methods will be strained by the progress of silicon IC technology over the next decade. For the next several generations of chips, however, devices will still be spaced by more than 0.5 μm, so that adequate resolution should be attainable with conventional optics. For the future, unconventional optical imaging methods should permit further extensions of the spatial resolution of optical methods for inspection and test, especially when the large value of the dielectric constant of Si is considered [11, 12].

3. Hot-carrier emission in silicon

Luminescence effects are weak in silicon [13]. Because of its indirect band structure, free electron–hole recombination with the emission of a photon can occur only through phonon-assisted or defect-induced processes. These band-to-band recombination processes are so inefficient in silicon that other luminescence processes, although also producing little light, may become significant. Of special interest in this paper is the phenomenon of hot-carrier emission by intraband transitions. We first define the concept of hot-carrier emission in silicon FETs and then provide experimental examples of this behavior for both isolated devices and CMOS gates.

- *Hot-carrier luminescence in silicon FETs—Concepts*

The usual case of room-temperature optical emission in direct-gap semiconductors such as GaAs involves the direct recombination of a free electron in the conduction band with a free hole in the valence band. The energy generated by this recombination is greater than the bandgap of the semiconductor (i.e., the minimum energy separation between conduction and valence bands). Therefore, the photon energy is greater than or equal to the bandgap energy, and emitted photons can be reabsorbed by the semiconductor. In indirect-bandgap

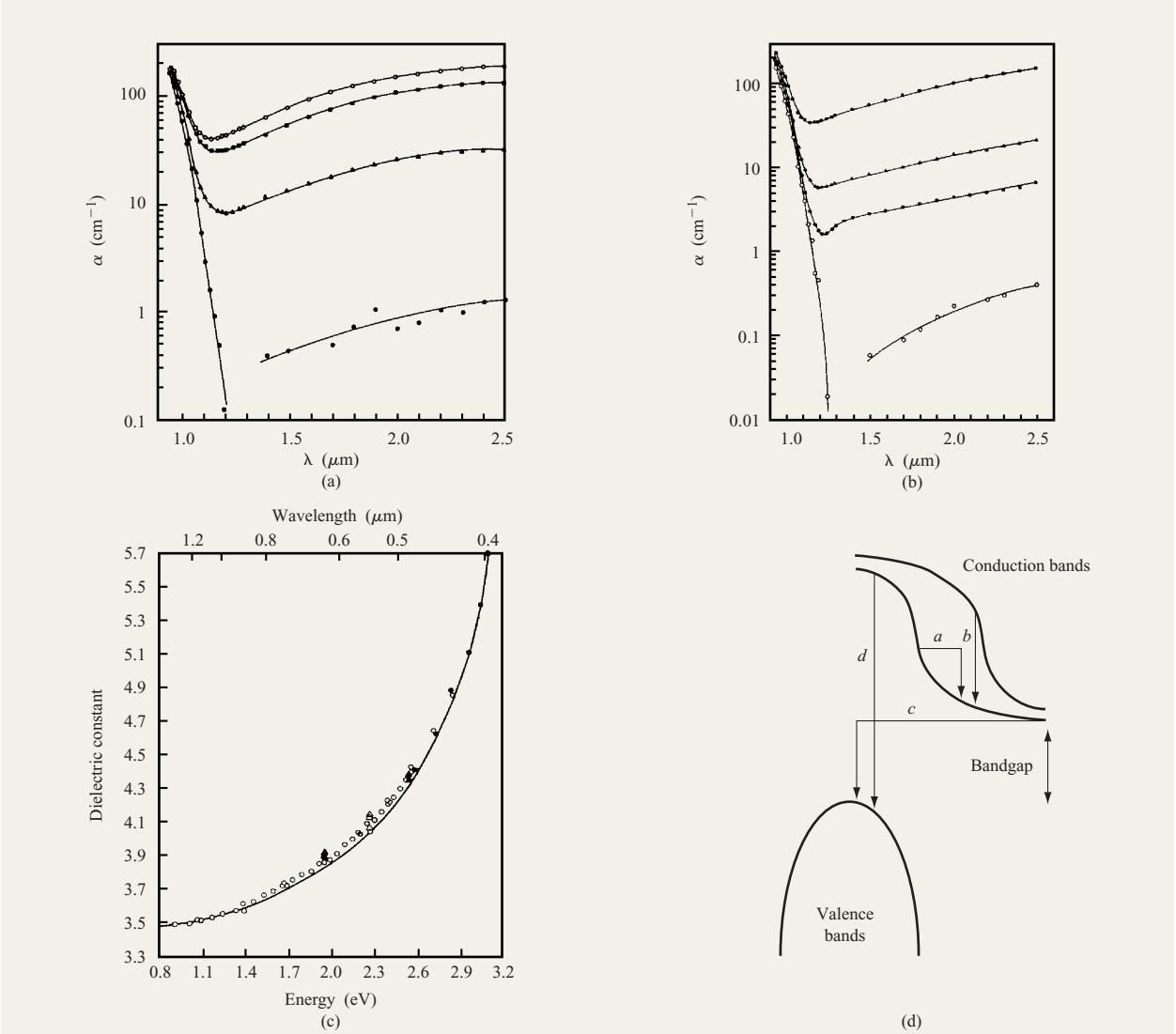


Figure 2

(a, b) Optical absorption coefficient α of n- and p-doped silicon for energies between 0.5 and 1.4 eV. (c) Dielectric constant of Si between 0.8 and 3.2 eV. (d) Band structure of silicon, indicating the energy positions of typical hot electrons. The symbols in (a) denote the following: (●) 6.5×10^{16} cm^{-3}; (▲) 1.6×10^{17} cm^{-3}; (■) 6.4×10^{18} cm^{-3}; (○) 9.2×10^{18} cm^{-3}. The symbols in (b) denote the following: (○) 1.5×10^{16} cm^{-3}; (●) 3.3×10^{17} cm^{-3}; (▲) 1.2×10^{18} cm^{-3}; (■) 7.3×10^{18} cm^{-3}. Parts (a) and (b) from [9], with permission; part (c) from [10], permission requested; part (d) from [16], with permission.

materials such as Si there are similar band-to-band emission processes, except for the participation of phonons or defects to satisfy the requirement for momentum conservation in optical transitions.

The process described as hot-carrier emission by intraband transitions is quite distinct from this free-electron–free-hole recombination. The first step of the process involves the generation of a "hot" carrier [14]. A hot carrier is defined as a free carrier with considerably more kinetic energy than expected from the temperature of the lattice surrounding the carrier. A hot electron, because of its kinetic energy, occupies a conduction-band state well above the conduction-band minimum, as illustrated in **Figure 2(d)** for carriers at the origins of the lines *a*, *b*, and *d*. Similarly, a hot hole occupies a valence-band state well below the valence-band minimum. The

term "hot" here is used by convention, and does not literally mean hot from a thermodynamic perspective.

Hot carriers can be generated by a variety of methods. In any semiconductor, absorption of a photon with energy well above the bandgap energy can generate a hot electron and a hot hole simultaneously. The excess energy of the photon (photon energy minus bandgap energy) is divided between the photogenerated electron and hole, with possibly a small amount of the excess energy going toward creation of a phonon in an indirect-gap semiconductor such as silicon [9]. Another way to create hot carriers, and the method which applies for PICA, involves adding kinetic energy to an already existing free carrier [14]. This energy is provided by accelerating the carrier in an applied electric field. If the electric field is **E**, the kinetic energy is just $e\mathbf{E}d$, where e is the intrinsic charge on an electron or hole, and d is the distance through which the carrier moves under influence of the field. For a carrier starting at zero kinetic energy, the time t required to move the distance d is just $\frac{1}{2}(e\mathbf{E}/m^*)t^2$, where m^* is the effective mass of the carrier. Provided the carrier does not lose this acquired kinetic energy by some other interaction, it should continue to accelerate and gain energy as long as it is subject to the electric field, leading to a kinetic energy at time t of $\frac{1}{2}(e\mathbf{E}t)^2/m^*$. However, if the carrier interacts with or is scattered by the lattice or other carriers, it will typically lose most of this kinetic energy, and so will have to start accelerating over again. Such scattering times are typically a few tenths of a picosecond. In order to acquire about 1 eV of kinetic energy in 0.2 ps, an electron in silicon must be subject to a sizeable electric field, about 10^5 V/cm.

When FETs in practical circuits are "on" (i.e., in a conducting state), many free electrons (or free holes in a p-FET) exist in the channel. Since the scattering process is statistical, at any given time the carriers have a distribution of energies. For the lattice temperatures and carrier densities of interest here, this distribution is the thermodynamic Maxwell–Boltzmann distribution, with the probability that a carrier has kinetic energy E proportional to $\exp(-E/kT)$, where k is the Boltzmann constant and T is the effective temperature of the distribution. With the values used above, a typical effective temperature is 3000 K, hence the use of the word "hot" to describe the carriers.

Theoretical calculations can generate detailed distributions of the carrier energies in FET structures under normal bias conditions. The agreement between different theoretical models is now good enough for some general conclusions to be drawn. Fischetti et al. [14] showed, for example, that the electrons in the near-drain pinch-off region of a deep-submicron silicon n-FET (0.15-μm effective channel length, 0.25-μm physical channel length, V_{DS} = 3.4 V and V_{GS} = 1.0 V), can have average

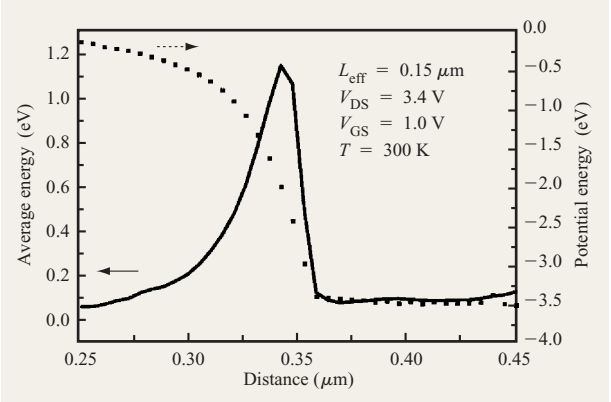


Figure 3

Predicted dependence on position in the channel of the energy of electrons in a 0.15-μm n-FET for V_{DS} = 3.4 V, V_{GS} = 1.0. From [14], with permission.

energies of more than 1 eV. The dependence of the average electron energy on its position in the channel calculated by Fischetti et al. is shown in **Figure 3**. The dotted line is the position of the band edge. Between the source and the pinch-off region of the FET (for $x < 0.3$ μm in Figure 3), almost all of the carriers are within 0.2 eV of the band edge. However, between the end of the pinch-off region, beginning at 0.35 μm in the figure, carriers with energies as high as 1 eV are found in Figure 3. For any FET, the high-field pinch-off region can exist only when the n-FET is in saturation. Within the section of the n-channel away from the pinch-off region, maximum electric fields and average energies can be an order of magnitude smaller than in the pinch-off region. Similarly, when the n-FET is operated in the linear regime, where the electric field in the channel is uniform across its length, the field and average carrier energies are low compared to those found in the pinch-off region of the FET in saturation.

In the qualitative discussion above, the electron distribution was assumed to be described by a thermodynamic state. The detailed calculations of Fischetti et al. [14] allow calculation of the electron distribution in the channel even if the distributions are nonthermal. For electric fields below 3×10^5 V/cm, the electrons are indeed characterized by a thermodynamic distribution, with a Maxwell–Boltzmann distribution for the high-kinetic-energy electrons. At higher fields, near 10^6 V/cm, nonthermalized distributions were obtained. Under extreme high-field conditions ($>10^6$ V/cm), electrons can acquire enough energy from the electric

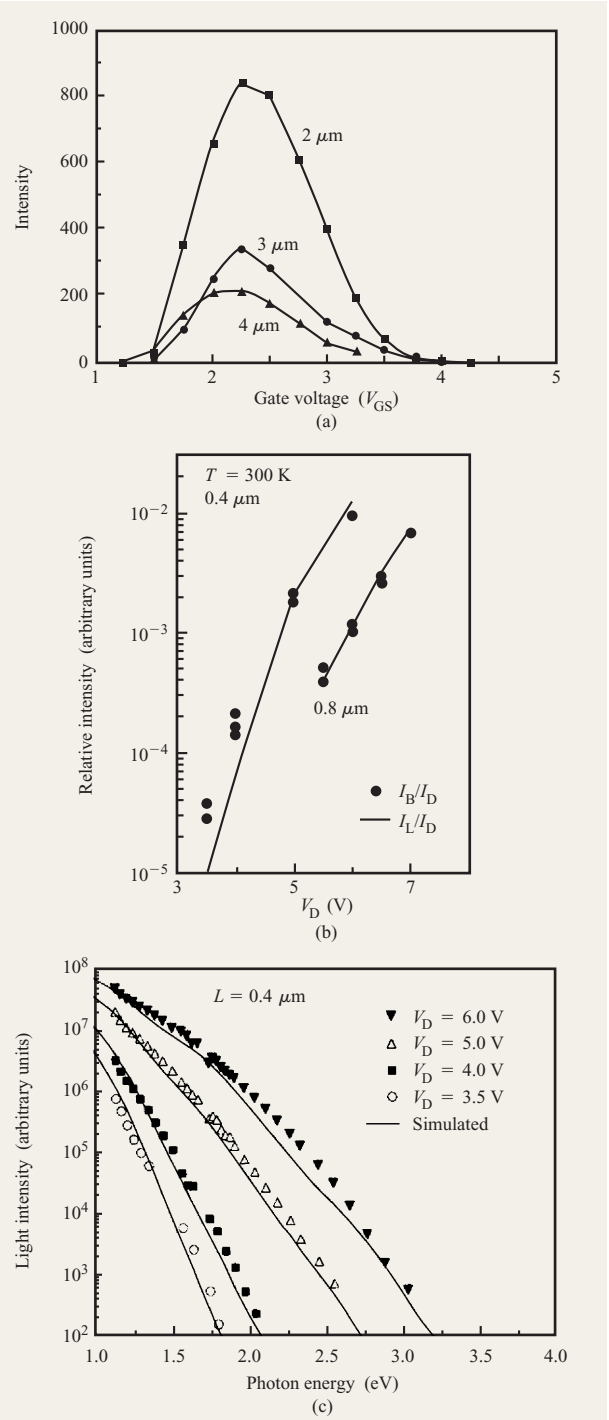


Figure 4

Experimental dependence of the intensity of hot-carrier emission in silicon n-channel devices on (a) gate-to-source voltage V_{GS} (for three different gate lengths and V_{DS} = 5 V); (b) source-to-drain voltage V_{DS} normalized to the drain current I_D; and (c) photon energy. Part (a) with permission from C. F. Hawkins et al., *Proc. ISTFA '90*, p. 55; parts (b) and (c) from [18], with permission.

fields to occupy states over 1 eV above the bottom of the conduction band.

Regardless of whether the hot carriers are characterized by a thermodynamic or nonthermalized distribution, a number of different types of transitions involving hot carriers can occur which result in the emission of light in the infrared and visible regions of the spectrum. For all transitions, a hot carrier ends up in a final state of small kinetic energy, accompanied by emission of a photon which contains most of the kinetic energy lost by the carrier. For most types of transitions, a portion of the lost energy goes to a lattice vibration (phonon or impurity or dopant atom). The quantitative theory of how such hot carriers can emit light in FETs has been considered by Tam and Hu [15], Bude et al. [16], Villa et al. [17], Selmi et al. [18], and others. These calculations require both the derivation of the hot-carrier distribution and the matrix elements for optical transitions involving these carriers. The hot-carrier distributions are generally similar. Significant controversy has existed over the exact mechanisms responsible for the optical emission. Possible mechanisms are 1) direct transitions between the higher and lower conduction bands; 2) intraband processes in which the requirements for conservation of momentum are satisfied by the presence of defects; or 3) intraband processes in which momentum conservation is satisfied by phonon emission or absorption. In all cases, however, the hot-carrier emission is weak, the radiative events being a tiny fraction of the phonon-dominated relaxation processes.

The spectral dependence of the hot-carrier emission is generally described by the same Maxwell–Boltzmann shape as the carrier distribution. Thus, the slope of the emission curve can be used to define an effective carrier temperature. For an n-channel device, the hot carriers can reside either within the lowest-lying X_1 symmetry conduction band or the next-higher-lying conduction band. The carrier–phonon and carrier–carrier scattering rates for these carriers are sufficiently high that the hot-carrier distributions can relax on subpicosecond time scales. The hot-carrier distributions and the associated light emission can respond essentially instantaneously to changes in the electric fields and currents that occur on the picosecond time scales characteristic of modern integrated circuits. The calculated intensity of the emission depends linearly on the current in the device. The hot-carrier distributions will thus reflect the electrical state of a gate in an FET on the picosecond time scale. The hot-carrier emission is therefore an effective means of monitoring electrical behavior in FETs on the picosecond time scale.

The above discussion has not made any distinction between hot electrons in the conduction band and hot holes in the valence band. The magnitude of the electric charge is the same for both, so they experience the same

force as a result of an applied electric field. On the other hand, differences in effective mass, and also shorter scattering times for holes compared to electrons, result in hole mobilities that are typically only half those of electrons. The result is that, for the same electric field, holes will gain on average significantly less energy before scattering than will electrons. Therefore, the temperature of a hot-hole distribution in a p-FET is substantially lower than the electron temperature in a comparable n-FET. This lower temperature means that the optical emission from p-FETs is typically more than an order of magnitude weaker than that from comparable n-FETs [19].

• *Hot-carrier luminescence in silicon FETs—Experimental results*

Under normal operating conditions either in the linear regime or in cutoff when no current is flowing, conventional FETs do not emit detectable amounts of light. However, it has been shown experimentally in many studies that when operating in saturation, both n-FETs and p-FETs emit detectable near-infrared light[2] [15, 19]. The intensity of the light is a strong function of both the source-to-drain voltage V_{DS} and the gate voltage V_{GS}. For a fixed value of V_{DS}, the intensity of the emission shows a peak near $V_{GS} = V_{DS}/2$, as shown in **Figure 4(a)**. If V_{DS} is allowed to change while the density of the carriers in the channel is kept constant, the intensity of the emission increases exponentially with V_{DS}, as shown in **Figure 4(b)**. The spectrum of the emission also demonstrates exponential (Maxwell–Boltzmann) behavior, as shown in **Figure 4(c)**. These experimental results are consistent with the theoretical treatments of hot-carrier emission in silicon FETs discussed above. Most emission measurements have concentrated on n-FETs. As discussed in the previous section, n-FETs are much brighter than p-FETs. Experimental observations indicate that p-FET emission is typically only about 2% of the emission from a comparable n-FET, and can be difficult to detect in many cases. Most of the data to be discussed subsequently will therefore concentrate on the more intense n-FET emission.

• *Switching-induced, pulsed hot-carrier light emission from CMOS logic gates*

In CMOS circuits, n-FETs and p-FETs are combined to perform various types of logical operations [20]. The "C" in CMOS stands for *complementary* in the sense that the conduction channels of the p-FETs are generally in series with the conduction channels of the n-FETs. **Figure 5**

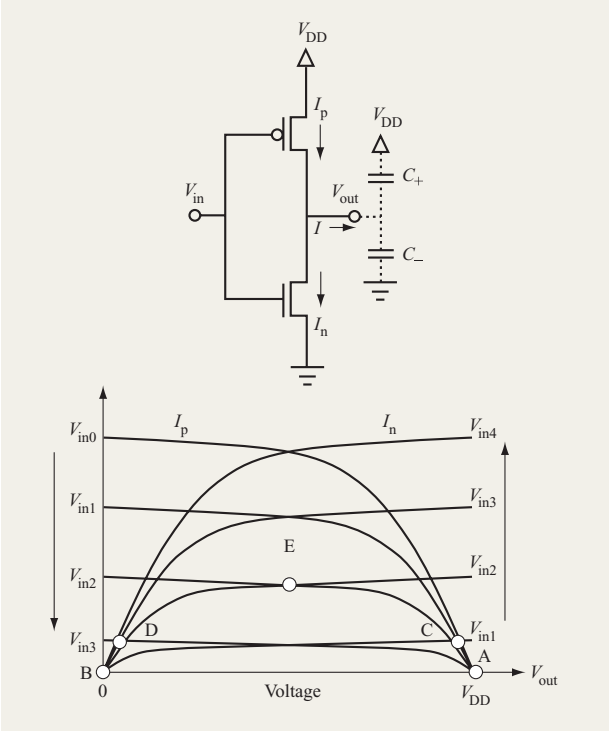


Figure 5

I–V characteristics of a p-FET and an n-FET in a simple inverter geometry showing the values of V_{out} and I_{DD} which occur upon switching. From [20], with permission.

includes the n-FET and p-FET *I–V* characteristics for the simple inverter at the top of the figure. Note that under dc conditions where a p-FET is fully conducting (i.e., gate voltage near zero), the complementary n-FET is switched off. In the reverse situation (gate voltage near V_{DD}), the p-FET is switched off and the n-FET is fully conducting. Because of this complementary character, as seen in Figure 5, under static conditions when the input voltage $V_{in} = 0$ or $V_{in} = V_{DD}$, no current flows in CMOS gates. The absence of current under dc conditions makes the CMOS technology low-power, facilitating the high circuit density of modern computer chips. This absence of current also means that there can be no light emission from CMOS circuits under dc conditions.

On the other hand, when a CMOS gate is changing its logic state, current does flow, leading to the transient emission of light. Our recent studies were the first to use this transient hot-carrier emission to measure intrinsic time-varying behavior in CMOS circuits. We now consider the case of a CMOS inverter which is not driving any other circuits. When the gate or input voltage, V_{in}, of this inverter is swept from ground to V_{DD}, there is a range of

[2] C. F. Hawkins, J. M. Soden, E. I. Cole, and E. S. Snyder, Proceedings of the International Symposium for Test and Failure Analysis (ISTFA '90, USA), 1990, p. 55; D. L. Barton, P. Tangyunyong, J. M. Soden, A. Y. Liang, F. J. Low, A. N. Zaplatin, K. Shivanadan, and G. Donohoe, Proceedings of the International Symposium for Test and Failure Analysis (ISTFA '96, USA), 1996, p. 9.

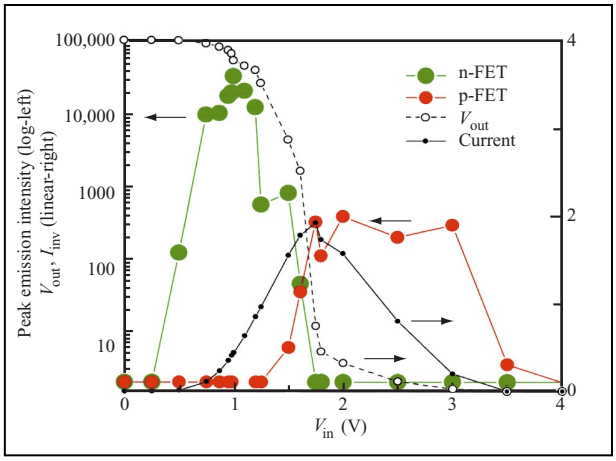


Figure 6

Dependence of V_{out}, I_{DD}, n-FET emission intensity, and p-FET emission intensity on V_{in}.

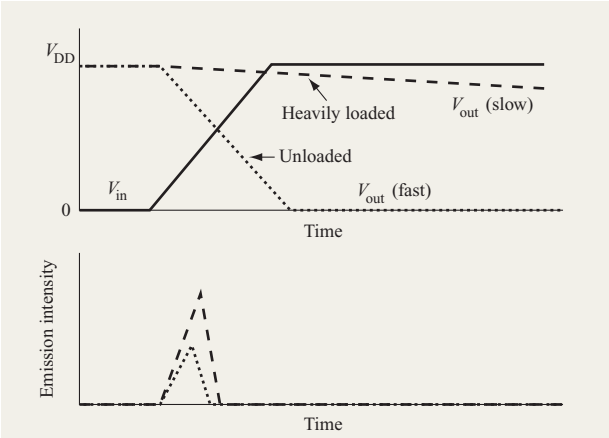


Figure 7

Schematic dependence on time of the intensity of light emission from an n-FET for unloaded and heavily loaded inverters.

values of V_{in} when both the n-FET and p-FET devices are conducting, and a "short circuit" current runs from the power supply at voltage V_{DD} to ground through the conducting channels of the n-FET and the p-FET. Switching of the input voltage V_{in} causes the output voltage V_{out} of the inverter to trace the path shown by the circles (ACEBD) in Figure 5. This figure shows that both the n-FET and the p-FET are briefly in saturation as V_{in} switches from 0 to V_{DD}. In both cases, as V_{in} switches, $V_{in} = V_{GS} = V_{DS}/2$ for a brief period. This gate-voltage condition produces maximum optical emission from the FETs.[3] As a result, optical emission from hot carriers is expected from both the n-FET and the p-FET of an unloaded inverter (although, however, the n-FET emission will be much brighter). In **Figure 6**, we show the V_{in}, V_{out}, I_{DD} characteristics of a simple unloaded inverter, and the intensities of the light emissions from the n-FET and the p-FET as V_{in} is varied from 0 to V_{DD} so that V_{out} goes from V_{DD} to 0. There is a finite current through the circuit only when the n-FET and the p-FET are both conducting. Light emission is observed from the n-FET during the first half of the transition as V_{out} begins to drop and V_{in} approaches $V_{out}/2$. Light emission is observed from the p-FET during the second half of the transition as V_{out} goes from $V_{DD}/2$ to 0, and the p-FET is in saturation. Thus, observation of the emissions from the p-FET and n-FET can be combined to trace both the start and the end of the switching transition. Current and planned CMOS circuits are characterized by gate-to-gate delays of 10–50 ps. The rise and fall times of these gates are twice the gate-to-gate delays. As indicated above, for example, the n-FET emission from a simple inverter for a $V_{out} = V_{DD}$ to 0 transition occurs during the initial rise of the input waveform. As a result, the switching-induced light pulse has a width of less than half the rise time of the gate. Therefore, the width of the emission pulse is comparable to the gate-to-gate delay.

The above discussion was for an inverter which was not driving any downstream logic gates. In a functional chip, of course, a logic gate usually drives one or more downstream circuits. In CMOS circuit chips, the load generated by these downstream circuits is due primarily to the metal lines going to the subsequent logic gates, and the gates of the FETs comprising the downstream logic. Such loads are largely capacitive with a small resistive component. For an inverter driving a capacitive load, the output voltage cannot instantaneously follow changes in the input voltage. Then, for example, V_{in} can reach $V_{DD}/2$ while V_{out} remains at V_{DD}. For a transition where V_{out} changes from V_{DD} to 0, p-FET emission will be weaker than for the unloaded inverter (or may not occur at all) if a voltage drop is developed across the p-FET only after the p-FET is turned off. This situation is represented schematically in **Figure 7**. Also under this situation, the n-FET emission can be significantly stronger than for the unloaded inverter. Current passes through the n-FET even after the p-FET is turned off. This current is not directly provided by the power supply, as for the unloaded inverter. Instead, the current originates from the charge stored on the load. The presence of the load therefore

[3] C. F. Hawkins, J. M. Soden, E. I. Cole, and E. S. Snyder, Proceedings of the International Symposium for Test and Failure Analysis (ISTFA '90, USA), 1990, p. 55.

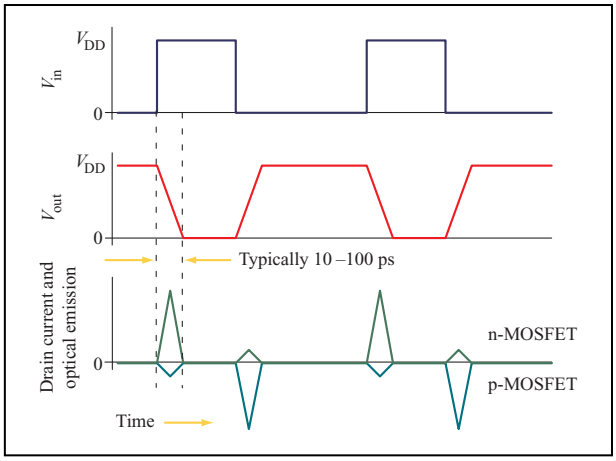


Figure 8

Schematic of the relationship between electrical and hot-carrier-induced optical emission waveforms for a CMOS inverter.

increases the amount of emission from the n-FET during a $V_{out} = V_{DD}$ to 0 transition. The p-FET emission remains solely due to the short-circuit current. For a $V_{out} = 0$ to V_{DD} transition, however, the opposite situation occurs. The n-FET can cut off while there is still a large voltage drop across the p-FET. The large voltage drop arises from pulling the load up to V_{DD}, which requires additional current from the power supply through the p-FET. The n-FET current in this case is just the short-circuit current. Therefore, for a loaded CMOS gate, $V_{out} = V_{DD}$ to 0 transitions will be accompanied by strong n-FET emission, while $V_{out} = 0$ to V_{DD} transitions will be accompanied by strong p-FET emission [1, 2].

In summary, for either loaded or unloaded CMOS logic gates, switching transitions generate optical emission from hot carriers. This is shown schematically for a simple inverter in **Figure 8**. No current flows through the n-FET and p-FET of the inverter except when there are time-varying voltages and currents present. The dominant emission is associated with n-FETs undergoing $V_{out} = V_{DD}$ to 0 transitions. The presence of current flow through the n-FETs and p-FETs is necessary for observable light emission, but does not guarantee it, since the presence of hot carriers in an FET depends on the magnitude of the source-to-drain voltage and the relative value of the gate voltage. In **Figure 9(a)**, we show a static image of an operating ring oscillator superimposed on a photograph of the circuit itself (barely visible). The oscillator consists of an odd number of inverters (47 in this case), which are connected in a ring. The full circuit also includes a divide-by-32 output stage which is located in the lower 60% of Figure 9(a). For an odd number of inverters in a ring, it is not possible for all inverters to have inputs at the opposite logic state of their outputs. One inverter will have the same input and output. The output of that inverter will

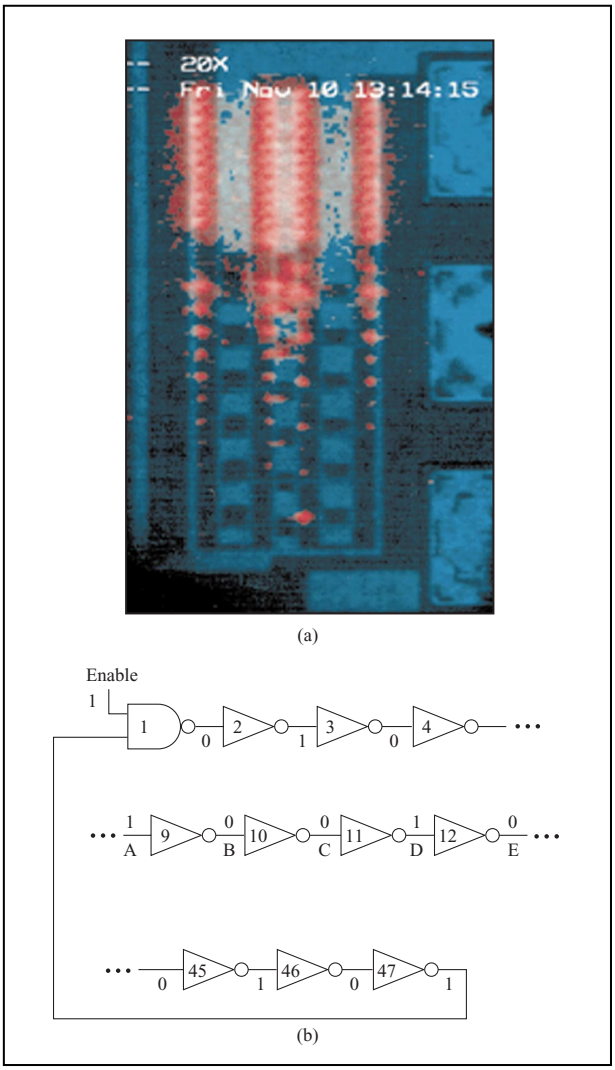


Figure 9

(a) Time-integrated commercial emission microscope image of switching-induced light emission from a CMOS ring oscillator circuit. This ring oscillator consists of 46 inverters and a NAND gate configured as an inverter, fabricated using a 0.6-μm process, and biased at 5 V. The figure shows a photograph of the circuit with an image of the light emission during operation superimposed on the photograph. The ring oscillator portion of the circuit is located in the upper 40% of the figure, where four uniform bands of emission can be seen. The lower 60% of the figure, where four columns of individual spots can be seen, is a divide-by-32 output section which allows the circuit to produce a 3-MHz output rather than a 100-MHz output. (b) Schematic of a 47-gate ring oscillator showing the states of the inverter gates at an arbitrary instant of time.

switch. Then the next inverter will have the same input and output, so it will switch, and so on. The result is the endless sequential switching of the inverters around the ring. The image was obtained with an intensified CCD from a commercial emission microscope and shows light emission from each of the 47 gates of the ring oscillator. This circuit is illustrated schematically in **Figure 9(b)**, which shows the symbols for ten of the 46 inverter gates in three of the four rows of the ring oscillator in Figure 9(a) as well as the enable gate. The ones and zeros next to the inverters capture the state of the inverters at a given instant in time and show how the use of an odd number of gates always results in a single gate having the same level of input and output (0s at gate 10). This is the condition that will propagate through the ring, generating the luminescence seen in Figure 9(a). If the ring oscillator is not operating and no electrical pulse is circulating through it, no switching-induced light emission is observed. We observed emission only from the n-FETs because of their higher mobility, which produced hotter carriers than the p-FETs, and because part of the p-FETs were buried under metal wiring which is opaque to any light that is produced. With a more sensitive detector, weak emission has been observed from the p-FETs.

Although in this paper we restrict the discussion to fully complementary MOS circuits, in the noncomplementary case, such as n-MOS circuits, emission may still be seen. However, the pattern of the emission may be different from the short pulses associated with CMOS switching. For example, with n-MOS circuitry, emission is seen whenever the logic state of the gate is 1 and disappears when the logic state is 0. It should also be possible to observe light emission from bipolar devices. Here, however, the presence of minority carriers means that the dominant emission is expected to be of the band-to-band type.

The above discussions have focused on the optical signals that can occur during the normal operation of CMOS gates, and other silicon devices. Noise signals on the input lines of a CMOS gate can also produce optical emission if they are large enough to switch on the normally open gate of a CMOS pair. Since the conducting partner has only a small voltage drop across it, the normally open complementary gate will be in saturation, and there will be light emission. Such noise-induced light signals can usually be distinguished from the switching-induced n-FET and p-FET signals by their amplitudes and shapes, and by the behavior of the gates both driving the gate of interest and the gates following the gate of interest.

4. The temporal resolution of weak light emission

The image in Figure 9 was obtained using a sensitive camera collecting the light from the sample over a period of several minutes. The image shows that light was emitted, but provides no information about the temporal characteristics of the light. In the previous section, we indicated that theoretical models of light emission in CMOS circuits imply that the light is emitted in the form of short pulses. The pulses should occur only during switching of the gates of the ring oscillator circuit under study, and their duration should be comparable to the gate-to-gate delay time. In the case of the circuit shown in Figure 9, this time is about 100 ps. The measured intensity of the emission in Figure 9 implies that each gate emits a single detected photon only once every several hundred thousand or million switching events, or only a few photons every second. In the next section, we describe how the picosecond time dependence of the intensity of the switching-induced optical emission can be measured with such weak optical signals.

- *Single-photon detectors*

The most sensitive commercially available detectors of light work on the principle of single-photon counting [21]. One important class of single-photon-counting detectors is the photomultiplier. In a photomultiplier, a photon incident on a photocathode has a finite probability, the quantum efficiency, of being absorbed with the emission of an electron, referred to as a photoelectron. The quantum efficiency is a function of both the wavelength of the light and the material used in the photocathode. Acceleration of the emitted photoelectron and resulting collisions within the photomultiplier lead to the creation of a pulse of typically 10^6 electrons, with a duration of typically 1 ns. This amplified pulse, which originates from the absorption of a single photon, can be both detected and analyzed. Electrical noise pulses are also generated in a photomultiplier; thermal generation of electrons by the photocathode produces random "dark" counts at a rate of 10–1000 counts per second for a one-in.-diameter photocathode cooled to −25°C, a typical operating temperature for such a tube. This dark count cannot be distinguished from that due to incident photons, and is the primary source of spurious photon counts in a photomultiplier. Other sources of false counts include electron pulses generated by cosmic rays or by thermal generation of electrons within the multiplying section of the photomultiplier. These pulses are typically larger or smaller than pulses associated with incident photons, and so can be electronically distinguished by simple pulse height discrimination.

- *Time-correlated single-photon counting*

The temporal properties of light detected by photon counting can be analyzed at the picosecond level using a measurement technique known as photon timing or time-correlated photon counting (TCPC). TCPC has been described in many review articles, most recently by Hungerford and Birch [21]. In brief, it relies on the fact

that (as mentioned in the previous section) in a single-photon-counting photomultiplier, each photon-generated electron burst proceeds through the gain stages as a short pulse, typically 1 ns in duration. The time at which any individual pulse arrives at any part of the multiplier is well defined as a fixed delay relative to the time at which the photon arrived at the photocathode. (This delay has a time jitter of 50–100 ps in a well-designed photomultiplier, which determines the ultimate time resolution of TCPC.) Therefore, measuring the time at which this pulse passes a chosen point in the detector can be used to determine when the photon arrived at the photocathode, and ultimately when it was emitted by the light source.

Since typically 10^6 switching events are required to produce one detected photon from an n-FET, the time dependence of optical emission cannot be determined from running a single set of switching events or instruction cycles through the chip circuitry. Multiple passes must be summed to recreate the time dependence. We refer to this recreated time dependence from a single FET as the "optical waveform" of the FET. Since far less than one photon is detected on average from each repetition of the instruction set, this optical waveform is not the actual intensity of emission vs. time from a single pass, but instead measures the relative probability of detecting a photon at any given time within the pass relative to its beginning. So, for each detected photon we need to determine the delay between the start of the instruction cycle constituting a test pattern, and the emission of the photon.

An electrical trigger pulse synchronous with the start of each pass is used to start an electronic stopwatch. This start pulse can be generated either by the chip under test or by the system exciting the chip under test. If a photon is detected by the photomultiplier, the stopwatch is stopped. (If no photon is detected, the stopwatch is simply reset at the end of the instruction cycle.) Typically, a time-to-amplitude converter followed by an A/D converter is used as the digital stopwatch for the time delay. This delay is the time from the start of the pass to the emission of the photon, plus a constant offset fixed by the optical and electronic delays of the apparatus. The delay is passed to a computer which stores it and updates a histogram of the number of detected photons for each possible time. As the histogram is developed, it ultimately (after 10^8 or 10^9 passes) provides a good representation of the optical waveform. The signal-to-noise ratio of the waveform increases as the number of passes is increased. For most cases when photon counting is used, the noise is due to the combination of the dark count of the detector and the statistics for the detection of single photons. The Poisson distribution describes these statistics, as discussed further in the subsection on data storage and signal-to-noise ratio.

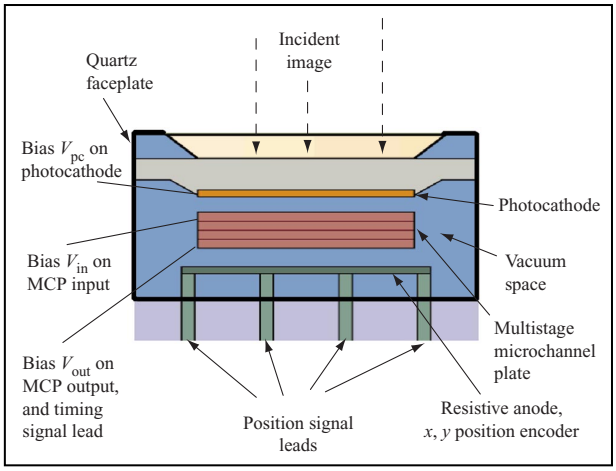


Figure 10

Cross-sectional view of an imaging microchannel plate photomultiplier. From [26], with permission.

Detectors consistent with TCPC include conventional single-channel photomultipliers, silicon avalanche photodiodes, and microchannel plate imaging photomultipliers. The properties, behavior, and capabilities of the first two families in the context of TCPC have been described by Hungerford and Birch [21]. The properties of the third have been described in the context of time-independent imaging by Lampton and Carlson [22] and in the context of time-correlated photon counting by McMullan et al. [23] and Charbonneau et al. [24]. All of the data presented next were obtained using a microchannel plate imaging photomultiplier. We choose this type of detector because it facilitates imaging individual devices and modules on a chip in space as well as time. Therefore, a single measurement of light emission from a working IC using a microchannel plate imaging photomultiplier makes it possible to obtain temporal and spatial information about the behavior of many different devices simultaneously. With such a parallel measurement of time-dependent optical emission from multiple devices, it is possible to make a detailed characterization of an entire circuit with a single measurement, hence the designation picosecond imaging circuit analysis, or PICA [25].

- *Imaging detectors for PICA*

A cross section of an imaging microchannel plate detector, a Mepsicron**, is shown in **Figure 10** [26]. Such a detector has been widely used in atomic physics, condensed-matter physics, and chemistry research [26] but is used here for

the first time to analyze light emission from silicon ICs. The optical emission from the sample is imaged onto the photocathode shown at the top of the figure, which converts an incident photon into a photoelectron. The quantum efficiency of the photocathode is uniform over the image area. Therefore, when integrated over time, the two-dimensional spatial dependence of the photoelectrons emitted from the photocathode replicates the intensity of the optical image on the photocathode. An electric field is applied to the small gap between the photocathode and the first microchannel plate. This field accelerates the photoelectrons emitted from the photocathode into the first microchannel plate and prevents the electrons from drifting laterally in the small gap. When an electron enters a channel of the microchannel plate, it is subject to a large accelerating voltage. Within that channel, it undergoes multiple collisions with the walls, ultimately producing electron multiplication of 10^5 or 10^6. There can be gaps between the microchannel plates in Figure 10 to allow the charge in the microchannel plates to expand slightly, reducing space-charge effects. The electrons provided by the microchannel plates for the multiplication of the original photoelectron arise from the high-voltage supply which biases the photomultiplier. The power supply provides a transient current pulse with the passage of each amplified electron pulse through the microchannel plate. This current pulse can be used as the timing pulse for TCPC. While the electron pulse is within the microchannel plates, it is confined to a few neighboring channels so that when the pulse emerges from the bottom of the microchannel plate and strikes the anode in Figure 10, it is spatially localized to the same location as the incident photon.

To determine this location, a special resistive anode is used. The amplified packet of electrons is captured by the resistive anode at the bottom of Figure 10. The packet spreads to the edges of the resistive anode and is collected by charge-sensitive amplifiers at each of the four corners. Because of its particular shape, and the uniformity of its resistance, the x and y coordinates of the point at which the electron packet was incident on the resistive anode can be calculated from the ratio of the charge collected at each amplifier [22].

The microchannel plate imaging photomultiplier therefore responds to each detected photon with three outputs [23]. One is due to the current pulse from the high-voltage power supply which signals the passage of an electron pulse through one of the microchannels and provides temporal information. In our measurements, we have achieved a time response of 90 ps [full width at half maximum (FWHM) response to an optical pulse of much shorter duration]. The other pair of outputs are used to obtain the spatial coordinates of the incident photon. Typical spatial resolution is about 60 μm with a 25-mm-diameter photocathode. Primarily because of the *RC* time constant of the resistive anode and charge amplifiers, the detector has a "dead" time of several microseconds between detected photons.

Although the instrumental response time of our imaging-time-correlated photon-counting system is of the order of 100 ps, it should be noted that we can temporally resolve delays between signals on a much finer time scale, especially if they occur at different spatial locations. Under these circumstances, the relative timing is inferred from the shifts in time of the observed emission peaks. The accuracy with which these shifts in position can be determined depends on both the widths and the signal-to-noise ratio of the detected emission peaks. For our 90-ps-wide FWHM signals, the signal-to-noise levels are sufficient to be able to identify, by simple visual inspection, 20-ps shifts of the instrumentally broadened peaks.

- *The PICA system*

Figure 11 shows a block diagram of our PICA system for the simultaneous spatial and temporal resolution of optical emission from an operational integrated circuit. The sample being measured is at the lower left. It is observed by a collection lens which images any hot-carrier light emission onto the photocathode of the detector at the upper left. The sample is excited by the circuit tester at the lower right, and either the sample or the tester generates a pulse at a fixed time during each electrical cycle of the circuit being investigated. In the description of TCPC above, this trigger pulse was used to start the timer, and the photon timing pulse was used to stop the timer. In actual practice, given the finite response time of the electronic system, the roles of these pulses are interchanged because there are many more trigger pulses than photon timing pulses. Since the detected photons are used to start the timer and the trigger pulses are used to stop the timer, the initially acquired data measures the time between the point at which an event occurs and the end of the test cycle. Earlier events are thus associated with longer times and later events with shorter times, so that the actual temporal evolution of the system is obtained by inverting the measured data in time. Outputs from the four charge amplifiers at the anode corners are fed to a position analyzer which generates voltage signals proportional to the x and y coordinates of the detected photon on the photocathode. Digital signals representing the output of the time-to-amplitude converter and the coordinates of the detected photon are stored in a three-parameter multichannel analyzer for subsequent display and analysis of optical waveforms and time-resolved images. It should be noted that different imaging systems have differing conventions for the viewing of electrically acquired images. In some cases the data is presented as

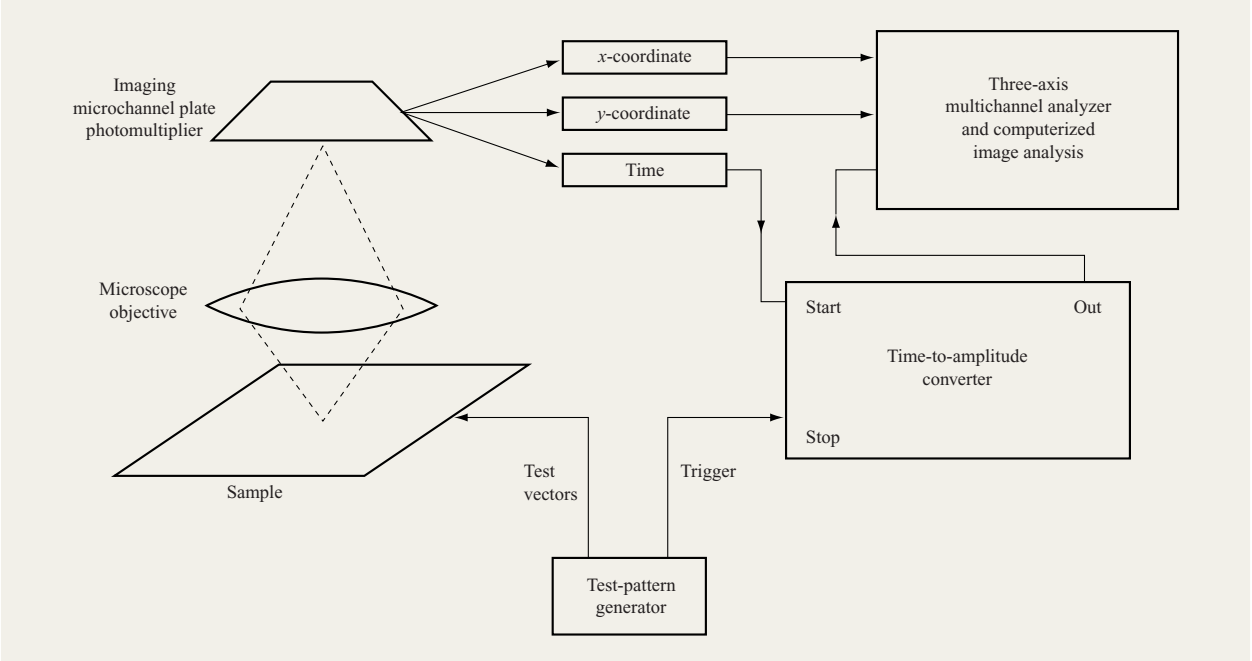


Figure 11

Block diagram of picosecond imaging circuit analysis system.

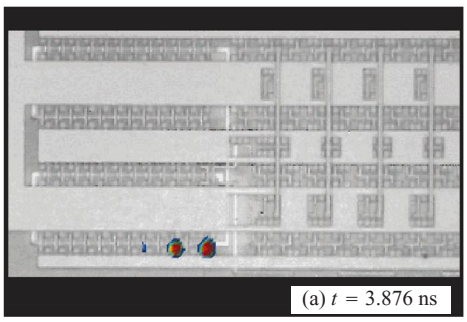
viewed from the face of the photocathode, and others, as viewed from behind its face.

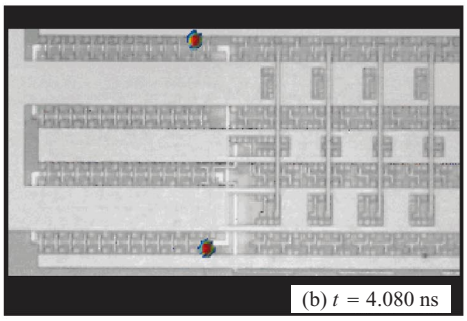
An example of the capabilities of the PICA system, and how it provides information that has been unobtainable with traditional emission microscopes, is shown in **Figure 12**. The PICA system was used to image a portion of an operating ring oscillator identical to the circuit of Figure 9(a). If the parameters defining the arrival times of the photons are ignored, a time-integrated image of the light emission similar to that shown in Figure 9(a) is obtained. If the data describing the time of arrival of the detected photons are used to bin these photons into 34-ps-wide time frames, we obtain sequences of images like those shown in Figure 12. At any particular instant of time, within the approximately 100-ps temporal resolution of the PICA system, only one or two gates in the ring oscillator are emitting light. Because the gate-to-gate delays are about 100 ps and the period of the ring oscillator about 10 ns, the time-integrated measurement in Figure 9 obtained over several minutes sums over many individual switching events. These individual events are temporally resolved by the PICA system. The different emission spots, which are separated by about 20 μm, can easily be resolved spatially by the PICA system. We show in the following sections that this behavior is completely consistent with our expectations for the behavior of the ring oscillator, and our model of light emission by hot electrons generated by switching in the ring oscillator.

The spatial resolution of the above PICA system is a function of the spatial resolution of the Mepsicron [26] detector and of the optical system used to image the test device on the photocathode of the detector. The combination of the 60-μm spatial resolution of the optical detector, an 80–100× microscope objective, and the diffraction limit results in the following: light emission at a wavelength of 1 μm from devices emitting light simultaneously separated by the order of 1 μm on a chip can be spatially resolved for a field of view of 0.25 mm. If a lower magnification is used, a larger field of view can be obtained, and in fact it is possible to image a full microprocessor onto the Mepsicron detector. However, for such "macro" images, it may not be possible to spatially separate emission from neighboring FETs. As described earlier, the temporal resolution of the PICA system is considerably better than the 90-ps full width at half maximum instrumental resolution of the TCPC system.

- *Data storage and signal-to-noise ratio considerations*

The full PICA histogram typically uses 512 x and y values, and 8192 time values, for a total of $(2^9)(2^9)(2^{13}) \simeq 2 \times 10^9$ (x, y, t) elements. Because of the large number of possible bins, the data is usually not stored as a full





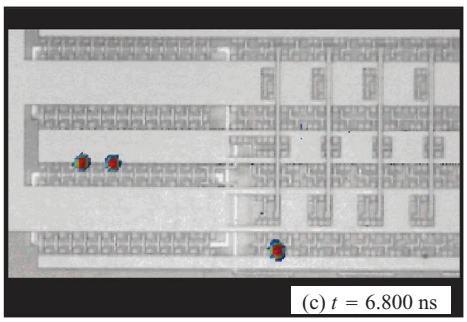


Figure 12

Example of spatial and temporal response of an imaging microchannel plate photomultiplier using a time-correlated photon-counting system to examine the emission from a ring oscillator identical to that of Figure 9. The pulse nature of the emission is clearly seen, as is the ability to spatially resolve light pulses from next-nearest-neighbor gates.

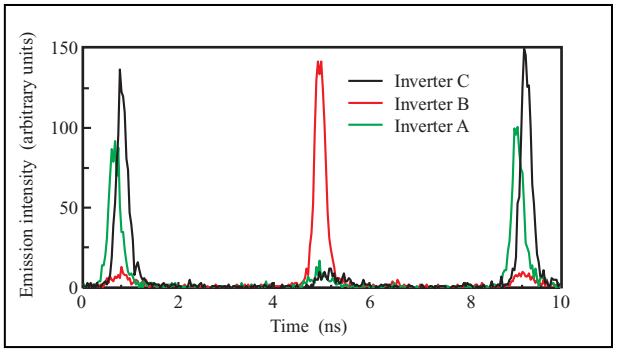


Figure 13

Three optical waveforms of switching-induced light emission from three neighboring inverters of the ring oscillator of Figure 9. Conditions: ring frequency = 236 MHz, or 1/4.24 ns; 47 inverters in ring; delay per stage = 90 ps.

histogram. Instead, other, more efficient storage methods are used. These methods can be exact: A typical PICA dataset contains less than 10^8 photon counts, so most of the (x, y, t) elements in the PICA dataset do not contain any photon counts.

There are two potential sources of noise in PICA measurements. The first arises from the fact that only a finite number of photons are detected in the measurements. Since the photon-counting process follows Poisson statistics, if there are n photons in any (x, y, t) element of a PICA dataset, the average number which would be obtained if the same dataset were taken many times is in the range $n \pm \sqrt{n}$. This statistical noise can be improved as much as desired by collecting more photons, limited only by the patience of the researcher. The other source of noise is that due to the spurious counts created by the dark noise of the detector. This dark noise is typically in the range of 100–1000 dark counts per second, depending on the material from which the photocathode is fabricated and its temperature. But these dark counts are uniformly distributed among the 2^{31} elements of the PICA dataset, resulting in a dark count at each element of 10^{-8} to 10^{-7} photons per second. Since the emission tends to be well localized to the actual devices on the chip, and is further localized to short pulses in time for each device, this very low rate of dark count tends in practice to be negligible. Thus, the statistical noise associated with the detection of the photons is the dominant source of noise for PICA.

5. Picosecond imaging circuit analysis

In this section, we demonstrate the ability of PICA to temporally resolve switching-induced emission from CMOS circuits. Our objectives are to demonstrate how the combined spatial and temporal resolution of PICA allow us to identify switching activity in individual gates of a CMOS circuit and also to describe some of the kinds of information that can be derived from such measurements. We first demonstrate these objectives using the simple ring oscillator of Figure 9(a), viewed from the front side of the chip. In Section 2 we identified the crucial importance of a back-side diagnostic tool. Therefore, we then show examples of back-side measurements on normally operating,

flip-chip-packaged microprocessors, including the identification of actual defects in an early version of a chip.

• *Front-side characterization of ring oscillators using PICA*
In Figure 9, we showed the time-integrated light emission from a ring oscillator [1, 2]. In Figure 12, we showed that each of the 47 bright spots on the left side of Figure 9 is actually a pulse of light emitted during the normal operation of the ring oscillator. In Curve A in **Figure 13**, we show the time dependence of the light emission from a single n-FET in the center of one of the rows in Figure 9, obtained using PICA. The plot of the time-dependent intensity of the light emission was denoted earlier as an "optical waveform." Our nomenclature clearly calls to mind the more common voltage waveforms displayed on oscilloscopes. The optical waveforms from each n-FET of the ring oscillator consist of pulses with a period of less than 10 ns, which coincide with the period of the ring oscillator. Within every period, there is one strong pulse and one weak pulse from each n-FET. These pulses correspond respectively to the $V_{out} = V_{DD}$ to 0 and $V_{out} = 0$ to V_{DD} transitions of the inverter. Two strong pulses from next-nearest-neighbor gates are shown in the figure. The spatial interval between two strong pulses includes a gate producing a weak pulse during the approximately 100-ps time resolution of our PICA system. When inverter A switches from $V_{out} = V_{DD}$ to 0, the next inverter, B, switches from $V_{out} = 0$ to V_{DD}, and vice versa. In Curve B, we show the optical waveform from this nearest-neighbor inverter. As previously described in Section 3, the emission shows the same period but is almost 180° out of phase with that of Curve A. In Curve C of Figure 13, we show the optical waveform from the next-nearest-neighbor inverter, C. It is very nearly in phase with the waveform for Curve A, but is delayed in time by about 200 ps. Because of the imaging nature of PICA, all of these optical waveforms have been measured simultaneously. In **Figure 14**, we show the optical waveforms from a series of next-nearest-neighbor n-FETs for this ring oscillator. In **Table 2**, we tabulate the next-nearest-neighbor gate-to-gate delays derived from the complete set of n-FET optical waveforms. All of the listed delays are about 190 ps except for a group of delays, about 225 ps, involving different rows of the ring oscillator.

The designed gate-to-gate delay in this ring oscillator is half of the delay between the next-nearest-neighbor delays of Table 2, i.e., about 95 ps. This delay is consistent with the behavior we observe for almost all of the gates of the oscillator. The long delays in Table 2 are explained by the effects of loading on the speed of the inverters. These anomalously long next-nearest-neighbor gate-to-gate delays are obtained when the relevant gates are present on different rows of the oscillator. The extra delay is due to the additional length of polysilicon lines connecting the inverters on different rows, which provides a larger load than the relatively short lines needed to connect inverters in the same row [25]. Our ability to detect the subtle effects of wiring runs on the gate-to-gate delays in a working circuit demonstrates the sensitivity and capabilities of PICA.

The presence of alternating weak and strong emission pulses from each inverter n-FET reflects the asymmetric n-FET currents for the $V_{out} = V_{DD}$ to 0 transition compared to the opposite transition, as discussed in Section 3 for a loaded CMOS circuit. The observation of an emission pulse allows a measurement of the exact time at which switching has occurred. The alternation of the emission intensities allows us to determine unambiguously the logic level of the gate before and after the switching occurs.

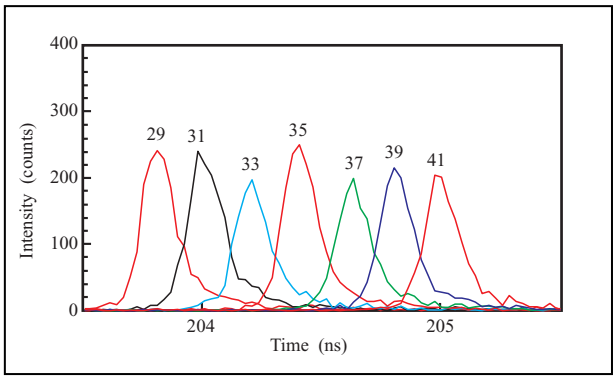


Figure 14

Series of seven next-nearest-neighbor optical emission waveforms from two different rows of the CMOS ring oscillator of Figure 9; inverter number designations are indicated.

Table 2 Next-nearest-neighbor gate-to-gate delays for eleven neighboring gates of the CMOS ring oscillator of Figure 9(a).

Stages	*Delay* (ps)
29 to 31	195
30 to 32	195
31 to 33	195
32 to 34	190
33 to 35	195
34 to 36	215
35 to 37	225
36 to 38	195
37 to 39	190
38 to 40	180
39 to 41	185

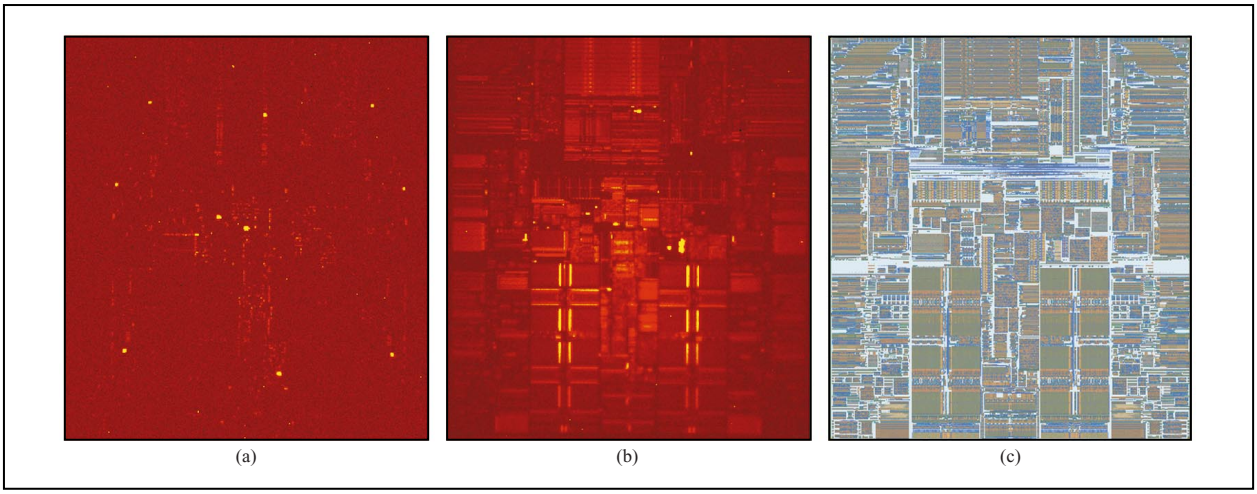


Figure 15

Time-integrated images of switching-induced light emission from a CMOS microprocessor chip under different excitation conditions (a) and (b). Image (c) is a computer rendering of the chip layout.

The intensities of the switching-induced hot-carrier emission from the individual gates of the ring oscillator of Figure 9 show measurable gate-to-gate variations within a given cycle of the ring oscillator, and, for particular gates, significant changes over the 32 cycles of the ring oscillator associated with one full period of the countdown circuit. The former variation arises from small differences in the physical structure of the inverters. The data shown were all obtained from the front side of the circuit; a substantial fraction of the n-FET emission is blocked by the metal layers above the channels. The later variation arises from the fact that there is only a single power-supply line to the circuit. The electrical activity of the countdown circuit loads the power line for the entire circuit, thus producing fluctuations in V_{DD}. The exponential dependence of the intensity of the emission on V_{DD} indicates that very small variations in the power-supply line produce measurable changes in the emission intensity. We have shown [1] that 0.1-V changes in the supply voltage can be easily detected in the intensity of the light emission.

- *Use of PICA for back-side characterization*

In **Figures 15(a)** and **15(b)**, we show two time-integrated CCD images of the switching-induced optical emission from a fully functional, flip-chip packaged, IBM S/390* microprocessor. The microprocessor contains about seven million gates, and was fabricated using a 0.35-μm process. The data were obtained from the back side of the microprocessor chip; the chip was thinned to about 200 μm and its back side was polished. The chip was imaged using a 1024 $\times$ 1024-pixel CCD detector. Hence, the number of pixels in the image was less than the number of active devices in the chip. As a result, even if the optical system responsible for imaging the device under test onto the photocathode had no distortions, it would not have been possible to resolve each individual gate in the chip. For the image in Figure 15, where the imaging was approximately 1:1, and the pixel size was about 25 μm $\times$ 25 μm, light emissions from the chip could not be localized to better than 25 μm $\times$ 25 μm. Such large areas generally contain several light-emitting FETs. The physical chip imaged in Figures 15(a) and 15(b) was 17 mm $\times$ 17 mm in size.

The microprocessor chip of Figure 15 contained a 64 000-latch scan chain. Such "scan chains" are used for characterization and debugging in the implementation of the level-sensitive scan design (LSSD) approach. To produce Figures 15(a) and 15(b), we operated the microprocessor in two different modes. In Figure 15(a), we operated only the clock system of the microprocessor. In Figure 15(b), we operated the scan chain in a clocked mode. In this mode, a word loaded into the scan chain is moved through the scan chain, a single step for every clock cycle. Figures 15(a) and 15(b) were obtained under the same exposure conditions. The images showed strong light emission from the basic clock network of a phase-locked loop, and local clock buffers in Figure 15(a), and the system of scan latches in Figure 15(b). Since the local clock buffers and the scan latches were attached to many

other devices in the chip, both figures also showed emission from other FETs. **Figure 15(c)** shows a layout image of the chip. This image was obtained from the computer-generated masks used to produce the chips so that all structures appeared with equal strength, in contrast to our light-emission images, which showed only electrically active gates. Although the images in Figures 15(a)–15(c) were derived from the same chip, the active gates were not the same for the two different excitation conditions of Figures 15(a) and 15(b). The image on the cover of this issue is closely related to Figure 15(a). In both figures, a false color scheme was used to represent the intensity of the light emission, with yellow representing high intensities and dark red low intensities. The mapping of intensities to color is fairly linear over the range of intensities in Figure 15(a). In the cover image, the mapping was saturated at a lower level to make the weaker emissions easier to see. As a result, the stronger emissions become regions of light rather than small spots.

Figure 16 shows three different time-resolved images of the optical emission obtained with the PICA system, with only the clock system of the chip in operation [as when Figure 15(a) was obtained]. The detector and imaging optics of the PICA system are different from those of the CCD emission microscope used to obtain Figure 15. The time-integrated PICA images generated from the frames shown in Figure 16 are rotated by +90° from the CCD image in Figure 15 owing to differences in the orientations of the cameras, and show only the left two thirds of Figure 15 (or about 12 mm × 12 mm of the chip) owing to differences in the field of view.

Figure 16 shows the operation of the clock system of this chip at three different instants of time; each frame corresponds to an interval of 34 ps. Figure 16(a) shows primarily a single area of light emission, corresponding to the central clock buffer of the chip. Figure 16(b), obtained 542 ps later, shows no activity in the central clock buffer. However, there was switching activity in the local clock repeaters which distribute the clock pulses to the different parts of the chip. Six of the nine repeaters are seen in this image; the additional three repeaters on the right side of Figure 15(a) were out of the field of view of the PICA camera. Figure 16(c) was obtained 798 ps after Figure 16(a), and shows activity in local clock latches throughout the chip.

If the frames of Figure 16 and all of the other time-resolved frames in the dataset used to produce Figure 16 are summed, the result is a time-integrated image of the light emission. The time-integrated image obtained from the PICA system is similar to that shown in Figure 15(a). Because the macro optical imaging system used with our initial prototype PICA camera had less spatial resolution than the optics of our CCD camera, and because the spatial resolution of the Mepsicron is only half that of the CCD, less spatial detail is present in the PICA image compared to that in Figure 15(a). Figures 16(a)–16(c) have been combined with other frames to produce a video of the operation of the clock system of this chip which can be obtained from the authors.

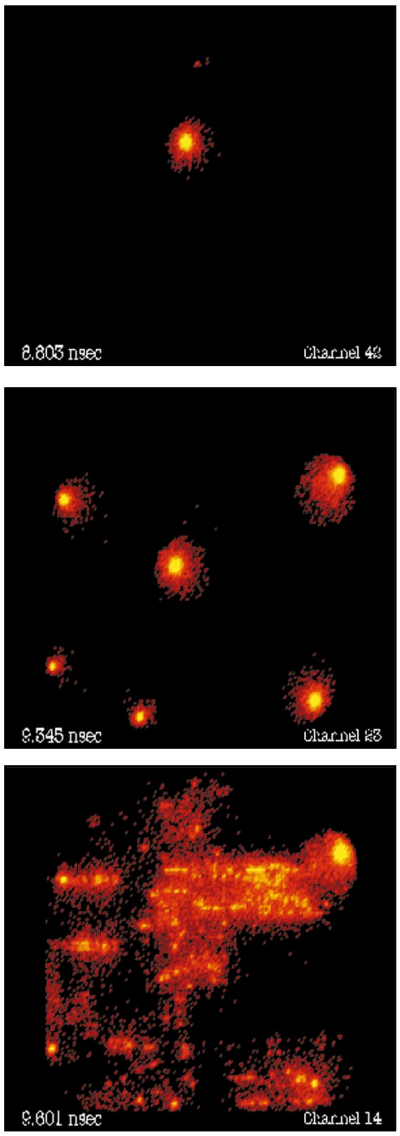


Figure 16

Three time frames of the operation of the clock system for the chip of Figure 15.

The back-side capabilities of PICA shown in Figure 16 have been applied to diagnose a timing issue [27] which

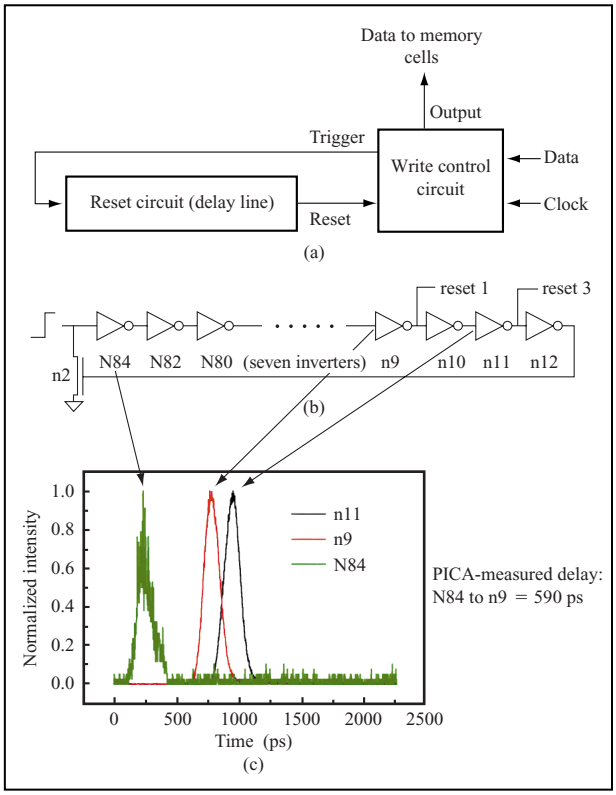


Figure 17

(a) Block diagram of the write control circuit for an L1 cache of a microprocessor including the data and clock inputs, the data outputs to the memory cells, and the reset for this circuit. (b) Schematic of the inverter chain which generates the delay for the reset of the write control circuit in (a). (c) PICA waveforms and timings for the indicated inverters of (b). From [27], with permission.

appeared during the development of an IBM System/390* G6 microprocessor chip. The chip operates at a 637-MHz clock frequency and is implemented in a 0.18-μm process. It was thinned to 60 μm and polished before measuring. Electrical tests showed that the thinning and polishing did not alter the performance of the chip. The normally flip-chip-packaged part was then measured by PICA under normal test conditions. A write control circuit for the L1 cache of the chip showed consecutive-write problems in electrical tests at the highest operating frequencies, though proper behavior at lower operating frequencies. The write-control circuit was reset after each write by the output of a chain of inverters, as shown in **Figure 17(a)** [27]. The chain of inverters is shown in detail in **Figure 17(b)** [27]. This chain of inverters is very similar to the inverter chains used earlier to demonstrate the capabilities of PICA. The back-side PICA measurements on the emission from the chain of inverters revealed that the

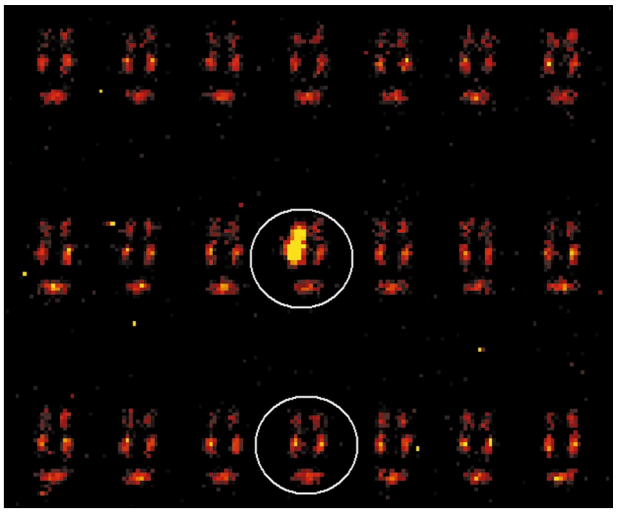


Figure 18

Time-integrated image of light emission from a register file of a microprocessor chip under excitation by a signal producing an error in its register file. From [28], with permission.

measured delay between inverter N84 and inverter N9 was 590 ps. The PICA waveforms are shown in **Figure 17(c)** [27]. The delay was longer than expected, indicating that the reset circuit was trying to reset the write-control circuit at the start of the next memory access cycle, which resulted in a timing collision and errors in the memory cells. The direct observation of this anomalously long delay in the reset of the write control circuit led to a design change which solved the problem [27].

Figure 18 shows a time-integrated image of light emission from a small area of a chip [28]. This chip was similar to that of Figures 15 and 16. In contrast to the original figures, which were obtained using a macro lens where the full chip or a large part of the chip was imaged onto our time-resolved imaging detector, in Figure 18 we used a 20× microscope objective to resolve individual logic gates within a small chip area. The time-integrated emission image of Figure 18 was due to switching in a set of nominally identical latches of a portion of the chip called a register file. Electrical tests of the chip indicated the presence of a fault in this register file. Figure 18 shows that the emission from one single latch pair of the many pairs in the register file is anomalously bright. In **Figure 19**, we compare the optical waveform from a normal latch pair of this register file to that of the anomalously bright latch, which is labeled as the "Faulty latch pair." The extra intensity of the light emission in Figure 18 is due to the fact that there were FETs which were emitting light for a much longer period of time than

normal. The instantaneous intensity of the emission did not appear to be unusual, but its duration was anomalous.

The excess emission occurred only when the B clock was active (high). In addition, the excess emission occurred in all clock cycles, regardless of whether the latch pair was storing a zero or a one. This ability to correlate emission data with circuit timing is a critical element that sets PICA apart from static-emission analysis. Of the suspected defect types and locations, only one could produce these time-resolved emission waveforms: a series resistance between the B clock input and the pull-up of the first inverter in the circuit. The location of this resistance is indicated by the X in **Figure 20**, which is a simplified schematic of a latch pair. A simulated PICA waveform of the circuit containing the defect matches the signature of the measured optical waveform. Physical failure analysis of the part showed a break in the line from the B clock input to the pull-up of the first inverter, which slowed the rate at which the input could vary and produced the observed anomalous PICA waveform.

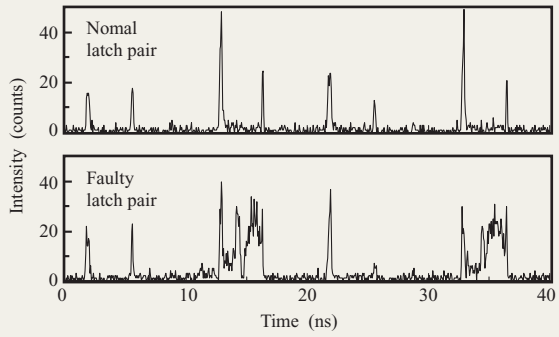


Figure 19

Optical waveforms from normal and faulty latch pairs (circled in Figure 18). The A clock was active from 2 to 5.5 ns and from 22 to 25.5 ns. The B clock was active from 13 to 16.5 ns and from 33 to 36.5 ns. During imaging, the latches were storing a zero in the first half of the waveforms and a one in the second half.

6. Laser probe tools for back-side waveform acquisition

In addition to PICA, the need for back-side-compatible test tools has recently produced a laser probe-based tool for obtaining electrical waveforms from operating flip-chip mounted ICs. The technique is the optical analog of the electron-beam probe technique described earlier. As in the case of PICA, the low absorption by Si of light having wavelengths near 1 μm [9] enables photons at these wavelengths to penetrate hundreds of microns into the Si. Using the technique, a laser beam is focused on a working chip, and the reflected light focused onto a detector. Electrical activity in the chip can produce modulation of the reflected light beam. The modulation is measured, and the scale of the device activity is inferred from the magnitude and phase of the modulation. The optical response of Si can be modulated by a variety of physical mechanisms, including applied voltages, heating, strain, and injected carrier density [29]. The modulated signal is proportional to one of the derivatives of the dielectric constant of Si, the particular derivative being dependent on the modulating mechanism. This technique was originally applied by Heinrich et al. [30] to bipolar transistors and later extended to a variety of devices and environments [31]. Paniccia et al. have shown that it can also be applied to packaged CMOS circuits [32]. Successful use of the laser-based voltage measurement technique requires the ability to match the optical probe size to the device dimensions; extract modulated signals five to six orders of magnitudes weaker than the probe signals; control the loading on a working circuit by carriers generated by the absorption of the probe beam; proper identification of the physical mechanism responsible for the observed modulated signal; and efficient synchronization of the operation of a pulsed laser to the clock of the device under test.

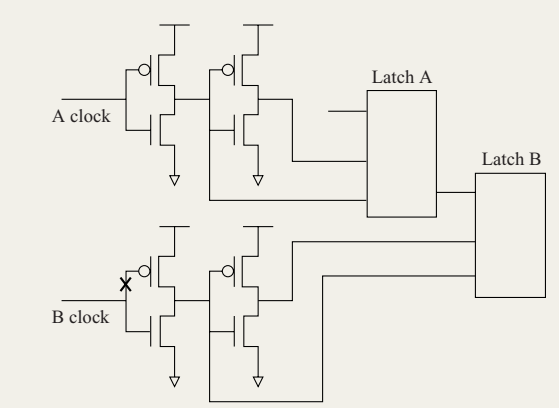


Figure 20

Schematic of a latch pair in the faulty register file of Figure 18. The location of the defective metal interconnection is denoted by an X.

Conclusions

We have covered the use of picosecond imaging circuit analysis for investigating the switching activity in fully functional silicon CMOS integrated circuits. The technique is compatible with modern circuit wiring methods and packaging practices. In particular, PICA can be used

directly on flip-chip-packaged circuits. It can spatially resolve the individual transistors in a CMOS circuit and provide a means to measure gate-to-gate delays with an accuracy of better than 20 ps. Although the switching-induced hot-carrier emission is extremely weak, the high speeds achieved by modern integrated circuits and the sensitivity of existing commercially available detectors and electronic measurement systems allow many useful measurements to be performed in reasonable periods of time. The success of the technique reflects the high level of performance now achieved by silicon integrated circuits, our firm understanding of the physics of electrons and holes in the submicron transistors in those circuits, and the impressive achievements of the physics, chemistry, and astronomy communities in recent years in developing detectors and electronic methods to detect and temporally resolve the weak optical signals found in phenomena relevant to laser fluorescence, Raman scattering, atomic trapping, and observational astronomy. The applications of the technique described in this paper make use of the capabilities of modern desktop workstations to readily store files containing the coordinates of 20–100 million photons, and, using commercially available software, to process these files into videos and groups of waveforms on the time scale of minutes.

The existing PICA system shows considerable power for chip and device analysis, even though the optics, detectors, and other equipment were developed in other fields of study. It is anticipated that when the equipment and methodologies are customized for PICA applications, substantial future improvements should be attainable. For example, the relatively low quantum efficiencies of the detectors used in current PICA systems can require integration times that can in the worst case approach hundreds of thousands of seconds, or as long as 10 to 20 hours. This is the case particularly when the repetition rates for the test patterns used to exercise a chip can be of the order of microseconds or longer. Use of higher-quantum-efficiency photocathodes in imaging photomultipliers should reduce integration times by orders of magnitude. This should greatly enhance the range of problems that PICA can be used to address. Improvements in the temporal resolution of PICA systems should allow detailed measurements of the shape of the optical emission waveforms, which should make it possible to use PICA to obtain additional quantitative information about the electrical waveforms present in a circuit.

Finally, the volume of PICA data poses interesting challenges for its analysis, especially in the context of the identification of failures associated with very long electrical tests involving tens of thousands of machine cycles. In such a case, hundreds of millions of switching events would have to be detected, generating billions of incident photons and creating data files of much more than a gigabit in size, all for a diagnostic problem that might involve the identification of a single event displaced in time by 100 ps. New methods of analyzing large datasets will be needed in order to address such "needle in a haystack" challenges.

Acknowledgment

This work would not have been possible without the cooperation and interest of many colleagues at the IBM Research and Microelectronics Divisions. These include J. C. McGroddy, S. E. Laux, M. Fischetti, D. Knebel, T. Jackman, B. Pezeshki, H. Heinrich, G. Northrop, M. McManus, R. Oldrey, J. Sylvestri, R. Rizzolo, G. Nuttall, R. Evans, P. Phoenix, P. Nigh, and H. Casals.

References

1. J. C. Tsang and J. A. Kash, *Appl. Phys. Lett.* **70,** 889 (1997).
2. J. A. Kash and J. C. Tsang, *IEEE Electron Device Lett.* **18,** 330 (1997).
3. Robert R. Schaller, *IEEE Spectrum*, p. 53 (June 1997).
4. C. W. Koburger III, W. F. Clark, J. W. Adkisson, E. Adler, P. E. Bakeman, A. S. Bergendahl, A. B. Botula, W. Chang, B. Davari, J. H. Givens, H. H. Hansen, S. J. Holmes, D. V. Horak, C. H. Lam, J. B. Lasky, S. E. Luce, R. W. Mann, G. L. Miles, J. S. Nakos, E. J. Nowak, G. Shahidi, Y. Taur, F. R. White, and M. R. Wordeman, *IBM J. Res. Develop.* **39,** 215 (1995).
5. David P. Vallett and Jerry M. Soden, *IEEE Spectrum*, p. 39 (October 1997).
6. J. L. T. Thong, *Electron Beam Testing Technology*, Plenum Press, New York, 1993.
7. "National Technology Roadmap for Semiconductor Technology—1997 Edition," Semiconductor Industry Association, 1997.
8. P. E. Schmid, *Phys. Rev. B* **23,** 5531 (1981).
9. S. E. Aw, H. S. Tan, and C. K. Ong, *J. Phys: Condens. Matter* **3,** 8213 (1991).
10. R. Hulthen, *Phys. Scripta* **12,** 342 (1975).
11. S. M. Mansfield and G. S. Kino, *Appl. Phys. Lett.* **57,** 2615 (1990).
12. Dieter W. Pohl, *IBM J. Res. Develop.* **30,** 417 (1986).
13. J. I. Pankove, *Electroluminescence*, Berlin, Springer-Verlag, 1977, p. 31.
14. M. V. Fischetti, S. E. Laux, and E. Crabbé, *J. Appl. Phys.* **78,** 1058 (1995).
15. S. Tam and C. Hu, *IEEE Trans. Electron Devices* **ED-31,** 1264 (1984).
16. J. Bude, N. Sano, and A. Yoshii, *Phys. Rev. B* **45,** 5848 (1992).
17. S. Villa, A. L. Lacaita, and A. Pacelli, *Phys. Rev. B* **52,** 10993 (1995).
18. L. Selmi, M. Mastrapasqua, D. M. Boulin, J. D. Bude, M. Pavesi, E. Sangiorgi, and M. R. Pinto, *IEEE Trans. Electron Devices* **45,** 802 (1998).
19. L. Selmi, *Microelectron. Eng.* **28,** 249 (1995).
20. Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge University Press, Cambridge, UK, 1998.

21. G. Hungerford and D. J. S. Birch, *Meas. Sci. Technol.* **7,** 121 (1996).
22. M. Lampton and C. W. Carlson, *Rev. Sci. Instrum.* **50,** 1093 (1979).
23. W. G. McMullan, S. Charbonneau, and M. L. W. Thewalt, *Rev. Sci. Instrum.* **58,** 1626 (1987).
24. S. Charbonneau, L. B. Allard, J. F. Yound, G. Dyck, and B. J. Kyle, *Rev. Sci. Instrum.* **63,** 5315 (1992).
25. J. A. Kash and J. C. Tsang, *Phys. Stat. Solidi (B)* **204,** 507 (1997).
26. M. Mellon, *Laser Focus World*, p. 521 (March 1997).
27. W. Huott, M. McManus, D. Knebel, S. Steen, D. Manzer, P. Sanda, S. Wilson, Y. Chan, A. Pelella, and S. Polonsky, *Proceedings of the International Test Conference*, Atlantic City, NJ, 1999, p. 883.
28. D. Knebel, P. Sanda, M. McManus, J. A. Kash, J. C. Tsang, D. Vallett, L. Huisman, P. Nigh, R. Rizzolo, P. Song, and F. Motika, *Proceedings of the International Test Conference*, Washington, DC, 1998, p. 733.
29. M. Cardona, *Modulation Spectroscopy*, Academic Press, Inc., New York, 1969.
30. H. K. Heinrich, B. R. Hemenway, K. A. McGroddy, and D. M. Bloom, *Electron. Lett.* **22,** 650 (1986).
31. H. K. Heinrich, *IBM J. Res. Develop.* **34,** 162 (1990).
32. M. Paniccia, T. Eiles, V. R. M. Rao, and W. M. Yee, *Proceedings of the International Test Conference*, Washington, DC, 1998, p. 740.

Received May 24, 1999; accepted for publication October 28, 1999

James C. Tsang *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (jtsang@us.ibm.com).* Dr. Tsang is a Research Staff Member in the Physical Sciences Department at the Thomas J. Watson Research Center. His B.S. and M.S. degrees are from MIT, and he received a Ph.D. degree in electrical engineering from MIT in 1973. He subsequently joined the IBM Research Division, where his research has involved the optical properties of solids, including Raman scattering, and time-resolved optical studies of materials and devices. Dr. Tsang has received three IBM Outstanding Achievement Awards, has written 134 papers, and holds several patents. He is a Fellow of the American Physical Society, a Fellow of the American Association for the Advancement of Science, and a Senior Member of the IEEE. In 1993–1994, he was an AAAS–Alfred P. Sloan Foundation Executive Branch Science and Engineering Fellow at the White House Office of Science and Technology Policy. He has served on the APS Panel on Public Affairs and its Forum on Industrial and Applied Physics.

Jeffrey A. Kash *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (jeffkash@us.ibm.com).* Dr. Kash is a Research Staff Member in the Physical Sciences Department at the Thomas J. Watson Research Center. He received undergraduate degrees, and a Ph.D. in physics in 1981, from the University of California at Berkeley. He then joined the IBM Research Division, where he has worked on problems involving ultrafast optical processes and hot electrons in semiconductors, with recent emphasis on optical emission from silicon devices. Dr. Kash has received several IBM Outstanding Innovation Awards, holds nine patents, and has authored more than 90 publications. He is a Fellow of the American Physical Society and a Senior Member of the IEEE. He is currently the Technical Assistant to the IBM Research Vice President for Systems, Science, and Technology.

David P. Vallett *IBM Microelectronics Division, Burlington facility, Essex Junction, Vermont 05452 (dvallett@us.ibm.com).* Mr. Vallett is an Advisory Engineer in the Analytical Services Laboratory at the Burlington, Vermont, facility of the IBM Microelectronics Division. After receiving a B.S. degree in electrical engineering from the State University of New York at Buffalo in 1982, he joined IBM as a product characterization engineer. He has held both engineering and management positions in failure analysis and is currently responsible for strategic tool and technique development. Mr. Vallett has received an IBM Outstanding Technical Achievement Award and holds three patents. He has authored several conference papers and journal articles, receiving three best paper awards, and has given graduate lectures in integrated circuit failure analysis. He is a member of the National Engineering Honor Society and the IEEE, and a past chair of the SEMATECH Product Analysis Forum.