Can Silicon-Based Heterodevices Compete with CMOS for System Solutions?

E. Kasper and G. Reitemann
Institut für Halbleitertechnik, Universität Stuttgart, Pfaffenwaldring 47, D-70569 Stuttgart, Germany

1. Introduction

In a rapidly developing field like microelectronics, both researchers at the cutting edge and manufacturers who develop commercial products feel the need for general rules that help to extrapolate from the past into the future. The most prominent rule is known as Moore's Law, which states that on a semilogarithmic scale the critical device dimensions shrink and the circuit complexity increases linearly with time. Complexity in this context is represented by the number of transistors in certain integrated circuits (ICs), like memories or microprocessors. It should be noted, however, that initial exponential growth is not specific to microelectronics alone, but applies to many emerging technologies. Still, one marvels at the steep slope and long duration (over 30 years) of the exponential growth, which has been accompanied by continuous price reduction per transistor in ICs. Eventually, rapid evolution of IC complexity will be enhanced by integrating different components, leading to systems on a chip.

System integration requires rethinking of the integration scheme beyond the current model of component integration, in which CMOS technology provides highly perfect logic circuits while input/output circuits lag behind. In this article we will consider systems with components that differ significantly either in their electrical functions (e.g. analog/digital) or in their nonelectrical functions (e.g. mechanical/optical) or in their power/voltage level (power/logic) or in their technology (e.g. bipolar/CMOS).

Key issues for system solutions concern the high system complexity as well as economics. Will monolithic integration be the ultimate choice? Will CMOS prove an economically viable technology in areas other than digital computation? Will novel device structures allow unified solutions for different functions?

The paper is organized as follows. First we examine the advantages and drawbacks of different paths to system solutions. Then we explore the idea of a common device structure for different functions, using the particular example of a novel device that combines HBTs with charge injection transistors (CHINTs) . Finally, we investigate the realisation possibilities of such devices using Si/SiGe heterostructures.

2. Different paths to system solutions

Today, when you look at a typical system you will find many ICs, discrete devices, and passive elements mounted and interconnected on a board. We will not discuss this conventional solution, except to note that packing density, interconnects, and mounting techniques are improving continuously. Integration can be achieved in hybrid multichip modules (MCMs) or monolithically.

3. Multi-chip modules

Different materials (laminates, ceramics, silicon) can be used as substrates, where the chips are mounted and interconnected. Very elegant solutions were obtained with silicon as substrate, since the thermal expansions of chip and substrate are the same and lithography and metallization for interconnects follows the same principles as in chip fabrication. A wide variety of different chips have already been integrated. As an example, Fig. 1 shows the integration of an active antenna for a mm-wave receiver (96 GHz) with a low-frequency CMOS circuit.[1] The multi-chip module technique has already obtained a good market acceptance and will be a serious candidate for system integration. Advantages are the separate optimization of chips and the sourcing of chips from different manufacturers; drawbacks are limited availability and testing of unpackaged chips, the interconnect length, and the additional mounting effort.

4. Monolithic integration

Monolithic integration of systems is probably restricted to silicon substrates, because of the importance of high complexity logic only available in silicon. A possible list of technical solutions includes:

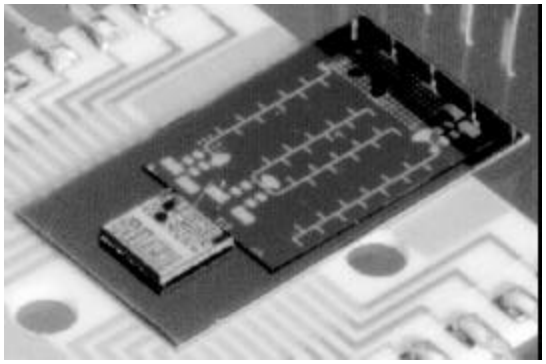

Figure 1. Multichip module with an active mm-wave rectifying antenna (rectenna) and a CMOS amplifier mounted on a high resistivity silicon substrate.[1]

- Chips entirely made in CMOS
- Bipolar and CMOS (BiCMOS)
- Subchip attachment techniques
- Heterointegration

Logic circuits are dominated by CMOS transistors, because in this technology the basic unit – the inverter – consumes very small power at static operation. In a simple theory, assuming field-independent mobility and operation in the saturation regime, the current I_D is given by

$$I_D = b(V_{GS} - V_{TH})^2 \qquad (1)$$

where $b = (\mu W \varepsilon_{ox})/(2Ld_{ox})$, and μ is the mobility, W is the device width, L is the gate length, ε_{ox} and d_{ox} are the oxide permittivity and thickness, V_{GS} and V_{TH} are the gate-source and threshold voltages. For the cut-off frequency f_T one then finds

$$2\eth f_T = \mu(V_{GS} - V_{TH})/L^2 = g_m/C_{ox} \qquad (2)$$

where g_m is the transconductance and C_{ox} the gate oxide capacitance. The gate length shrinkage leads to smaller footprint, higher transconductance, and higher frequency limits. Today, production of microprocessors utilizes CMOS with 0.25 µm gate length.[2] Projected downscaling of CMOS devices with the corresponding technical specifications and performance is described in roadmaps.[3] The future L = 0.1 µm CMOS transistor will be an impressive device with low voltage and respectable speed. Also, MOS transistors with high voltage and power handling capability can be built, but on different substrates and with different technology, making system integration entirely within CMOS a none-too-easy task.

As a result, a combination of bipolar and CMOS technology (BiCMOS) was suggested, realized and produced.[4] Bipolar transistors can deliver large drive currents, operate with small logic swings, and have high noise immunity. A schematic diagram of a BiCMOS device is shown in Fig. 2.

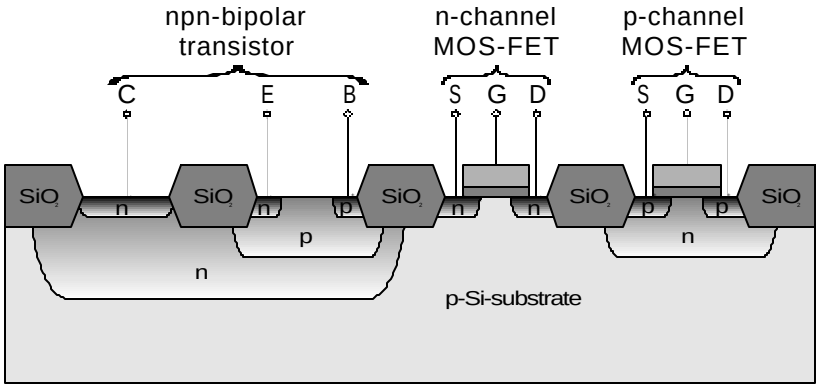

Figure 2. Schematic diagram of a BiCMOS circuit.

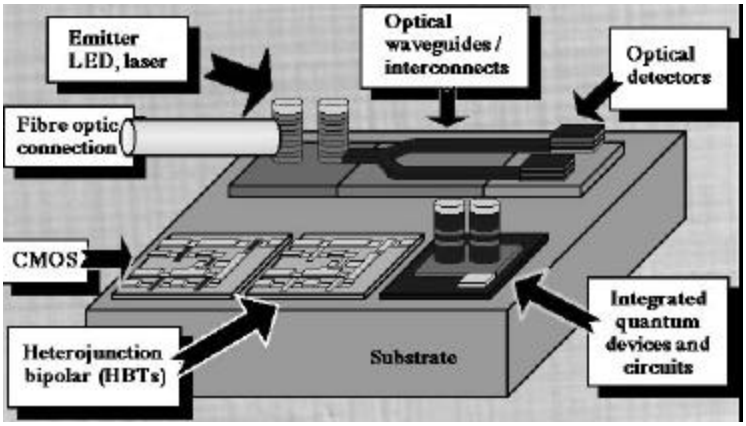

Figure 3. Sketch of a potential future chip with different integrated components.[5]

BiCMOS offers the benefits of both bipolar and CMOS circuits by appropriately trading off the characteristics of each technology, but at the cost of added process complexity. If the system were to integrate different semiconductor technologies in addition to silicon, the process complexity would increase further. A sketch of a future chip[5] integrating many different devices and circuits would be fascinating indeed — see Fig. 3.

However, the limited success of BiCMOS should be a serious warning that adding process complexity will not easily be accepted. The devices from different materials could be added by subchip attachment techniques or directly fabricated by epitaxial heterostructure growth on top of the silicon device level (heterointegration). A possible scheme for heterointegration would consist of three main steps.[6] In the first step, the silicon devices are fabricated without the metallization level and areas for the later heterodevices are defined, e.g. by oxide windows. In the second step, the heterostructures are grown within the windows and the heterodevices are fabricated with a small thermal budget process. In the final step, the common metallization and passivation is formed.

From the viewpoint of process complexity – and BiCMOS is teaching us how important this topic is – a common structure for different devices would be a highly desirable target. Obviously this target cannot be realized with a single material system, however let us think about possible solutions with heterostructures.

5. The search for a common device structure

Heterostructures offer numerous possibilities for enhancing device performance and progress is ongoing in many areas. The idea behind this section is to point to the unique potential of heterostructures for a unified system technology. This direction is a new one for heterostructure research, with system performance and

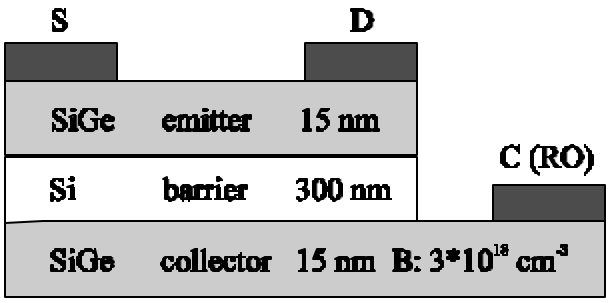

Figure 4. Diagram of a *p*-type SiGe charge injection transistor (CHINT).[7]

low process complexity as the primary goals. The idea is explained with a specific example, but we hope to stimulate other suggestions.

Let us combine as a primary building block a field effect transistor or a real space transfer device and a bipolar transistor. We have chosen a charge injection transistor (CHINT) as real space transfer device and an HBT for the bipolar side. A schematic diagram of a CHINT[7] is presented in Fig. 4.

On top of a Si substrate there is essentially a three layer structure SiGe/Si/SiGe called collector/barrier/emitter. The source/drain contacts are on the top emitter, the collector contact is on the bottom collector layer. By real space transfer hot carriers accelerated by the drain-source voltage cross the barrier to the collector when an appropriate voltage is applied. In the following we denote the CHINT collector contact with RO (real space transfer output) to avoid confusion with the bipolar collector contact. In a common-source configuration of the CHINT the usual input is the drain and RO is the output, therefore the drain in the CHINT is more comparable to the gate electrode in a field effect transistor. A variety of logic functions can be realised with CHINT transistors.[7] The HBT layer structure on the silicon substrate[8] consists also of a three layer structure, however in the order Si/SiGe/Si. The proposed common layer structure is shown in Fig. 5. It consists of a four layer *n*-Si/p^+-SiGe/Si/SiGe structure on a Si substrate with an n^+ buried layer beneath the bipolar area. The p^+-SiGe layer is the base of the HBT and the collector of the CHINT. As shown in Fig. 5, the transistor areas can be defined by trench etching and the base (B), emitter (E) and RO contacts are performed by implantation or diffusion from a polysilicon source, while the HBT collector is contacted via a buried layer subcollector (C).

6. Silicon based heterostructures

Amongst several silicon-based heterostructures (e.g. silicides NiSi and $CoSi_2$, insulators like CaF_2, semiconductors like GaP or SiC), the silicon-germanium on

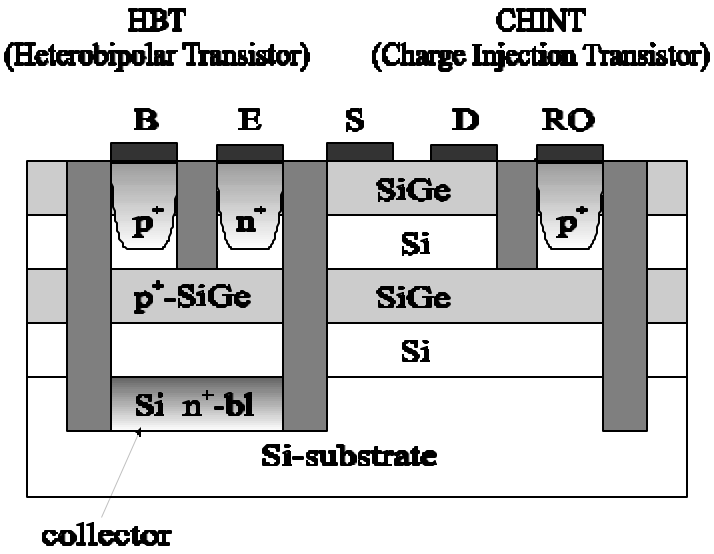

Figure 5. Common layer structure for an integrated HBT and CHINT on a Si substrate. HBT contacts are on the emitter (E), base (B) and on the collector outside of the plane via the buried layer subcollector. CHINT contacts are on the source (S), drain (D) and on the CHINT collector (RO). This contact is denoted RO to distinguish it clearly from the HBT collector contact.

silicon (SiGe/Si) turned out to be most practical because of its chemical similarity, its complete miscibility and its moderate lattice mismatch.[9] As a result, SiGe/Si heterostructures are compatible with Si technology. These heterostructures act as carrier filters in single-barrier structures, exploit quantum effects in double-barrier structures, extend the optical sensitivity into the infrared, and promise fabrication of sub-100 nm structures by self-ordering. The obstacles to widespread technological adoption are the relaxation of layers above a critical thickness h_c and reduced thermal budget processing. Layers below the equilibrium critical thickness are elastically strained and stable under all process conditions. With growth at reduced temperatures (e.g. 550 °C) and processing at reduced temperatures (e.g. 750–850 °C for Si capped structures), significantly higher critical thickness can be maintained (metastable regime). For instance, for a 25% Ge content alloy on Si the equilibrium critical thickness amounts to 6 nm, whereas metastable growth at 550 °C is possible up to 70 nm.[9]

Recent device research into SiGe has focused on SiGe HBTs for mobile communication applications. High frequency limits (f_T , f_{max} > 100 GHz) and low noise levels were obtained by making use of the high current drive and very narrow highly-doped base layers possible in HBTs. The next attractive device target would be a hetero-CMOS with high symmetrical mobilities (1500–2500

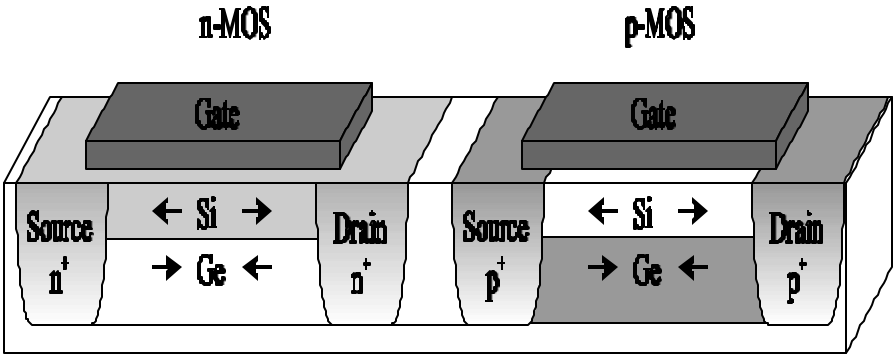

Figure 6. Scheme of a hetero-CMOS device. The strain in the layers is indicated by arrows (Si under tensile strain, Ge under compressive strain). The *n*-MOS (left) has a electron channel in Si, the *p*-MOS (right) has a hole channel in Ge.

$cm^2/V{\cdot}s$) for holes and electrons.[10] In Si CMOS the mobilities are moderately high but asymmetrical, with electron μ 800 cm^2/Vs and hole μ 250 cm^2/Vs. Elastically strained SiGe is the only material system that offers the desirable property of high symmetrical mobilities for both carriers (holes in strained Ge, electrons in strained Si). Figure 6 shows a schematic layout of a hetero-CMOS device.

In strained Si/Ge (with Ge under compressive and Si under tensile strain) a type II heterointerface lets the electrons jump in the Si channel and the holes in the Ge channel, respectively.[10] The strain status is obtained by a virtual substrate and relaxed SiGe layer on top of it. The realization of this high performance hetero-CMOS requires research on improvement of the crystal quality of the virtual substrate and on growth of SiGe layers with high Ge content.

7. Outlook

Within a few years the SiGe-HBT has emerged as the fastest Si-based transistor and several companies are now prepared to produce HBT based ICs. A scheme for a superior hetero-CMOS is available. Strain adjustment by virtual substrates and growth of highly strained layers need additional research efforts. Silicon-germanium based optoelectronics will also benefit from research in this area. System integration with demand for different devices will be of increasing importance. A unified heterostructure technology with a common layer structure for different devices could solve obvious problems for system integration and may compete successfully with CMOS in this growing segment of the semiconductor market.

References

1. K. M. Strohm, J Buechler, and E. Kasper, "SIMMWIC rectennas on high-resistivity silicon and CMOS compatibility," *IEEE Trans. Microw. Theory* **46**, 669 (1998).
2. P. Singer, "The dawn of quarter micron production," *Semicond. International*, January 1997 issue, pp. 50-56 (1997).
3. Semiconductor Industry Association Roadmap, 1994.
4. C. Y. Chang and S. M. Sze, *ULSI Technology*, New York: McGraw-Hill, 1996.
5. D. J. Paul, "Silicon-germanium heterostructures in electronics: the present and the future," *Thin Solid Films* **321**, 172 (1998).
6. E. Kasper and J.-F. Luy, "Molecular beam epitaxy of silicon-based electronic structures," *Microelectronics J.* **22**, 5 (1991).
7. M. Mastrapasqua, C. A. King, P. R. Smith, and M. R. Pinto, "Functional devices based on real space transfer in Si/SiGe structures," *IEEE Trans. Electron Dev.* **43**, 1671 (1996).
8. E. Kasper, "Prospects of SiGe heterodevices," *J. Crystal Growth* **150**, 921 (1995).
9. E. Kasper, ed., *Properties of Strained and Relaxed Silicon Germanium*, EMIS Datareviews Series No. 12, London: IEE, INSPEC, 1995.
10. F. Schäffler, "High mobility Si and Ge structures," *Semicond. Sci. Technol.* **12**, 1515 (1997).