Materials Science in Semiconductor Processing xx (xxxx) xxxx–xxxx

Contents lists available at ScienceDirect

Materials Science in Semiconductor Processing

journal homepage: www.elsevier.com/locate/mssp

Ion implantation of advanced silicon devices: Past, present and future

Michael I. Current

Current Scientific, San Jose, CA, USA

ARTICLE INFO

Keywords:
Ion implantation
Dennard scaling
MOS channel doping
Fully-depleted channel
Quantum confinement
Single-ion implanters

ABSTRACT

Ion implantation has been a key enabler, along with improvements in lithography, for the 40+ year evolution of MOS and then CMOS devices. Alterations in the channel doping levels followed the template developed by Dennard in the mid-70's from feature sizes of mm to tens of nm. When increasing channel doping in bulk planar CMOS created unacceptably high leakage current problems, ion implantation process developed past the "End of the Roadmap" to the "geometry-controlled" channels of fully-depleted finFETs and FDSOI of the present day. Future applications for nm-scale devices call for new understanding of ion damage accumulation in fin and nano-wire materials, consideration of effects of quantum confinement on channel conductivity, development of new implantation tools for efficient operation in the 100 eV range with ion and neutral species and soon after, for single-ion doping for quantum entangled, "atomic" electronics.

The structure of this paper frames the discussion in three time periods; (1) the 4 decades since the mid-1970's with the introduction and growth of Si-based ICs, first as bipolar and then MOS and CMOS devices, (2) the present day applications of ion implantation, focused on the process conditions related to the doping of finFETs and other 3D structures, and (3) the near-future (next 5–10 years) applications for ion (and energetic neutral) beams for processing of nm-scale semiconductor structures.

1. The past: ion implantation as enabler of CMOS scaling: 1970–2020

By the early 1970's it was recognized, at IBM, Intel and elsewhere, that, even though the ICs of the day were almost all based on bipolar transistors, the manufacturing "simplicity" and low-power operation of metal-oxide-semiconductor (MOS) transistors will greatly facilitate the development of increasingly complex IC circuits. This trend, recognized by Gordon Moore at Intel in 1965, that the number of transistors and memory bits in IC devices will double in approximately 2-year intervals, became a driving goal for the IC industry for half a century. By 1974, Robert Dennard at IBM systematized the changes in MOS transistor structure and operating characteristics that would follow from improvements in patterning technology. "Dennard scaling" for MOS transistors, built on the use of ion implantation to provide steadily shifting junction depths and dopant concentrations as lateral device dimensions were reduced, or "scaled", provided a template for design of next-generation transistors and fabrication process conditions. With the wide-spread introduction of CMOS transistors in the mid-1980's, replacing higher operating power bipolar, the combination of the planned regular scaling of lateral transistor dimensions by 0.7 (leading to a device area shrink of 0.49) for each new "node" and the corresponding shifts in junction depths and doping densities, guided by Dennard scaling, resulted in a global IC industry "roadmap" based on planar CMOS designs on "bulk" Si wafers. The exponential shrinkage of gate lengths, oxide thickness and junction depths (Fig. 1) for bulk CMOS continued until 2011.

Although technologically demanding, involving the use of vacuum processing, high voltages and requiring high temperature (> 1000 C) annealing for recovery of lattice damage and dopant diffusion, ion implantation brought key process and economic advantages to both bipolar and MOS transistor fabrication. For bipolar transistors, precise control of base implant energies and doses enabled much tighter controls on transistor gain with corresponding increases in device yield. For MOS devices, ion implantation was essential for setting channel doping levels for threshold voltage control and, when combined with self-aligned designs using poly-Si gates, reduced the source/drain-gate dopant overlap, resulting in faster and more controlled switching speeds. The ability of implant doping to use photoresist films as patterning masks, replacing the dielectric "hard masks" used for diffusion-doping, was a fundamental process simplification and positive economic driver. By the end of the 70's, ion implantation tools were well established as direct enablers of continued device scaling, especially for MOS and then CMOS chips.

E-mail address: currentsci@aol.com.

http://dx.doi.org/10.1016/j.mssp.2016.10.045
Received 25 August 2016; Received in revised form 19 October 2016; Accepted 27 October 2016
Available online xxxx
1369-8001/

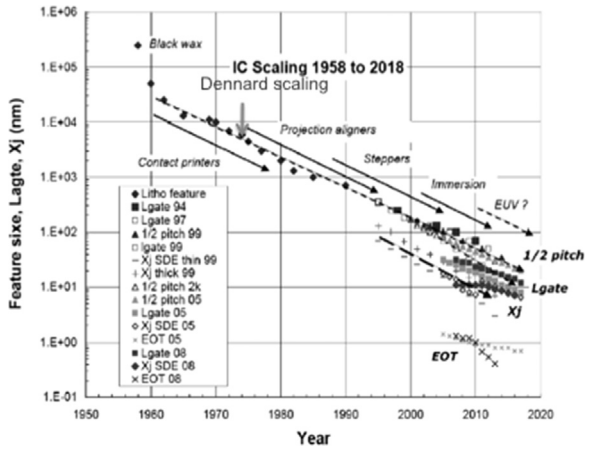

Fig. 1. IC transistor feature sizes (gate lengths, gate oxide thickness and source/drain junction depths) versus time as tracked by a succession of "roadmaps" for planar bulk CMOS.

1.1. Dennard scaling to the "End of the Roadmap"

Dennard's analysis of the doping requirements for increasing "scaled" MOS transistors highlighted the use of ion implantation for doping of the channel and source/drain regions as a key enabler of systematic shrinkage of transistor size [1]. The key doping step was the use of implant to set the channel doping level for a controlled depletion width, w_d, and corresponding threshold voltage. The channel doping levels, initially in the mid-e16 dopants/cm^3, were far below the controlled capability of diffusion doping. Dennard then developed a description of the coordinated changes in doping and device dimensions (L_{gate}, t_{ox}, w_d, etc.), in both lateral and vertical directions, and the corresponding shifts in drive voltages that maintained a *constant local electric field* as the transistor size is shrunk. This progression of "well-tempered" MOS (CMOS) transistors resulted in systematic improvements in switching speed and power requirements (Fig. 2).

However, one crucial feature of the "well-tempered" CMOS scaling plan was universally ignored from the start. To maintain constant local electric fields within the transistor, the drive voltage should also be reduced by the scale factor 1/k, where k≈1.4. The general design goal for operating circuits at the highest available switching speeds led to the delay in reducing the supply voltage from its 1974 value of 5 V to the present day value of ≈1 V. If Dennard scaling were followed in detail, circuit drive voltages would all be in the vicinity of 0.2 V (Fig. 3) [2]. Scaling physical dimensions of MOS transistors while holding the drive voltage constant has the desired effect of decreasing the switching delay by 1/k^2 instead of the constant field (Dennard) case of 1/k. However a consequence of "constant voltage" scaling is the increase in channel doping by k^2 rather the linear, k, scaling with constant field designs. This will accelerate the conditions that eventually led to the breakdown of continual scaling for bulk planar CMOS transistors.

Other than the pace of reduction of supply voltages, "Dennard scaling" provided a detailed template for CMOS transistor designs for the decades of dimensional scaling shown in Fig. 1. For ion implantation process, the decades of 1970's until 2010 saw a steady, highly predictable, shrinkage of source/drain junction depths, leading to the use of lower ion energies, pre-amorphization implants to suppress channeling effects for light ion (Boron) doping and increasingly constrained thermal budgets to minimize dopant diffusion during damage annealing and activation.

1.2. Red Brick Walls

The evolution of CMOS transistor designs along the seemingly peaceful path shown in Fig. 1 was punctuated by a number of difficult challenges, referred to at the time as "Red Brick Walls". The early challenges for formation of "ultra-shallow" source/drain extension junctions was addressed in the late 1990's by the introduction of sub-keV ion implanters and a continued evolution of "rapid" annealers. Throughout the 2000's, the introduction of strained channels to increase carrier mobility using SiGe epi and other methods, then use of Cu metal lines and low-k inter-metal dielectrics to reduce RC signal delays and then introduction, for 45 nm CMOS, of high-k dielectrics and metal electrodes in the gate stack continually increased transistor drive currents with minimal dependence on scaling other than aggressive shrinkage of S/D contact pitch.

However one "Brick Wall" remained for continued scaling of bulk planar CMOS transistors, the consequences of increased channel doping levels to scale the channel depletion depth, w_d, and suppress lateral leakage currents. By 1998 it was recognized that the increased doping levels in the channel and "halo" regions would lead to unacceptably high leakage currents due to band-to-band tunneling (BTBT) for scaled MOS gate lengths of ≈25 nm (Fig. 4) [3]. Throughout the late 1990's and 2000's numerous process designs, such as "super-steep retrograde wells", were used as "workarounds" to BTBT leakage current effects. But by 2010, as gate lengths approached 25–30 nm, the "End of the Roadmap" for bulk, planar CMOS as described by Dennard in 1974 was reached.

1.3. Transition from dopant-controlled to geometry-controlled depletion widths

The industry response, again led by Intel in 2012, to the limitations of "dopant-controlled" channel depletion widths was to shift CMOS evolution to "geometry-controlled" channel lengths. Two design options were developed, both with channel widths considerably less than the ≈10 nm limit for bulk planar CMOS; (1) finFETs, where conduction in an ≈8 nm wide vertical channel is controlled by a "high-k/metal gate" electrode wrapped over the fin channel and (2) the use of planar "fully-depleted SOI" (FDSOI) where conduction in a ≈6 nm thick channel is controlled by bias on a top gate electrode and "back gate" bias on n and p-type doped wells below a relatively thin (≈25 nm) buried oxide (BOX). In both of these "fully-depleted" architectures, the ideal case is an un-doped channel. Gate lengths for both finFET and FDSOI transistors are expected to approach ≈10–15 nm in "7 nm" node devices. Further development of CMOS channels into nm-scale "gate-all-around" nano-wire arrays provides pathways for CMOS evolution well beyond the "End of the Roadmap".

2. The present: doping challenges for finFETs

The first commercial utilization of finFET devices was introduced by Intel in their “22 nm” products. Each CMOS transistor was formed from single or multiple fin-channels etched in bulk Si. The challenges for doping CMOS junctions in these vertical arrays are significantly different than planar devices. These include: (1) the need for conformal (uniform in depth) doping in the SD contact and extension regions so that the carrier conduction is uniform in the body of the fin-channel bounded by the gate electrodes, (2) the use of little or no doping in the channel region to take obtain high carrier mobility and avoid threshold voltage variations, (3) the tight pitch (≈42 nm for the Intel "14 nm" designs) of the fins limits the beam incidence angles that can be used and still avoid shadowing from neighboring fins to ≈10° off-vertical or less, (4) the combination of the increased surface area of a multi-fin array over the equivalent planar area and the ion reflection, recoils and sputtering for grazing angle ion incidence increases the effective dose and desired beam currents for doping finFET devices, and (5) damage accumulation and annealing in high-aspect ratio vertical fins are significantly different from the well-known challenges for planar junctions and need to be re-learned in this new context.