TWIN-TUB CMOS - A TECHNOLOGY FOR VLSI CIRCUITS

L. C. Parrillo, R. S. Payne, R. E. Davis*
G. W. Reutlinger and R. L. Field

Bell Laboratories
Murray Hill, New Jersey 07974

ABSTRACT

CMOS technology has been developed through several generations of design rules with an n-type substrate (where p-channel transistors were formed) and with a p-tub implanted and diffused region (where n-channel transistors were formed). In order to enable a separate optimization of both transistors and to utilize the dopant control available with implanted layers, a two-tub approach was adopted. Utilizing lightly doped epi on an n^+ substrate (for latch-up protection), nitride-masked self-aligned tubs, $10^{16}cm^{-3}$ surface doping and 600Å gate oxides, an 8-mask CMOS process (named "Twin-Tub") was formulated. The combination of n on n^+ epi and careful I/O layout renders the circuits latch-up free. Novel aspects of the process, the devices it produces and finally the resultant circuit performance are herein described.

I. INTRODUCTION

CMOS provides an inherently low power static circuit technology which has the capability of providing lower power-delay product than a comparable design-rule NMOS technology. In the VLSI era, large random logic devices (like microprocessors) are often power limited, i.e., the total power dissipation can limit the gate count of a chip and hence limit its scale of integration and ultimate performance.

The n-channel transistor in the classic bulk CMOS process was formed in a p-tub which is diffused into a moderately doped ($2-4\times10^{15}$ cm^{-3}) n-substrate where the p-channel transistor was formed. For control purposes, the p-tub had to be doped approximately an order of magnitude greater ($\sim2\times10^{16}cm^{-3}$) than the n-substrate. In scaling to finer design rules, the doping in the n-substrate has to increase to support voltage in the shorter channel length p-channel transistors. Correspondingly, the p-tub doping has to be raised proportionately, and the n-channel transistor will suffer from excessive source/drain to p-tub capacitance. In order to avoid these scaling problems, a new Twin-Tub CMOS process was designed to provide separately optimized tubs and hence a lower capacitance n-channel transistor than could be obtained with the conventional process. With n^+ polysilicon doped with a separate phosphorus diffusion the parasitic thresholds of the field regions were designed to be 12V (without extra channel stoppers) with the n-channel active transistor thresholds at 0.6V and the p-channel active transistor thresholds at -1.9V. A p-channel threshold adjustment implant was added to reduce the threshold to -0.6V and enable circuits to be designed with operation down to 2V. In an effort to save masking steps, all the sources and drains are implanted nonselectively with boron and n^+ sources and drains are selectively implanted with a higher dose of phosphorus. Both implants are done through the gate oxide, obviating the need to expose and possibly contaminate the gate edges. The structure is completed with a flowed p-glass, aluminum metallization and plasma deposited SiN cap layer.

* Present Address: Analog Devices, Wilmington, Mass. 01887

The resultant technology has 1 μm junction depths for the source and drain regions, 2 μm electrical channel lengths, $\sim10^{16}cm^{-3}$ surface concentration with ±0.6V active device thresholds, 12 and -20V parasitic n- and p-channel device thresholds, and the ability to shrink in an evolutionary way as patterning technology improves. The remainder of this paper will describe the process, the device properties and examples of circuit performance demonstrated in the Twin-Tub technology.

II. THE PROCESS

Figures 1a-e show the the wafer in cross section as the process develops. Key features of the process are the tub formation, thinox etching, source and drain implantations, window definition in the p-glass, and metallization.

1. Tub Formation

The tubs are formed in a self-aligned manner by first defining an oxide-nitride sandwich (MASK 1). The n-tub is implanted (Fig. 1a) and selectively oxidized. The remaining nitride is removed and the p-tub is implanted self-aligned to the oxide masked n-tub (Fig. 1b). Both tubs are driven to a depth of 5-6 μm. The field oxide is then grown to a thickness of ~1 μm.

2. Thinox Definition

The 1 μm thick field oxide is nonselectively implanted with argon at a low energy(1). This generates a fast etching surface layer which in turn generates a 40° taper of the field oxide during BHF etching (MASK 2) of the areas where gate oxide will be grown (Fig. 2). The definition of the 1 μm oxide with sloped walls requires a 2 μm compensation of the masks. Hence, it is envisioned that this is probably the last generation of design rules where a tapered field oxide will be used. Anisotropic, steep-wall definition or a selective oxidation process for the formation of the gate regions along with steep-wall-tolerant lithography will be required in future technologies.

3. Polysilicon Processing

The gate oxide of 600Å is grown and the resultant structure has a surface doping of $\sim10^{16}cm^{-3}$ in each tub. The fact that the two tubs have comparable surface concentration is made possible by the Twin-Tub approach. Since n^+ polysilicon is used, the unadjusted thresholds would be 0.6V for the n-channel and -1.9V for the p-channel. The desired p-channel threshold is -0.6V and hence boron (MASK 3) is selectively implanted to reduce the threshold to the desired value. The sensitivity of the threshold to the implant doses is shown in Fig. 3. The polysilicon is then deposited and doped with phosphorus. The polysilicon is defined and etched (MASK 4) with an anisotropic plasma technique(2).

4. Source and Drain Formation

In order to save another mask, the source and drain formation has been done by first implanting boron into *all* sources and drains (Fig. 1c) and then masking (MASK 5) everything but the n^+ (n-channel) sources, drains and contacts. Phosphorus is implanted at

29.1

CH1616-2/80/0000-0752 $00.75 © 1980 IEEE

4X the boron dose, overcompensating the existing boron and forming the n^+ sources and drains (Fig. 1d). Subsequent glass flow processing will drive the junctions to ~1 μm deep.

5. Intermediate Oxide

Phosphorus doped glass is deposited (low pressure CVD), flowed, gettered (front of glass and back of wafer) with a phosphorus diffusion and patterned (MASK 6) with plasma etching(3). The windows are reflowed, cleaned and aluminum is then deposited.

6. Aluminum

The aluminum is patterned (MASK 7) with plasma etching, sintered and capped with plasma deposited SiN (4). Finally, vias are opened in the SiN for bonding (MASK 8). The completed device cross-section is shown in Fig. 1e.

III. RESULTANT DEVICE PROPERTIES

1. Active Transistor Properties

The polysilicon gates for both types of transistors are defined in the same step and are designed to have the same nominal polysilicon gate length. Since the n^+ and p^+ source/drain junctions are diffused to the same junction depth (1.0-1.2 μm), the channel lengths are approximately equal. This physical symmetry aids in the control of the ratio of current drive per unit gate length for the two types of transistors. Small variations in etching or thermal drive have proportional effects on both types of transistors and hence cause little variation in their beta ratio. It is notable that while the transistor betas of both types of transistors vary ±35% about nominal, the beta ratio varies less than ±15% for minimum length transistors.

The active channel length for each type of device is nominally 2.0 μm, and is consistent with a 7V maximum supply voltage. For higher voltage operation, the avalanche current in the n-channel transistor becomes significant and longer channel lengths must be employed.

To first order, the threshold voltages of the n- and p-channel transistors are independent. The tubs are independent of the epi doping and determined to first order by the separate tub implants. Second order effects like dopant contamination during tub oxidations and drive-ins can effect one, the other or both of the tub surface concentrations. Furthermore, the p-channel devices have their thresholds reduced by more than 1V with a threshold adjust implant. Only gate oxide thickness and Q_{SS} affect both devices in a correlated way.

2. Parasitic Transistor Properties

The two tubs, outside of the active transistor regions, are covered by ~1 μm of field oxide. The n^+ polysilicon runners over the tubs define parasitic transistors with 13V thresholds over the p-tub and -19V thresholds over the n-tub. The most critical region of the devices is the border of the tubs. Inter-diffusion of the two tubs makes a region of almost 10 μm wide where the surface concentration of both tubs is well below its value far from the edge. The lightly doped region can be inverted at low voltage and hence an adjacent drain must be well removed from the edge. A 7 μm design rule for n^+ diffusion to p-tub edge has been established to account for the out-diffusion of the p dopant at the edge of the tub. Both two-dimensional calculations (Fig. 4) and experimental measurements (Fig. 5) have justified the need for this much space.

IV. CIRCUIT PERFORMANCE AND SIMULATIONS

The circuit performance of the technology has been initially characterized by measuring the propagation delay through, and simulating, ten strings of circuits of various types. The ability to match the simulations of these simple strings of devices with measured propagation delays has been utilized to provide confidence in the computer simulations of complex circuits which proceeded in parallel with the process development. A procedure was established whereby simulation parameters were automatically extracted from test chips on specific wafers. The circuit operation was then simulated using these extracted model parameters. The simulated results were then compared with the actual circuit performance measured on the specific wafers from which the model parameters had been extracted. This procedure was carried out on a large number of wafers(5). The agreement between simulations and measurements was typically 10%, with the simulation being faster.

Distributions of values for model parameters in this technology have been established from routine parameter extraction on many lots. From these distributions, a worst case, nominal, and best case set of technology simulation files was created to predict the circuit performance variations which reflect processing variations. An example of the distribution of circuit performance displayed against the technology simulation files is shown in Fig. 6 for a simple inverter with fan-out of two. Figure 7 illustrates the results for a 3 input NAND gate with fan-out of 6. In each figure, the arrows represent the simulated performance using the worst case, nominal, and best case simulation files.

SUMMARY

A new 8-mask Twin-Tub CMOS technology has been developed which allows separate optimization of each type of transistor. The initial applications are in the field of microprocessors, microcomputers and custom logic. The device properties are well controlled and designed to be scaled smoothly into the improved lithography of the future.

ACKNOWLEDGMENT

The development of a new generation of IC technology requires a large team effort. From the first debates on performance needs to the final transfer to manufacture, many individuals have contributed to the Twin-Tub CMOS technology development. To these colleagues we extend our thanks.

References

(1) R. A. Moline, et al., *IEEE Trans.* **ED-20**, No. 9, p. 840 (September 1973).
(2) T. M. Mayer, J. H. McConville, "Linewidth Control in Anisotropic Plasma Etching of Polycrystalline Silicon," *IEDM*, Paper 3.2, December 1979.
(3) R. A. Porter, W. R. Harshbarger, J. T. Clemens and J. D. Cuthbert, "Plasma Etching of Phosphosilicate Glass," *Electrochem. Soc. Mtg.*, Extended Abst. No. **193**, October 1978.
(4) A. K. Sinha, H. J. Levinstein, T. E. Smith, G. Quintana, S. E. Haszko, *J. Electrochem. Soc.*, **125**, p. 601 (April 1978).
(5) H. H. Lehner (to be published).

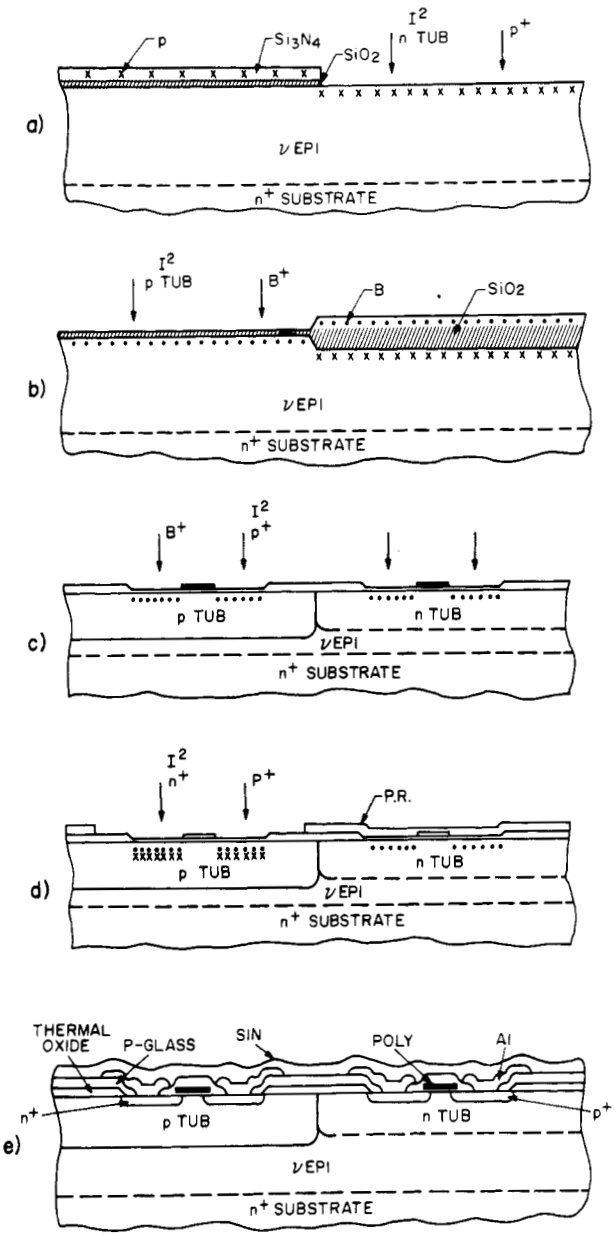

Fig. 1 Cross sections of Twin-Tub devices at several stages of the process.

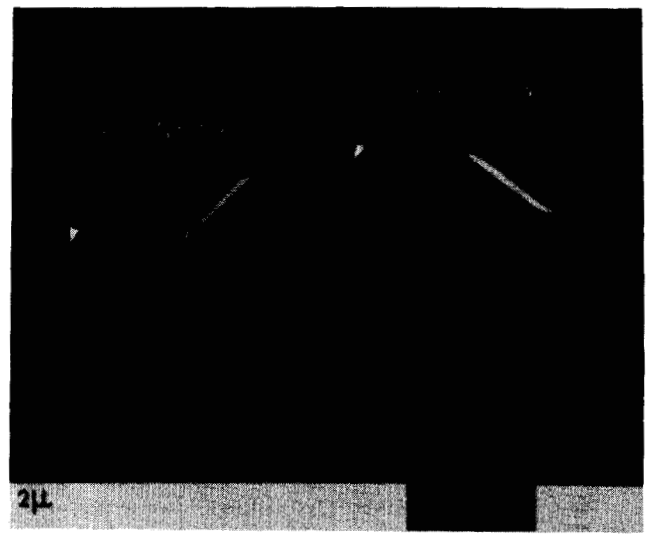

Fig. 2 SEM cross section of a device after the etching of the thinox regions.

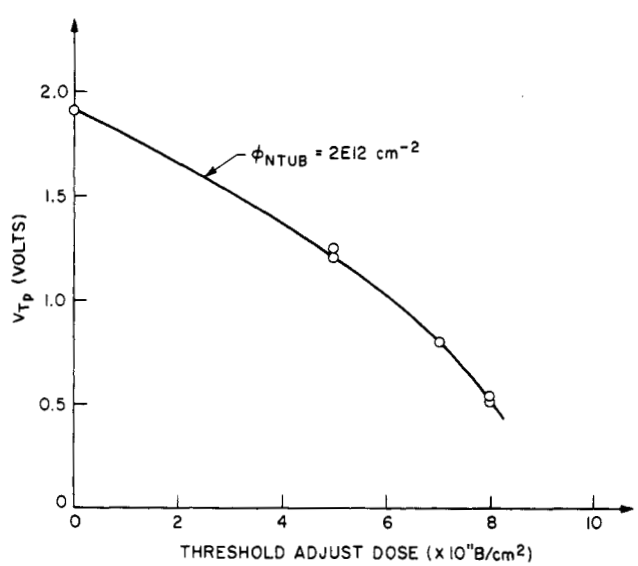

Fig. 3 Sensitivity of the p-channel transistor threshold to n-tub dose and threshold adjust dose.

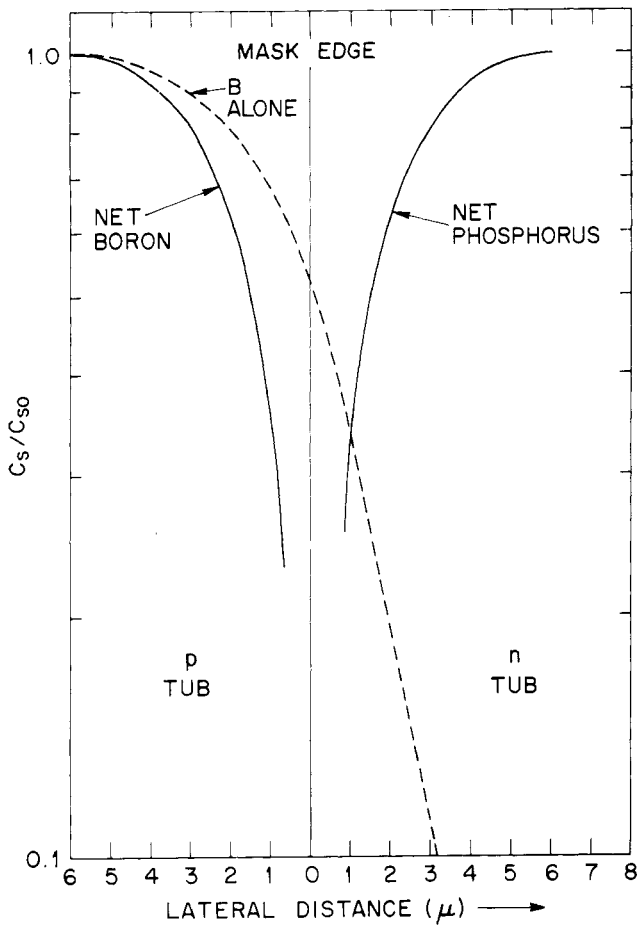

Fig. 4 Two dimensional diffusion profiles of boron and phosphorus at the border of the two tubs.

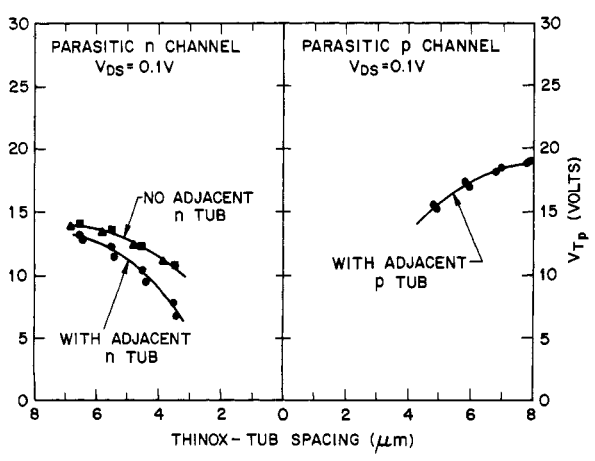

Fig. 5 Threshold voltage of the parasitic n- and p-channel transistors as a function of thinox to tub-edge spacing.

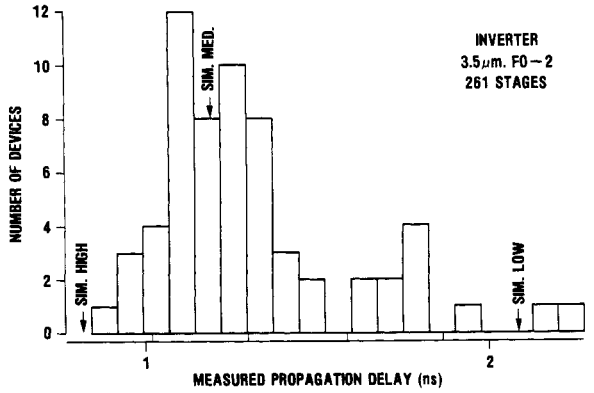

Fig. 6 Distribution of circuit performance displayed against the simulation file limits, inverter with fan-out of 2.

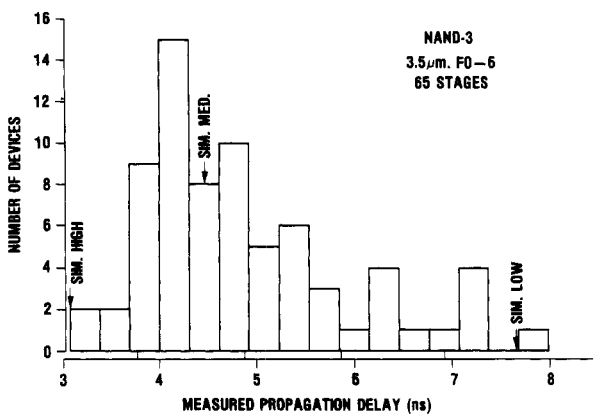

Fig. 7 Distribution of circuit performance displayed against simulation file limits, 3 input NAND with fan-out of 6.

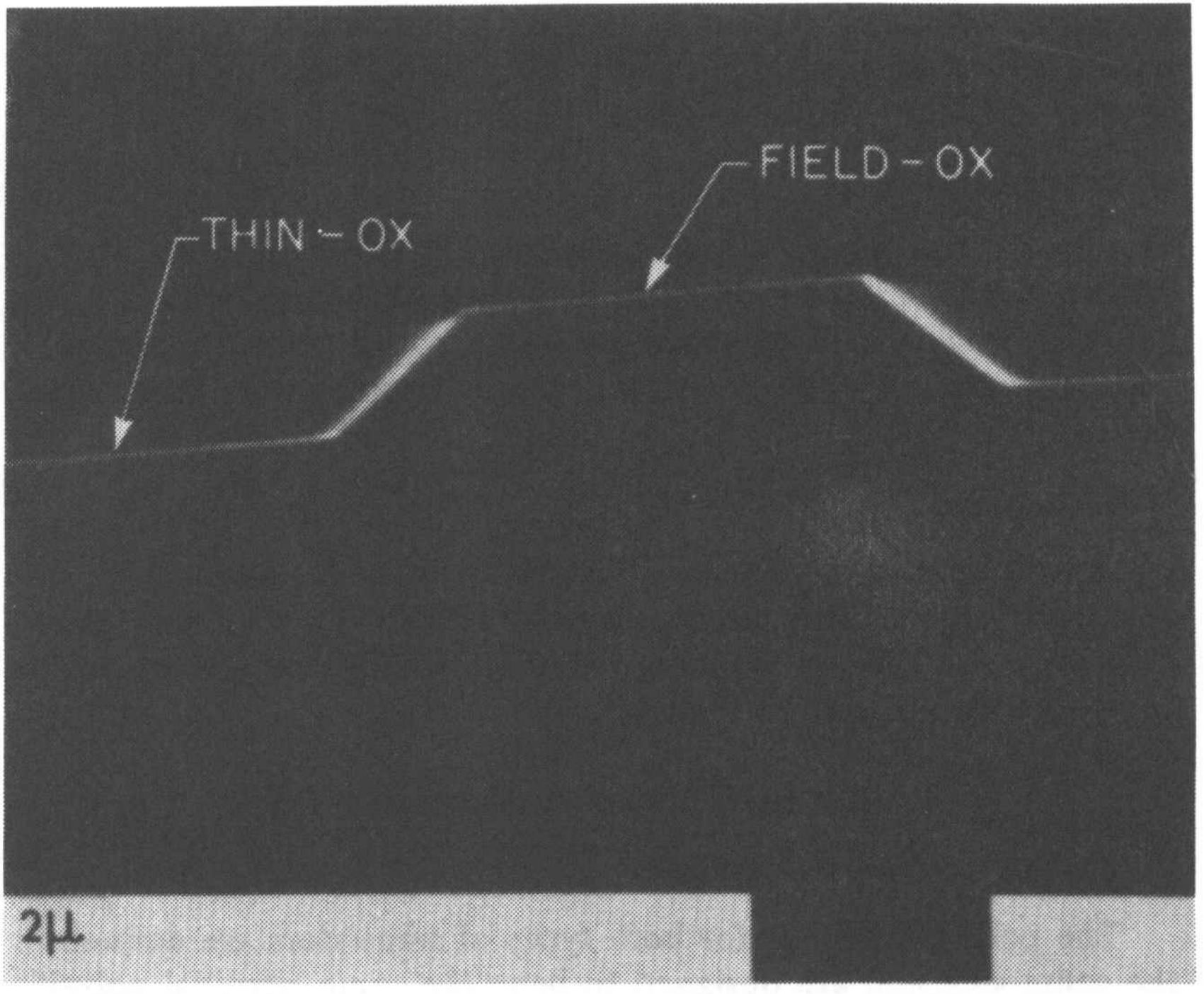

Fig. 2 SEM cross section of a device after the etching of the thinox regions.