

# Origins of Weak Inversion (or Sub-threshold) Circuit Design

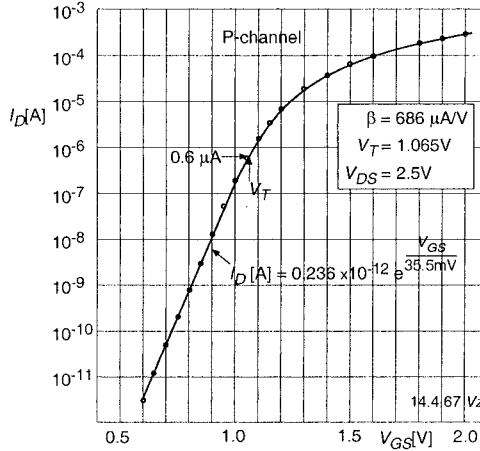
by Eric A. Vittoz

The state of weak inversion at the silicon surface in a metal-insulator-silicon structure was already implicitly mentioned as the “parabolic region” in the early paper of Garrett and Brattain on the MIS diode [21] (see also [22]). This particular situation is characterized by the fact that majority carriers have been repelled away from the surface, leaving a depletion charge of fixed atoms. The density of minority carriers is increased with respect to the distant bulk, but it is still negligible in the overall charge balance, and does therefore not affect the capacitance-voltage curves of the MIS diode. However, these minority carriers are the only mobile charge available at the surface. Hence, as soon as some voltage is applied between the source and the drain of a MOS transistor structure, they move by diffusion, thereby producing a drain current.

This current was ignored for years, since it was at the useless sub-microampere level, even for rather wide transistors. Indeed, MOS transistors were used, and are still mostly used, with a strongly inverted channel, i.e. with a density of inversion charge comparable to, or larger than, that of the depletion charge.

The very first application that needed to limit the power consumption of integrated circuits at the microwatt level was the electronic watch. Early developments at CEH (Watchmakers’ Electronic Center, Switzerland) started in bipolar technology [23], but it soon became obvious that the newly proposed CMOS technology [24] was ideally suited. The main problem in developing the logic part of the system (dominated by frequency dividers) was to obtain threshold voltages below 1 volt to ensure strong inversion at the low supply voltage of 1.3 volt in order to reach the required speed. But the heart of a watch is its crystal oscillator, for which a sufficient continuous transconductance must be created with the fraction of microampere of available current.

The characteristics of MOS transistors were then measured at this very low current level, and they showed the unusual exponential dependency of the drain current on the gate voltage depicted in Figure 2.1. Weak inversion then came to the attention of the digital design community under the name “sub-threshold current”. Indeed, it was the residual “leakage current” that kept flowing through a MOS transistor when it was supposed to be blocked by imposing  $V_{GS} = 0$ .



**Fig. 2.1.** Early measurement of the  $I_D(V_{GS})$  characteristics of a P-channel metal-gate MOS transistor (cleaned-up plot from E. Vittoz’ notebook, CEH, 1967).

This sub-threshold current was modeled in a succession of papers. In 1972, Barron [25] showed that this current was exponentially dependent on the surface potential (as can be expected from the same dependency of the surface charge), but he had no simple relationship with the gate voltage.

Immediately after, Swanson and Meindl [1] explained that the surface potential is related to the gate voltage by a capacitive divider, the ratio of which is approximately constant for a wide range of current. They characterized this ratio by a factor  $n$  that became the slope factor in subsequent models. In this pioneering paper, the authors applied their model to find the transfer characteristics of a CMOS inverter in weak inversion, showing that CMOS logic circuits can operate at a supply voltage as low as  $8kT/q$ . It is the very first publication on weak inversion (or sub-threshold) digital circuits.

In 1973, Troutman and Chakravarti [26] introduced the effect of non zero source voltage and extended the model to include short-channel effects. Troutman later derived an explicit expression of the sub-threshold slope [27]. The model presented in 1974 by Masuhara, Etoh and Nagata [28] used several equations to describe the current in the whole range of operation, but the relation between gate voltage and surface potential remained complicated.

A limited small-signal model was published in 1976 with a proposal to apply it to amplification [29], but the first analog experimental circuits exploiting weak inversion were presented the same year at the second European Solid-State Circuits Conference (ESSCIRC) [30]. It is interesting to mention that one comment from the audience was that such circuits, which use the “leakage current” of transistors, could not be reliable. Among these circuits was an amplitude-regulated crystal oscillator that has since been integrated by the tens of millions in electronic watches. Another was a current reference circuit that was directly inspired from its known bipolar version. The compact model used to describe the drain current

$$I_D = I_{D0} \exp \frac{V_G}{nU_T} \left( \exp \frac{-V_S}{U_T} - \exp \frac{-V_D}{U_T} \right) \quad (2.1)$$

was inspired from all previous publications and was derived in the extended paper published the following year [31]. In this model, all voltages are referred to the (local) substrate in order to preserve the intrinsic symmetry of the device. As was pointed out later [32], this model is very similar to the Ebers-Moll model [33] for a bipolar transistor having negligible base current. A related small signal AC model was first presented in 1977 [34]. It included the experimental evidence that channel noise is indeed the shot noise that can be expected for a barrier-controlled device.

During the following decade the interest for exploiting weak inversion in analog circuits was very limited world-wide [35, 36], and mostly concentrated in Switzerland [37, 38, 32, 39, 40, 41, 42, 43, 44, 45, 46] for the development of micropower circuits used in a variety of portable systems. These circuits had to use combinations of transistors biased at various current levels from weak to strong inversion. This was the motivation for the development of a continuous model, which was started by Oguey and Cserveny [47, 48] and became the EKV model [49].

In the late 80’s, a new wave of interest for weak inversion (or sub-threshold) circuits was triggered by Mead at Caltech. He promoted them as the best way to implement analog VLSI systems that mimic the operation of the brain [50]. Mead and Maher also introduced the charge-potential linearization in a charge-based model [51], which was later adopted for the EKV model [52].

Sub-threshold digital circuits remained totally ignored until the mid-90’s, with the newly recognized need of limiting the power consumption, in particular for portable systems. With modern deep submicron processes, it should be possible to reach clock frequencies much beyond 10 MHz with a supply voltage of just a few hundred millivolts and to drastically reduce thereby the power-delay product [53].

The reduction of supply voltage imposed by scaling-down process dimensions implies a reduction of the saturation voltage of transistors used in analog circuits. This is only possible by reducing the amount of channel inversion, thereby entering moderate inversion, or even weak inversion for supply voltages below half a volt.