

Compact Model Methodology for Dual-Stress Nitride Liner Films in a 90nm SOI ULSI Technology

R. Q. Williams, D. Chidambarrao*, J. H. McCullen, S. Narasimha*, T. G. Mitchell, D. Onsongo*

IBM Corporation, Essex Junction, Vermont, USA

*IBM Corporation, Hopewell Junction, New York, USA

Email contact: rqw@us.ibm.com

This work presents a novel methodology for a physically-based, layout-dependent nitride liner stress model that works with readily-available compact models. Nitride liner films with intrinsic stress are now being used to extend technology performance scaling trends[1]. While stress liner films can provide impressive performance gains, the strength of the effect is dependent on the physical structures adjacent to each transistor. Therefore compact models that encapsulate layout-dependent stress are essential for predicting circuit behavior and for facilitating circuit performance optimization. The methodology includes a data-calibrated, semi-empirical model and is tightly coupled to circuit netlist extraction for accurate adjacent structure information. The model formulation is summarized and simulation results for an SOI implementation are presented.

In dual-stress liner (DSL) technologies the fabrication process for the contact etch stop liner film (typically a nitride) is altered to deliberately change the stress seen in the MOS channel. Stress is transferred from the liner to the channel near the spacer edges and alters the underlying band structure and transport properties. Tensile stress is used to induce electron band shifts, improving electron transport, while compressive stress is used to induce valence band shifts, improving hole transport. The crystal stress response is relatively small compared to other crystal straining methods (e.g., strained silicon on relaxed silicon germanium buffers) and so the impact can be described with a linear stress-to-transport relationship. In this work, the stress response is reflected in the compact model by shifting mobility. The exact channel stress is in general a complex multidimensional tensor that depends on the precise details of all aspects of the physical transistor and its immediate neighborhood. To make this situation tractable for compact modeling, the stress is represented by two approximate one-dimensional components: a longitudinal term (parallel to current flow) and a perpendicular transverse term, both of which adjust mobility. Equation (1) illustrates how the stress components can be linearly decomposed

$$\sigma^L = \sigma_{self} - \sigma_{contacts} - \sigma_{poly} - \sigma_{DSL}^L \quad \sigma^T = \sigma_{DSL}^T + \sigma_{BC\ contact} \quad (1)$$

where the superscripts L, T represents longitudinal and transverse, respectively. The longitudinal stress component consists of a self term, which reflects a maximum stress with no adjacent shapes present, and stress degradation terms for the presence of local interconnect contacts, adjacent polysilicon gates, and the DSL interface. The transverse component consists of the effect of the transverse DSL interface and the effect of an optional SOI body contact (BC subscript).

In this work, dimensional information is provided from netlist extraction using the IBM Efficient Rapid Integrated Extraction (ERIE) tool [2] which preserves circuit hierarchy (see Fig 1). Distance measurements for the longitudinal and transverse stress terms are measured directly. The longitudinal stress term uses information about the leading and trailing edges of all shapes seen in the longitudinal direction out to a search distance of approximately 1 μm . Since a potentially large and unpredictable amount of shape information could be present, ERIE breaks the search regions on either side of the gate into “buckets” and returns the enclosed shape area, perimeter, and vertice count in each bucket to the compact model through the FET instance parameter list (see Fig. 2). The stress model is implemented in a modified version of the UC Berkeley BSIMPD model[3]. Alternatively, since the stress model contains no temperature or voltage-dependent expressions, the stress algorithm can be encapsulated as a stand-alone dynamically loadable library (DLL) and can be invoked during the netlisting phase for netlist simplification.

The stress model was calibrated using median I_{eff} [4] hardware data from structures designed to isolate the stress model components in equation (1). The starting point for the stress model fit is a compact model that is extracted from a reference device, and offsets are expressed as a normalized mobility scale factor. Fig. 3 shows the effect of interconnect S/D strapping, an effect due to the cut in the liner film. Fig. 4 illustrates the normalized degradation caused by DSL longitudinal edge proximity. Figs. 5 and 6 give the composite stress response for various layout styles. In conclusion, a novel integrated compact model-netlist extraction methodology for layout-dependent liner stress is successfully demonstrated.

References

- 1) Yang, et al., IEDM 2004, pg. 1075.
- 2) Habitz, et al., IBM Micronews, Vol. 7, No. 1, pg. 32.
- 3) UCB BSIMPD version 2.2.3 compact model, 2002.
- 4) Na, et al., IEDM 2002, pg. 124.

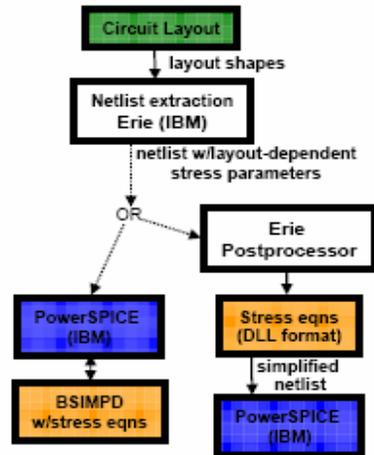


Fig. 1 Netlisting/compact model tool flow.

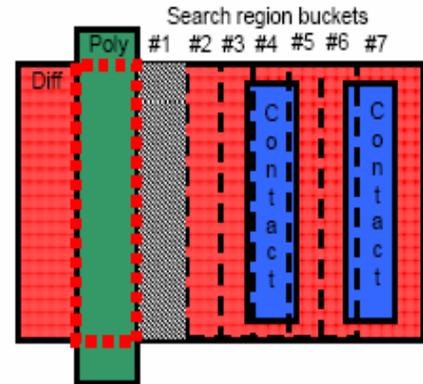


Fig. 2 Adjacent shape analysis. Hatched area (bucket #1) is the poly gate area intersected with the diffusion area and expanded towards the right side contacts.

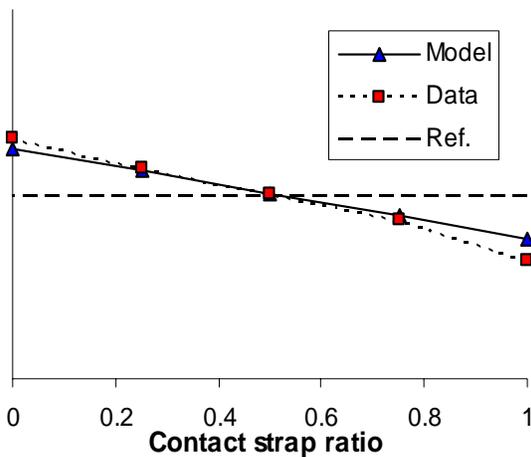


Fig. 3 Effect of contact strapping ratio in normalized units.

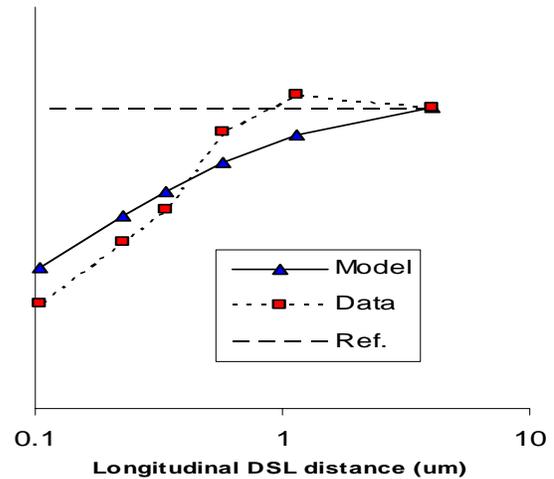


Fig. 4 Proximity effect of the longitudinal DSL boundary in normalized units.

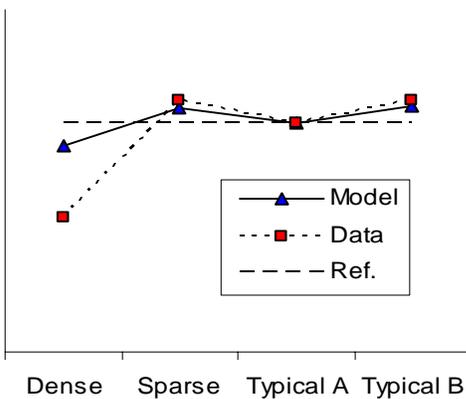


Fig. 5 Nfet composite stress model response for different layout styles (normalized units)

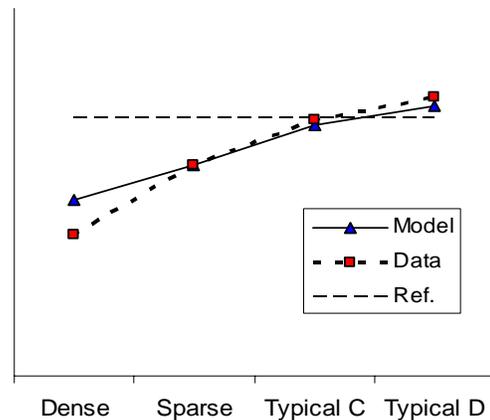


Fig. 6 Pfet composite stress model response for different layout styles (normalized units)