# Intel's 0.25 Micron, 2.0Volts Logic Process Technology

A. Brand, A. Haranahalli, N. Hsieh, Y.C. Lin, G. Sery, N. Stenton, B.J. Woo
California Technology and Manufacturing Group, Intel Corp.

S Ahmed, M. Bohr, S. Thompson, S. Yang
Portland Technology Development Group, Intel Corp.

Index words: CMOS, shrink, interconnect

## Abstract

Process 856 is a 0.25µm-generation logic technology currently in volume manufacturing, which has been optimized for high performance, yield, and density. This process is being used to manufacture high performance products including the Intel® Celeron™ and Pentium® II microprocessors. The process has a high equipment re-use rate to reduce cost. Using the older equipment has increased the challenge of scaling to smaller pitch, particularly in the interconnect process. Transistor optimization allows volume production of Pentium II microprocessors at 450 MHz. High yield has also been achieved, both before and after a 5% linear shrink of the initial 0.25µm design rules.

## Introduction

Process 856 (P856) is Intel's quarter micron (0.25µm) logic technology. In developing P856, the important goals were to achieve low cost through high equipment re-use, deliver a gate delay improvement of 30%, and deliver high yield. An equipment re-use goal of 70% was set: the actual level achieved was 85% [1]. A performance goal of 30% transistor delay improvement was set: this was exceeded by 18%. The yield improvement curve for the P856 is the fastest of any Intel process so far.

Each generation of high-performance, low-power microprocessor products requires progressively faster transistors with lower operating voltage, produced with higher density. Historically the rate of improvement in gate delay has been 30% per generation. Normally it takes two to three years to develop a new technology, and each technology generation is progressively more expensive. Through scaling and the introduction of key architectural features such as halo NMOS, P856 delivered a better than 30% delay improvement at certification, the key checkpoint for volume manufacturing.

A second post-certification technology enhancement project delivered a 5% linear shrink with an additional 18% delay improvement, using the same equipment set. This represents nearly a half technology generation improvement in performance and yield, and it was delivered at very low cost. The post-certification improvement was achieved through control improvement and further transistor scaling, including a reduction of gate oxide thickness, enhanced halo processing, and general optimization of transistor implant conditions. This transistor enhancement has been critical in achieving good binsplit for Pentium II processors at 450 Mhz.

In this paper, we describe the important architectural features in P856 that enabled scaling of the interconnect process and transistor enhancement. The transistor improvements made in the pre- and post-certification stages are described. We discuss some of the important issues for interconnect processing with quarter micron features. We also describe the approach used to achieve a 5% shrink of the initial design rules.

## Transistor Integration

### P856 Architectural Enhancement

A fundamental constraint for short channel length transistors is that as the channel length is reduced to improve drive current, the barrier to off-state leakage is decreased. Throughout the development of P856, the transistor was optimized to achieve the best Idsat at a given margin to leakage, while also striving for low capacitance. High transistor performance in P856 was achieved through aggressive scaling to 40.8A electrical

gate oxide and sub-quarter micron poly dimensions, and through the addition of the following architectural enhancements, to be described in detail:

- Silicon pre-amorphization implants
- NMOS and PMOS halo implants
- Junction compensation implants

Like the previous generation P854 (0.35µm) CMOS process, the P856 process flow uses 200mm P-/P+ epi wafers and begins with shallow trench isolation followed by implantation of N and P wells. The gate oxide thickness is scaled from 60A on P854 to 40.8A on P856. Complimentary doped polysilicon is used to obtain matched Vt in N- and P-MOS devices. Nitride spacers are used to separate the deep source drain regions from the shallow source drain extensions. TiSi2 is selectively formed on polysilicon and source drain regions, obtaining a worst-case sheet resistance of 5 Ω/sq. The transistor structure is illustrated in Figure 1.

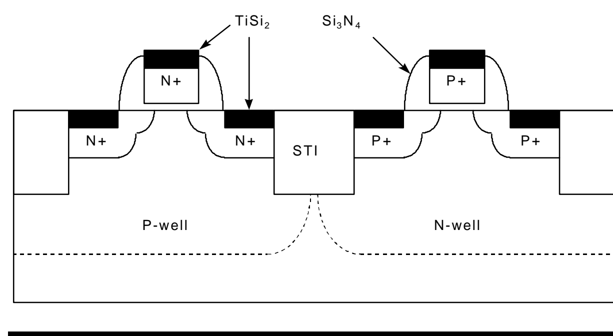



**Figure 1:** Schematic cross section of transistors

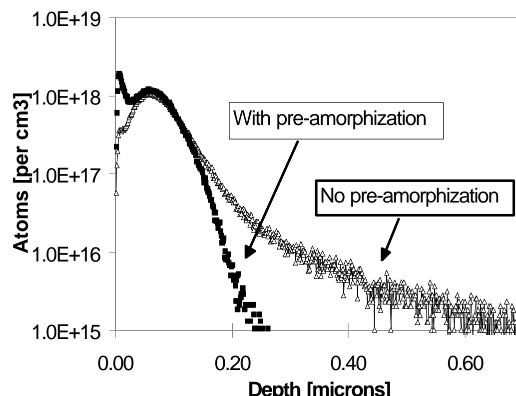



**Figure 2**: SIMS depth profile shows reduction in As implant tail due to Si pre-amorphization

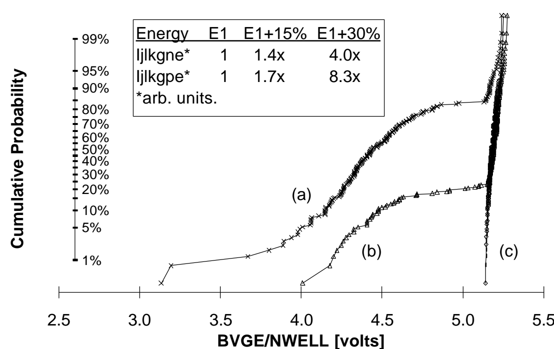



**Figure 3**: Increased Si pre-amorphization dose reduces the gate breakdown voltage. BVG failure rate is shown vs. a) 2X PA dose, b) 1.5X PA dose, and c) nominal PA dose.

## Silicon Pre-Amorphization Implants

A silicon implant is introduced in P856 after poly gate definition. It is used to create an amorphous layer in the polysilicon gate and source/drain regions of both the N and P devices. The amorphous layer reduces the channeling tails of subsequent implant steps resulting in abrupt implant profiles (see Figure 2). Reducing the lateral implant tails under the poly gate region is key to controlling the sub-threshold leakage in short channel devices. The dose and energy of the Si implant need to be high enough to amorphize the underlying region without degrading the gate oxide. Figure 3 shows that gate oxide leakage increases for higher energy implant, and that gate oxide failure, as measured by lower breakdown voltage (BVG), can occur when the dose is too high. The table inset in Figure 3 shows the impact on gate leakage.

## NMOS and PMOS Halo Implants

The short channel behavior of both NMOS and PMOS transistors was further enhanced by the introduction of halo implants. The halo implant is a high-angle implant introduced after Si pre-amorphization in the same lithography step used to dope the source/drain extension regions. Since the halo implant uses a high angle it must be done in four 90-degree rotations in the implant tool to ensure both sides of the channel are doped and that transistors oriented in both X and Y directions get doped. The halo implant uses the same implant type as the original well dopant (for example, N type dopant for the Nwell of the PMOS device).

The halo implant, together with the well implant, sets the threshold voltage of the transistor. By reducing the initial well implant dose and introducing the halo implant after

gate patterning, a non-uniform channel doping profile is achieved. Due to the angled implant, short channel devices receive a higher dopant concentration than do longer channel devices. There are several benefits when these implants are optimized. The halo implant reduces the Vt roll-off in short channel devices as shown in Figure 4. Since the same Vt is achieved with lower average channel concentration, the Vt with substrate bias is reduced as shown in Figure 5. Most important, higher Idsat at target is achieved because with a given Vt, the halo device has a more abrupt drain-channel junction and higher channel mobility than a non-halo device.

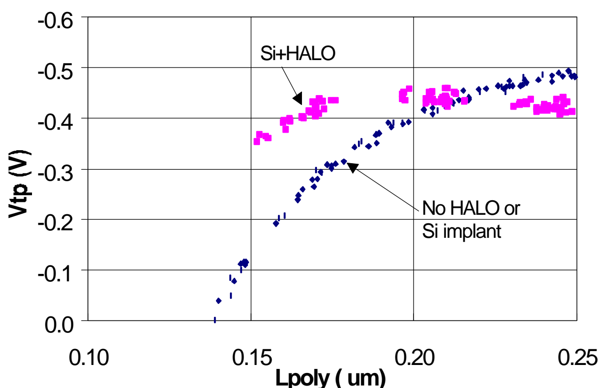



**Figure 4:** Reduction in PMOS threshold voltage roll-off with Si pre-amorphization and halo implant

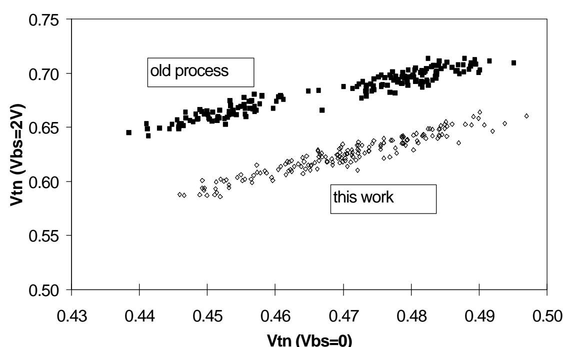



**Figure 5:** Reduction in substrate bias Vt effect

## Junction Compensation Implants

The third major transistor modification on P856 is the use of compensation implants to reduce junction capacitance. AC parameters play an increasingly important role in overall transistor performance, and junction capacitance was a high leverage parameter contributing to the performance of P856. A compensation implant is introduced in both N and PMOS devices during the same lithography sequence used for source and drain (S/D) implants. This implant uses the same type species as the S/D implant but with a lower dose and higher energy to give a more graded implant profile at the junction (see Figure **6**). The compensation implant conditions were chosen to give approximately a 20-30% reduction in junction capacitance (see Table 1) with no degradation of the isolation performance or the implant penetration of the gate oxide.

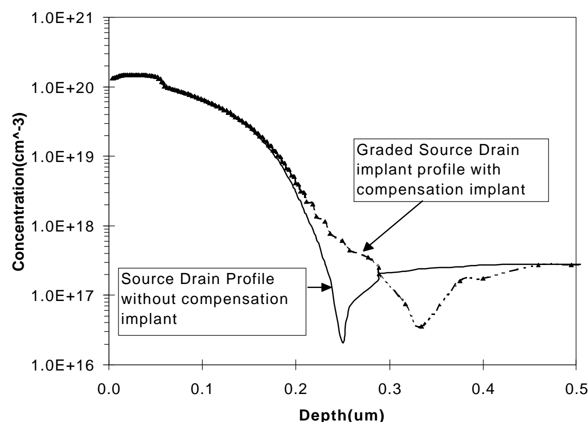



**Figure 6:** Junction doping profile with the addition of a compensation implant to reduce junction capacitance

| Type | Before | With | Change |
|---|---|---|---|
| N | 1.0 fF/µm2 | 0.7 fF/µm2 | -30% |
| P | 1.25 fF/µm2 | 1.0 fF/µm2 | -20% |

**Table 1:** Junction capacitance area component reduction attributed to compensation implants

## Transistor Performance Results

P856 was certified in Q3 1997 using halo implants, Si pre-amorphization implants, and n+ junction compensation [3]. Based on the common industry metric of 1nA/µm worst-case device leakage, the Idsat target of 0.585mA/µm for NMOS and 0.250mA/µm for PMOS was achieved. A simulated transistor delay metric known as FEM95 showed that the performance goal of a 30% delay improvement over P854 had been achieved.

| Time | NMOS Idsat | PMOS Idsat | FEM95 vs P854 | FEM95 |
|---|---|---|---|---|
| Certification | 0.585 | 0.250 | -33.2% | ref |
| Cert+2Q | 0.670 | 0.295 | -45.8 | -18.8% |
| Cert+4Q | 0.700 | 0.310 | -49.9 | -23.6 |

**Table 2:** Idsat target and FEM95 benchmark results as a function of time (in quarters) from certification (the FEM95 reference is P854)

To rapidly deliver significant additional performance, two process revisions were developed and implemented within a year of certification. The enhancement involved further thinning of the gate oxide to 40.8A, scaling of the poly target due to improved poly control, implementation of a p+ junction compensation implant, and re-optimization of the NMOS and PMOS halo, well, and S/D implants.

The halo implant re-optimization allowed a reduction in the N and P well surface implant, favoring an increase in the halo implant. The resulting transistors have well behaved sub-threshold characteristics (see Figure 7). As shown in Figure 8, we achieved Idsat at 1nA/µm of 0.755mA/µm for NMOS and 0.350mA/µm for PMOS. Accounting for the channel length control margin, we achieved industry pace-setting Idsat at target of 0.700mA/µm for NMOS and 0.310mA/µm for PMOS [4],[5]. These results are summarized in Table 2.

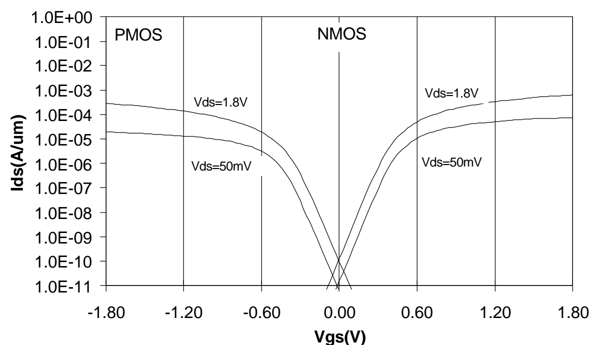



**Figure 7:** IV sub-threshold characteristics for NMOS and PMOS devices for target devices

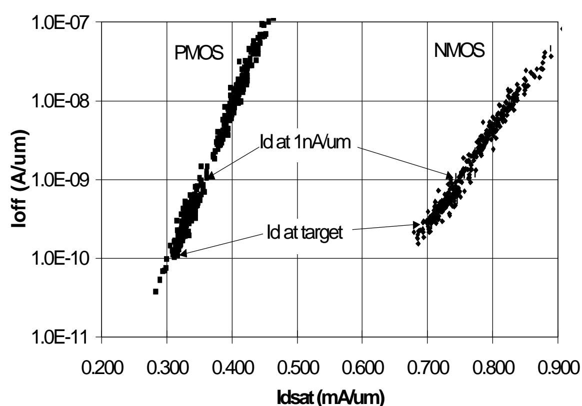



**Figure 8:** NMOS and PMOS drive current vs. leakage (the reference leakage current is 1nA/µm)

The improvement in performance has been demonstrated using the Pentium II microprocessor. Maximum speed measurements made at low-voltage and low-temperature conditions primarily show the improvement made in transistor performance. Under these conditions there is little influence from interconnect RC delay, because the interconnect sheet rho is reduced at low temperature. Figure 9 shows the progression in microprocessor path delay (period) as a function of time from certification. (In this figure, the data is smoothed for clarity, and the same stepping and test program is used in all cases.) A net 18.1% delay improvement has been observed on the same stepping of the Pentium II microprocessor. While there is dilution of the transistor improvement due to RC limited paths, with this enhanced process, better than 50% Fmax improvement has been achieved in microprocessor speed compared to the prior 0.35µm technology [6].

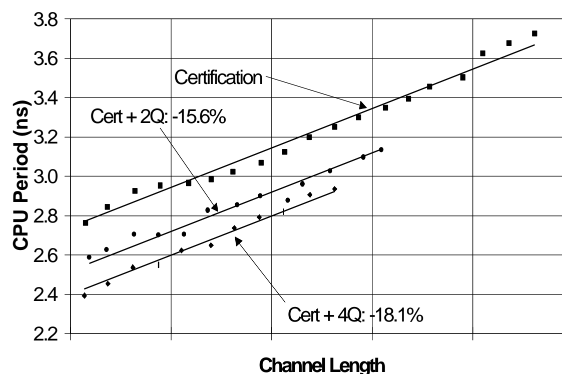



**Figure 9:** Microprocessor low voltage/low temperature delay improvement from post-certification process enhancement

All of the benchmarks discussed in this section are based on 1.8V transistor test conditions, and the P854 reference assumes the P854 and P856 run under nominal 2.5V and 1.8V conditions. To enable further performance enhancement, the reliability characterization of P856 was converted to a 2.0V nominal criteria. On products that can tolerate higher power consumption due to increased supply voltage, the 2.0V operation improves performance. Microprocessor characterization shows that there is an additional 9-10% frequency enhancement at 2.0V compared to 1.8V. At certification, P856 met the reliability goals for 2.0V operation.

## Interconnect Integration

P856 uses five metal layers that are optimized for microprocessor performance and density. Table 3 shows the intended functions for each layer. Intel's technologies for logic are optimized for high aspect ratio to provide the most competitive RC performance at the

best density. The M1 to M3 layers use tight pitch, which is necessary for good SRAM and logic cell routing density. The M4 and M5 layers use wide pitch and high thickness, resulting in the low sheet rho needed for power distribution and cross die interconnect.

As with previous Intel processes, the metal stack is Ti/Al-Cu/Ti/TiN, which provides low line and via resistance while meeting electromigration requirements. Also, as before, the first inter-layer dielectric (ILD) above poly is Boro-Phosphosilicate-Glass (BPSG). The BPSG is planarized using chemical-mechanical polishing (CMP). The remaining ILD layers are PTEOS oxide that use a deposition followed by an etch-back process followed by CMP planarization. The CMP steps improve layer planarity, which is necessary for the uniform lithographic and etch processing of multi-layer interconnects. Contacts and vias are all filled with tungsten plugs formed by blanket tungsten deposition followed by CMP.

| Layer | Pitch | Thickness | AR | Purpose |
|---|---|---|---|---|
| M1 | 608 nm | 480 nm | 1.6 | local connections |
| M2 | 882 | 900 | 2.0 | intermediate length RC |
| M3 | 882 | 900 | 2.0 | intermediate length RC |
| M4 | 1520 | 1325 | 1.7 | power / long RC |
| M5 | 2432 | 1900 | 1.6 | power / long RC |

**Table 3:** Metal layer pitch, aspect ratio, and intended applications

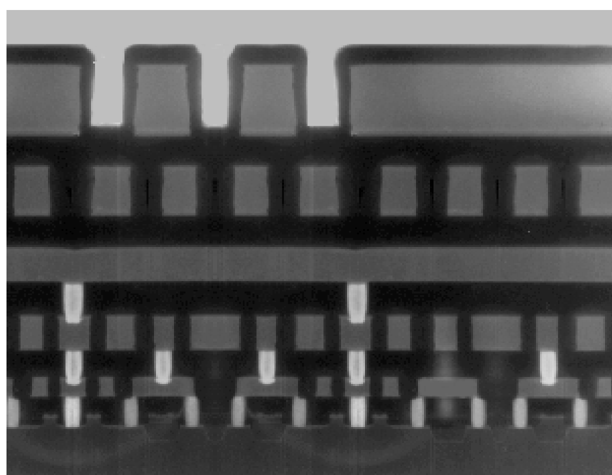


**Figure 10:** Five-layer metal interconnect cross section

To achieve cost savings, most of the metal-processing tools used in P856 were used in P854. The same stepper, metal deposition, contact etcher, metal etcher, and planarization equipment are used. A key challenge in the P856 interconnect has come from optimizing the lithographic and etch processes to work with the 20% smaller pitch of P856.

Just as Poly stretches the line width capability of DUV tools, Metal 1 patterning challenges the DUV lithography for space-limited capability, as the minimum space required is beyond the wavelength limits. This tight pitch (608 nm) demands thin photoresist for resolution, which in turn degrades the margin for metal etch due to resist erosion. The resist erosion results in poor metal line profile (shelving) and poor metal line critical dimension (CD) control.

Stringent control in depth of focus is also needed to ensure the integrity of the lithographic patterning. In order to achieve a planar surface for metal lithography, CMP is used prior to metal deposition for both ILD0 and contact plug steps. However, density variation causes local ILD erosion during CMP, which can result in severe variation in topography. For example, a depression as deep as 180 nm has been seen on the surface near a boundary between a dense memory array area and a loose periphery area. This depression causes a local area to be printed out of focus and results in a distorted metal line, as shown in Figure 11. Improved oxide and tungsten polishes that reduce the topographical step have been developed to ensure enough depth of focus on the surface.

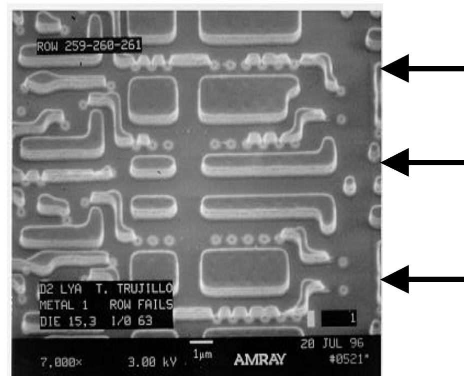


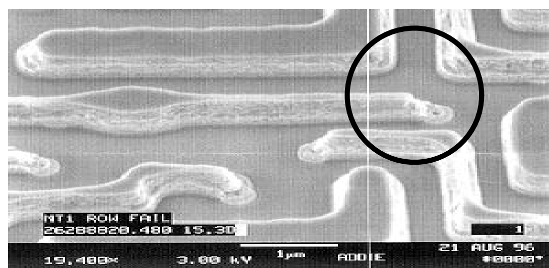

**Figure 11:** Metal 1 line distortion caused by ILD erosion induced out of focus lithography

Another limitation of lithographic capability is evident in the pullback at the end of a metal line. This pullback can cause a reliability problem when it is so severe that the metal line does not adequately cover a contact at the end of the metal line. Figure 12 shows a Metal 1 void bake

failure due to Metal 1 pullback and improper contact coverage. Twenty to 160 nm pullback has been detected in Metal 1 lines, depending on whether the structure is nested or isolated and on the location on the wafer. One solution to this pullback problem is to use optical proximity correction (OPC). These features compensate for the lithographic pullback effect at the end of the metal line. Contact coverage better than 75% has been achieved with OPC improvement, and the failure has essentially been eliminated.

**Figure 12:** Example of a Metal 1 void failure due to pullback, before process optimization

The P856 technology also places stringent demands on the metal etch control. Magnatron current and RF power are optimized to reduce the erosion of photoresist during etch and to provide enough sidewall passivation to protect the metal profile. A vertical Metal 1 profile without undercut and shelving was achieved while providing good metal CD control (Figure 13). In the high aspect ratio M1 process, incomplete etching due to cross wafer thickness and CD non-uniformity can result in metal stringer defects. This is addressed by limiting the M1 sputter deposition target lifetime, controlling the M1 grain size through minimum deposition chamber heating, improving the uniformity of metal thickness, reducing the metal electrical CD, and slightly increasing the over-etch time.



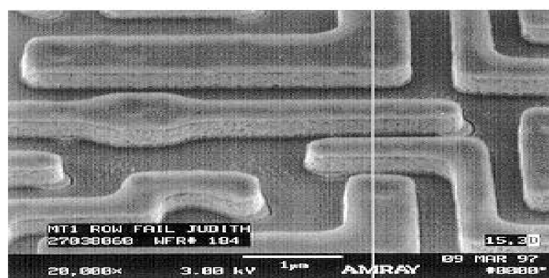




**Figure 13:** Metal 1 profile before optimization (top) showing shelving and M1 pullback, and after optimization (bottom) with good vertical profile & good contact coverage

Just as Metal 1 challenges the DUV limits, the Metal 2 and Metal 3 patterning with very thick films stress the I-line lithography limits. Both metals required significant improvements on processing issues, such as shelving of the metal profile, pullback at the end of metal lines, and bridging between narrow spaces. Optimized operating conditions have been determined for individual I-line lithographic tools to provide the best focus and exposure window. Together with an optimized metal-etching recipe, shelving is eliminated in the metal profile. Optimized reticle sizing for narrow spacing is used to provide adequate margin for the metal bridging. OPC is also used in Metal 2 and Metal 3 reticles to reduce pullback. The combination of these enhancements has successfully provided needed process capability in a production environment.

## 5% Shrink

A 5% linear shrink, known as P856.5, was applied to P856.0 in order to reduce die cost. A 5% technology shrink has been used in Intel in many generations as a standard means of cost reduction. Due to high equipment re-use in P856, the margin for shrink initially appeared tighter than in previous technologies. This required increased optimization of individual layers. Process margins and design rule margins were examined closely in order to achieve a "smart shrink" for minimum margin loss on the tightest part of the technology.

The smart shrink strategy uses optimum sizing for all critical layers and optimum targeting for critical

dimensions. For example, the high performance interconnect has high aspect ratio metal spacing as well as lines. The lithographic and etch margins are more critical for spacing than for lines in manufacturing. Therefore, a strategy of avoiding or minimizing shrinking metal spacing was adopted. Whenever possible, the metal line width rather than the space is shrunk. This helps ensure no degradation in speed due to the increased cross-talk if spacing were shrunk.

The reduced metal line width in the shrink technology does reduce certain design rule margins such as metal overlap of underlying via and metal enclosure of top via. To overcome this difficulty, OPC techniques were used creatively to systematically prevent the degradation of the design rule margin. For the contact layer, the proximity effect from clustered contacts became much worse after shrink, and a special selective sizing method was used on the reticle to restore the process margin. Due to very tight constraints, many layers required re-characterization. Table 4 shows the approaches used for critical DUV and I-line layers.

| Layer | Shrink Strategy | Re-Characterized | Mask Fix/OPC |
|---|---|---|---|
| Isolation | Line / Space | Yes | No |
| Poly Gate | Space | Yes | OPC |
| Contact | Space | Yes | Selective sizing |
| Metal 1 | Mostly line | Yes | Improved OPC |
| Via 1 | Line /Space | Yes | No |
| Metal 2 | Mostly line | Yes | Improved OPC |
| Via 2 | Space | No | No |
| Metal 3 | Mostly line | Yes | Improved OPC |

**Table 4:** Shrink strategy

The shrink reduces SRAM cell size from 10.26$\mu m^2$ to 9.26$\mu m^2$. With the 5% shrink, there is a 15% increase approximately in sorted good die due to smaller die size. The shrink technology went to production four months after product tape-out thereby setting a new benchmark. All quality and reliability requirements were met, and products were synchronized just in time for the volume production ramp.

## Conclusion

At certification, P856 met its principal performance, yield, and density goals, while achieving an 85% equipment re-use rate. Within one year of certification, and with only low-cost changes, a further 5% shrink was implemented. With the same equipment set, re-optimization of the transistors combined with control enhancement has allowed an 18% improvement in gate delay, more than a half technology step.

## Acknowledgments

Many people contributed to the results discussed in this paper. Key people include Max Wei, Brian Johnson, Maurice DeCourcy, Domenic Pipitone, Karen Lubic, Brett Huff, Haiping Dun, Sam Hu, Bill Kavanaugh, Yung-Huei Lee, Wallace Lin, Steven Soss, Andrew Stack, John Mardinly, Li-Jia Ma, K.C. Patel, Tom Castro, Nevine Malek, Mike Maxim, Melinda Hoppe, and Ajay Chatterjee. In addition to those mentioned, we acknowledge the contributions of many others from the CTM, PTD, and virtual factory module and integration groups.

## References

[1] M. Bohr, Y. El-Mansy, "Technology for Advanced High-Performance Microprocessors," *IEEE Transactions on Electron Devices*, March 1998, pp. 620-625.

[2] S. Wolf, *Silicon Processing for the VLSI Era, Volume 3: The Submicron MOSFET.*

[3] M. Bohr, S.S. Ahmed, S.U. Ahmed, M. Bost, T. Ghani, J. Greason, R. Hainsey, C. Jan, P. Packan, S. Sivakumar, S. Thompson, J. Tsai, and S. Yang, "A High Performance 0.25um Logic Technology Optimized for 1.8V Operation," *IEDM Technical Digest*, 1996, pp. 847-850.

[4] S. Venkatesan, A. Gelatos, B. Smith, R. Islam, et.al., "A High Performance 1.8V, 0.20um CMOS Technology with Copper Metallization," *IEDM Technical Digest*, 1997, pp. 769-772.

[5] M. Chang, J. Ting, J. Shy, L. Chen, "A Highly Manufacturable 0.25 um Multiple-Vt Dual Gate Oxide CMOS Process for Logic/Embeded IC Foundry Technology." *1998 Symposium on VLSI Technology Digest*, pp. 150-151.

[6] J, Schutz, R. Wallace, "A 450MHz IA32 P6 Family Microprocessor," *ISSCC Technical Digest*, 1998, p. 236-237.

## Authors' Biographies

Adam Brand received his BSEE and his MSEE from the Massachusetts Institute of Technology in 1991. He joined Intel in 1991 and is currently working in the California Technology and Manufacturing 0.25µm Device Group. His interests include transistor

performance optimization, high voltage device development, and circuit modeling. His email address is adam.d.brand@intel.com .

Aravinda Haranahalli received an MS in Physics in 1976 and a Ph.D in Materials Engineering 1980 from the University of Florida. He joined Intel in 1984 and has held various management positions in technology, manufacturing, and business development. He currently manages interconnect technology development for 0.2µm. Before joining Intel he held technology positions at Texas Instruments and Fairchild. His current interests include technology, manufacturing, and business management. His email address is aravinda.r.haranahalli@intel.com .

Ning Hsieh received a Ph.D. in Materials Science from Northwestern University in 1979. He worked for various semiconductor companies including IBM, Fairchild, and DEC. He joined Intel in 1993 and has worked in CTM Technology Development since then. His work experience is mostly in process integration. He has published six external papers and has six patents. His email address is ning.hsieh@intel.com .

Yi-Ching Lin graduated from the University of California, Berkeley with a Ph.D. in EECS in 1981. Prior to joining Intel in 1987, he was with Texas Instruments and Monolithic Memories, Inc. He has been working in the area of process integration for microprocessor, Flash and EPROM memories. He had also worked on technology transfer from D2 to foreign foundries, including those located in Taiwan and Japan. His email address is yi-ching.lin@intel.com .

George E. Sery is an Intel Fellow and director of Device Technology Optimization in Intel's California Technology and Manufacturing group. Mr. Sery is currently responsible for directing process characterization, performance improvement, and capability enhancement for Intel's 0.25 micron CMOS logic technology. He received a B.S. and M.S. in electrical engineering from the University of Minnesota in 1976 and 1978 respectively. He joined Intel in 1978 as part of the SRAM Technology Development group. He has been involved with the development of NMOS and CMOS technologies for logic, SRAM, and Flash memory applications. For each technology, he has led the device physics team responsible for device development and process characterization. His email address is george.sery@intel.com .

Nicky Stenton received a M.S. in Materials Engineering from Lehigh University in 1982. She joined Intel in 1982 and most recently has been working on transistor process development in the California Technology and Manufacturing P856 Integration group. Her email address is nicky.stenton@intel.com .

Been-Jon Woo received a B.S. in Chemical Engineering from the National Taiwan University in 1975 and a Ph.D. from USC in 1979. She joined Intel in 1984 after working at Fairchild. She has worked in EPROM, Flash, and logic technology integration in the California Technology and Manufacturing group. She is currently the 0.25µm transistor integration manager. Her email address is been-jon.k.woo@intel.com .

Shahriar Ahmed joined Intel in 1985, initially as a interconnect device engineer working on Process 448. He subsequently was part of the team that developed P648 and coordinated the final transfer to high-volume manufacturing. Shahriar then moved on to be the device engineer for Intel's first bi-CMOS process. His next project was P856, which he developed together with a team from California Technology and Manufacturing. Currently he is in working on 0.18µm process development. His email address is shahriar.ahmed@intel.com

Mark T. Bohr joined Intel in 1978 after receiving a MSEE from the University of Illinois. He has been a member of the Portland Technology Development group since 1978 and has been responsible for process integration and device design on a variety of DRAM, SRAM, and logic technologies, including recently 0.35µm and 0.25µm logic technologies. He is an Intel Fellow and director of process architecture and integration. He is currently directing development activities on 0.18µm and 0.13µm logic technologies. His email address is mark.bohr@intel.com .

Scott Thompson joined Intel in 1992 after completing his Ph.D. under Professor C. T. Sah at the University of Florida on thin gate oxides. He has worked on transistor design and front-end process integration on Intel's 0.35, 0.25, and 0.18µm silicon process technology design for the Pentium® and the Pentium® II microprocessors. Scott is currently managing the development of Intel's 0.13µm transistor design. His email address is scott.thompson@intel.com .

Simon Yang received his B.S. in Electrical Engineering from the Shanghai University of Science and Technology (Shanghai, PRC). He then received his M.S. in Physics and a Ph.D. in Materials Engineering from the Rensselar Polytechnic Institute in New York. He joined Intel after graduating in 1987 and is currently leading transistor and yield improvement for Intel's 0.18µm logic technology. His email address is shi-ning.yang@intel.com.