

Optimization of Interconnect Geometry for High-Performance Microprocessors

The goals of the work presented in this paper were to estimate quantitatively the impact of interconnect technology parameters on the performance of high-end microprocessors and to use this information to optimize the interconnect geometry within the constraints imposed by the process. The 64-bit PA 8000 microprocessor was used as a test case.

by **Khalid Rahmat and Soo-Young Oh**

For the past two decades the driving force for integrated circuits has been scaling of both the devices and the interconnect. This has yielded faster and denser chips with ever increasing functionality. Today's high-performance microprocessors have 4 to 5 million logic transistors^{1,2,3} and operate in the frequency range of 200 to 250 MHz. To stay on the trend for CPU performance gains, the next-generation processors will likely have clock speeds of 400 to 500 MHz and have up to 10 million transistors for the logic implementation.⁴ The design and manufacture of systems that provide such performance will require careful attention to all process, device, and circuit issues.

In particular, the interconnect system for the next-generation CPUs needs to sustain the burden of connecting 10 million transistors with 5 to 6 metal layers over a 2-cm die with a clock cycle of 2 to 2.5 ns. Because the interconnect system affects the manufacturing yield, reliability, performance, and design complexity of a CPU, it cannot be optimized in isolation but rather must take into account all these aspects.

The goal of the work presented in this paper was twofold. First, we wanted a quantitative estimate of the impact of interconnect technology parameters on the performance of high-end microprocessors. Second, we wanted to use this information to optimize the interconnect geometry within the constraints imposed by the process. It is well-known that interconnect will become an increasingly significant factor in the design of future high-performance processors.^{4,5} Therefore, it is important to benchmark the current technology with realistic process parameters and circuit designs. The detailed layout and extracted data for the HP PA 8000 microprocessor¹ provided an ideal testbed to perform both of these tasks.

Two criteria were used for optimization: delay or rise time (they have very similar characteristics) and cross talk noise.^{6,7} The optimization was performed only for the global (block-to-block) routing consisting of metal layers 2, 3, and 4. Because the longest signal wires are usually at this level of routing, optimizing these layers has the most direct impact on system performance. For optimization, we only considered nets that had marginal timing based on the expected performance target for the processor.

Optimization Methodology

Our optimization approach is as follows:

1. Determine the timing-critical nets using a global static timing simulator.
2. From this set of nets select a sample that has significant global interconnect.
3. Extract the driver sizes and wire lengths for this interconnect-dominated subset of the global nets.
4. For these global interconnect dominated nets, determine the sensitivity to all interconnect parameters: metal thickness, width, and space and dielectric thickness.
5. For two of the most sensitive parameters perform two-dimensional optimization to determine the best design point for each metal layer considered (metal 2, 3, and 4).

The total wire capacitance and cross talk are calculated assuming a dense geometry with neighboring lines within the plane of the signal line and ground planes above and below, as shown in Fig. 1. The wire capacitance is obtained using the field solver RAPHAEL in 2D, and the cross talk and delay are obtained from SPICE simulations using the coupling and total capacitances generated from RAPHAEL. The cross-talk noise and delay were simulated in worst-case situations. For delay this means that the neighboring lines were switching simultaneously in the direction opposite to the signal line, and for cross talk the victim line was quiescent while the neighboring lines switched simultaneously in the same direction.

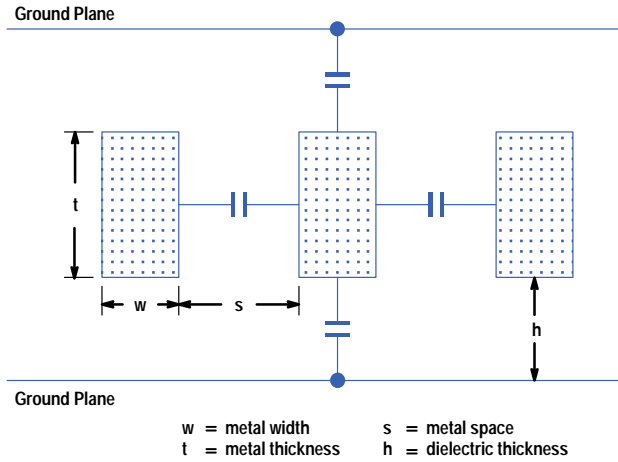


Fig. 1. Schematic of the interconnect parameters considered for optimization.

Results

Fig. 2 shows the distribution of metal lengths among the nets that failed the timing criterion at the target frequency. A significant number of critical nets have long metal 2 and metal 4 lines while metal 3 is usually much shorter. This is, of course, a direct consequence of the orthogonal channel routing scheme used by the HP-proprietary automatic router, PA-ROUTE. The other salient feature of Fig. 2 is the large number of nets that have 5 to 6 mm of metal 2. Similarly, there is a secondary peak of 4 to 5 mm for metal 4. Thus, an obvious target for optimization is wires that have this typical long length. We did not select the longest wires for optimization because they were not typical.

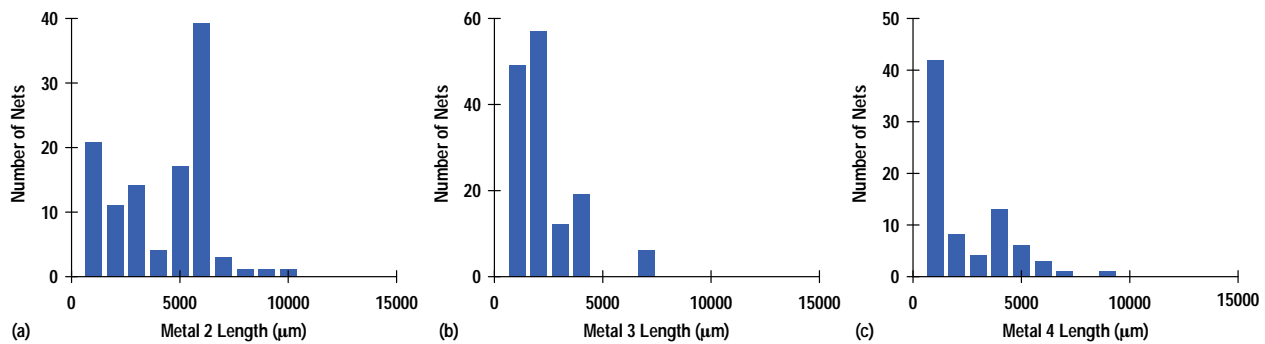


Fig. 2. Wire length distribution for critical nets. (a) Metal 2. (b) Metal 3. (c) Metal 4.

The drivers for all of the nets studied were approximately $44 \mu\text{m}$ wide (W_n = width of n-channel transistor = $44 \mu\text{m}$; W_p = width of p-channel transistor = $2W_n$). The drivers were large enough that the delays and rise times were determined primarily by the interconnect and not the driver size.

Metal 2 Optimization. Using the strategy outlined above, we considered net1, which has a length of 6.8 mm in metal 2, 3.6 mm in metal 3 and 0.9 mm in metal 4. Fig. 3 shows the impact on cross-talk noise and rise time (at the load) of varying only one interconnect parameter while keeping all others constant. The nominal values for the metal 2 parameters are:

- Metal width = $1.2 \mu\text{m}$
- Metal thickness = $0.7 \mu\text{m}$
- Metal space = $1.2 \mu\text{m}$
- Dielectric thickness = $0.95 \mu\text{m}$.

It is clear from Fig. 3 that the two metal 2 parameters to which the rise time is most sensitive are the metal width and thickness. Thus, the next step in our optimization strategy varies these two parameters simultaneously to generate a two-dimensional contour plot of rise time and cross-talk curves, as shown in Fig. 4. When we vary the metal 2 width we keep the pitch constant, which is a more realistic situation than allowing the pitch to vary, too. To reduce the rise time, either the width or the thickness (or both) of metal 2 lines must be increased. For example, going from width = $1.2 \mu\text{m}$, space = $1.2 \mu\text{m}$, and thickness = $0.7 \mu\text{m}$ to width = $1.4 \mu\text{m}$, space = $1.0 \mu\text{m}$, and thickness = $0.8 \mu\text{m}$ will yield a reduction in rise time of almost 150 ps (14%). From Fig. 4, it can be estimated that this change will produce an increase in worst-case cross-talk noise of less than 100 mV to about 0.52V which may still be acceptable.

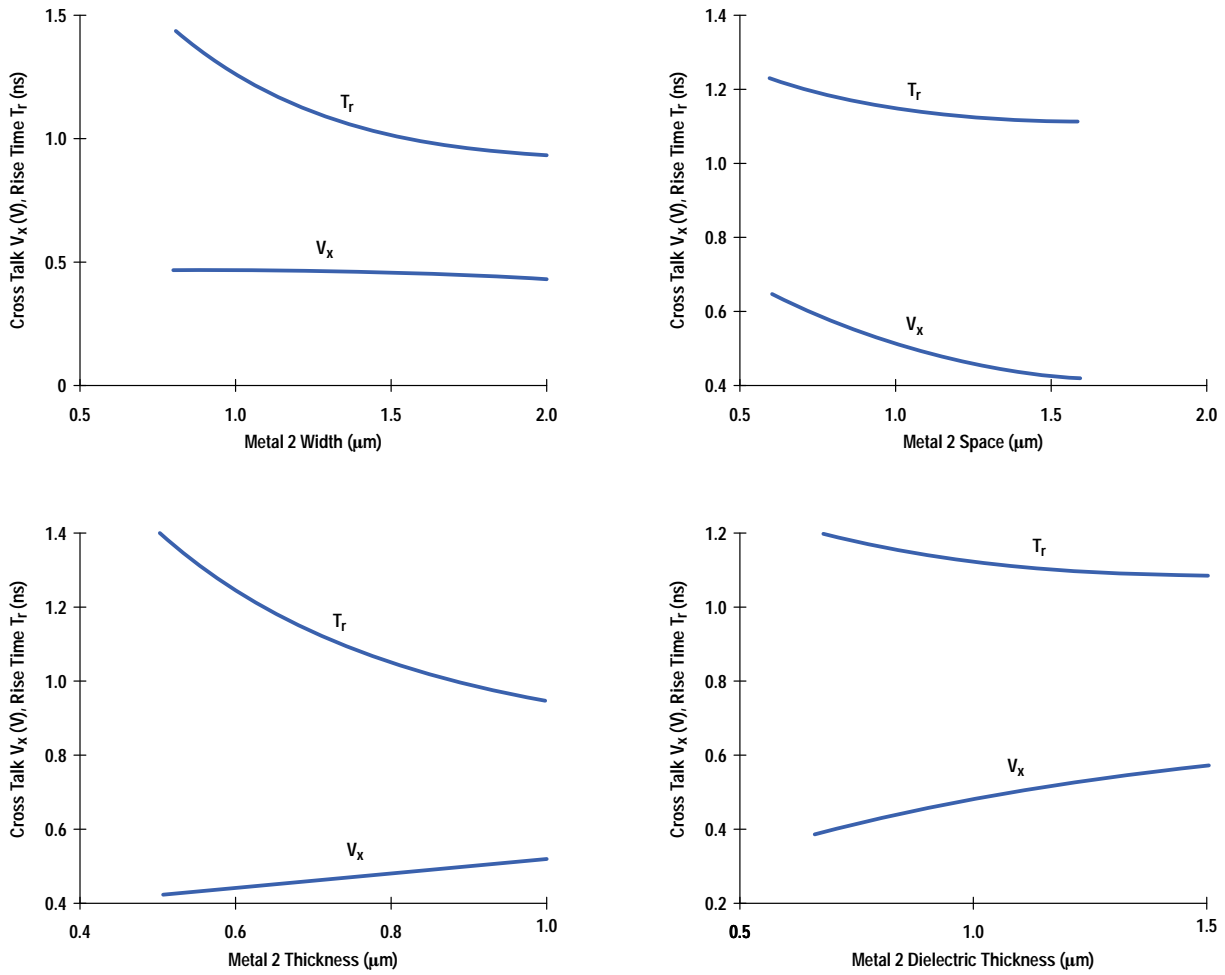


Fig. 3. Sensitivity of net1 rise time T_r and cross talk V_x to metal 2 parameters.

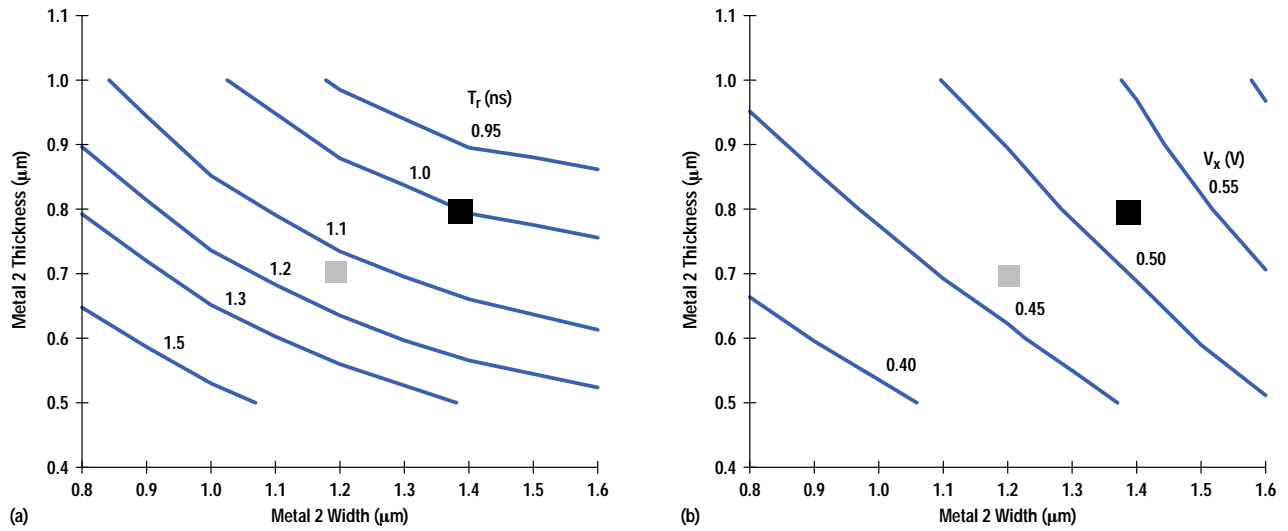


Fig. 4. Contours of constant rise time (a) and cross talk (b) for the metal 2 dominated net1. The nominal design point is indicated by the gray square while the possible optimized point is shown in black. The pitch (width + space) was kept constant at 2.4 μm .

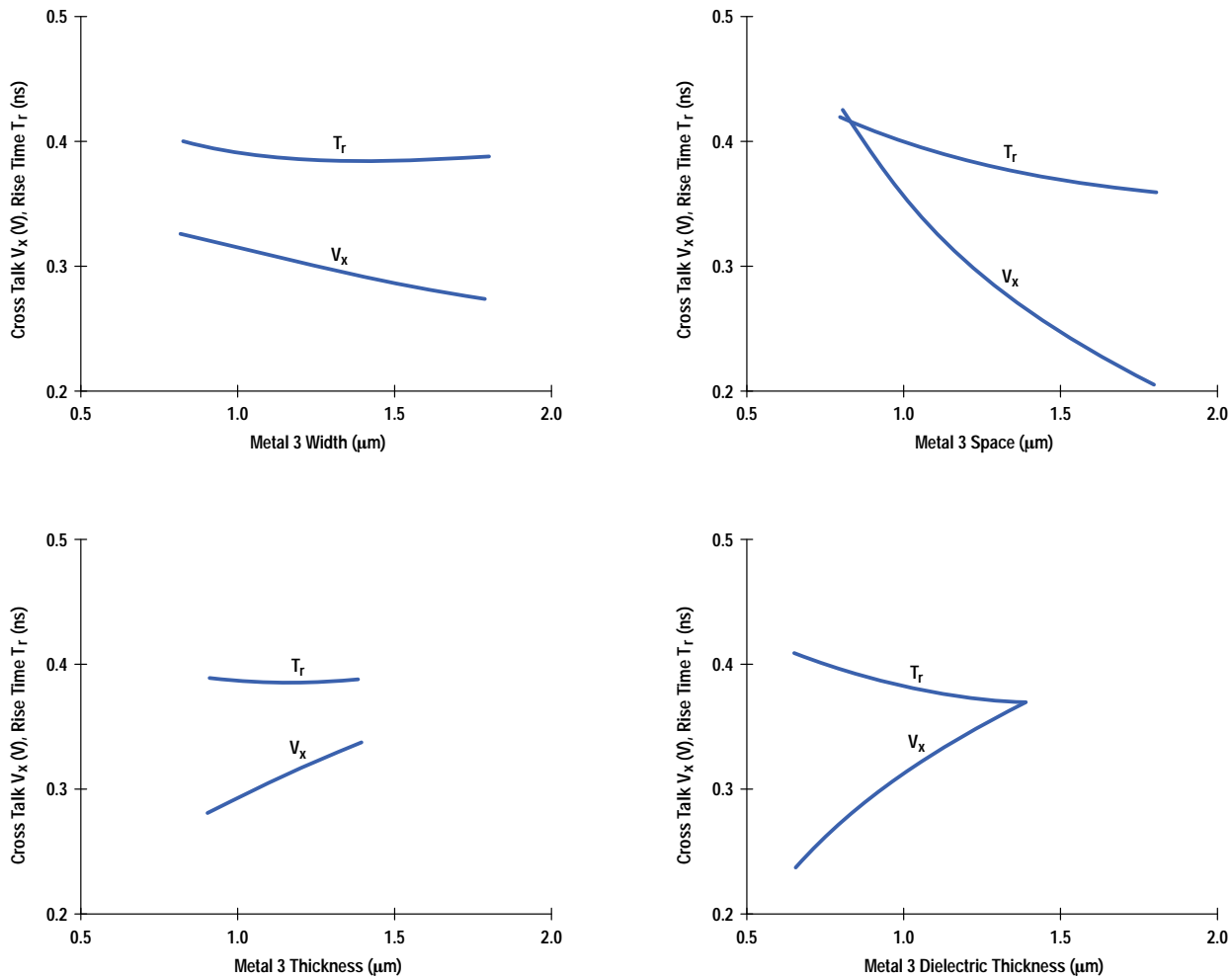


Fig. 5. Sensitivity of net2 rise time T_r and cross talk V_x to metal 3 parameters.

Metal 3 Optimization. We proceed here just as in the case of metal 2 by first choosing a net that has significant length in metal 3. In this case this is net2 which has 0.7 mm of metal 2, 3.3 mm of metal 3, and 2.3 mm of metal 4. The nominal metal 3 parameters are:

- Metal width = 1.2 μm
- Metal thickness = 1.1 μm
- Metal space = 1.2 μm
- Dielectric thickness = 0.95 μm .

Because of the relatively short length of the metal 3 lines, the rise time is not very sensitive to metal 3 parameters, as shown in Fig. 5. Of the four parameters, the sensitivity is greatest to the metal 3 spacing and the thickness of the dielectric. As before, we choose these two parameters (while keeping the metal 3 pitch constant) to perform a two dimensional optimization as shown in Fig. 6. The rise time can be improved by using a slightly narrower width (width = 1.0 μm , space = 1.4 μm), but the absolute improvement is small (20 ps).

Metal 4 Optimization. For metal 4 we used net3, which has 1.3 mm of metal 3 and 6.2 mm of metal 4. The nominal metal 4 dimensions are:

- Metal width = 1.6 μm
- Metal thickness = 1.2 μm
- Metal space = 2.4 μm
- Dielectric thickness = 1.05 μm .

For metal 4, the rise time is most sensitive to the metal spacing and the dielectric thickness as seen in Fig. 7. However, for the nominal metal 4 spacing of 2.4 μm , the rise time is a fairly smooth function of spacing, so the spacing between metal 4 lines could be reduced (while keeping the width constant) by up to 0.4 μm without increasing the delay significantly. Thus, the pitch could be reduced without adversely impacting the delay. If the goal is to reduce delays, then two-dimensional

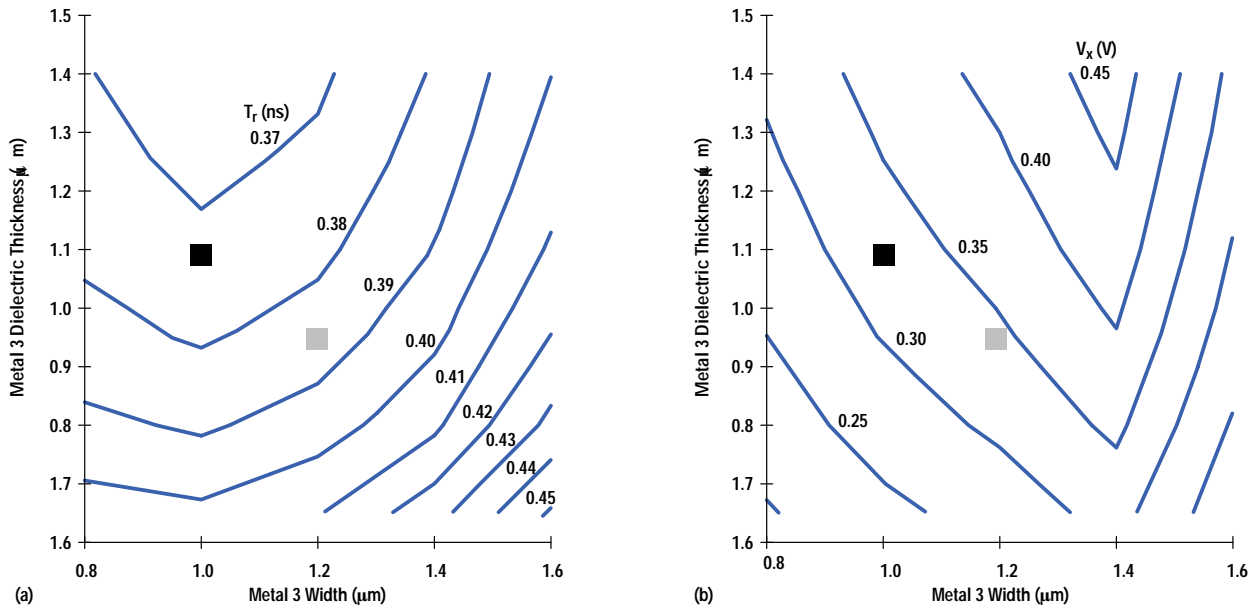


Fig. 6. Contours of constant rise time (a) and cross talk (b) for the metal 3 dominated net2. The nominal design point is indicated by the gray square while the possible optimized point is shown in black. The pitch (width + space) was kept constant at $2.4 \mu\text{m}$.

optimization reveals that the rise time can only be improved by using a narrower metal 4 and a thicker dielectric as shown in Fig. 8. Cross talk is a much smaller concern here because of the wide spacing between metal lines.

Conclusion

The main results of this study are as follows:

- Global routing in metal 2 wires causes significant delays because of the high resistance associated with this layer. The delays can be reduced by increasing the metal thickness and by increasing the metal 2 width even if the metal 2 pitch is kept constant.
- Typical lengths of global interconnect in metal 3 are relatively short. Therefore, delays incurred because of routing in metal 3 are minimal. Hence, the optimization for metal 3 should be primarily based on performance within functional units.
- For metal 4, the delays are primarily affected by the capacitance of the lines rather than the resistance, since the thickness is substantially greater than for metal 2. Therefore, shorter delays require thicker dielectrics and narrower metal lines when constant pitch is assumed.

In our study the gains in performance of up to 200 ps can be obtained by wire optimization within reasonable process parameters. A simple example will illustrate the impact of an improvement of this magnitude on the cycle time of a microprocessor. Consider a hypothetical processor fabricated in a manufacturing process that yields a nominal cycle time of 4.0 ns (250 MHz), with a standard deviation, assuming a normal distribution around the mean of 0.3 ns. In this case, only 7.6% of the devices will have a speed of 280 MHz or greater. If the nominal design is changed to reduce the cycle time by just 200 ps, then the fraction of the devices yielding above 280 MHz will increase threefold to 22.2%. Thus, a 5.0% reduction in the nominal cycle time produces a 200% increase in the number of the fastest devices, which command a premium because of their high performance.

Acknowledgments

The PA 8000 data to which the optimization method was applied was provided by Clay McDonald and Mark Hammer, whose cooperation and help are gratefully acknowledged.

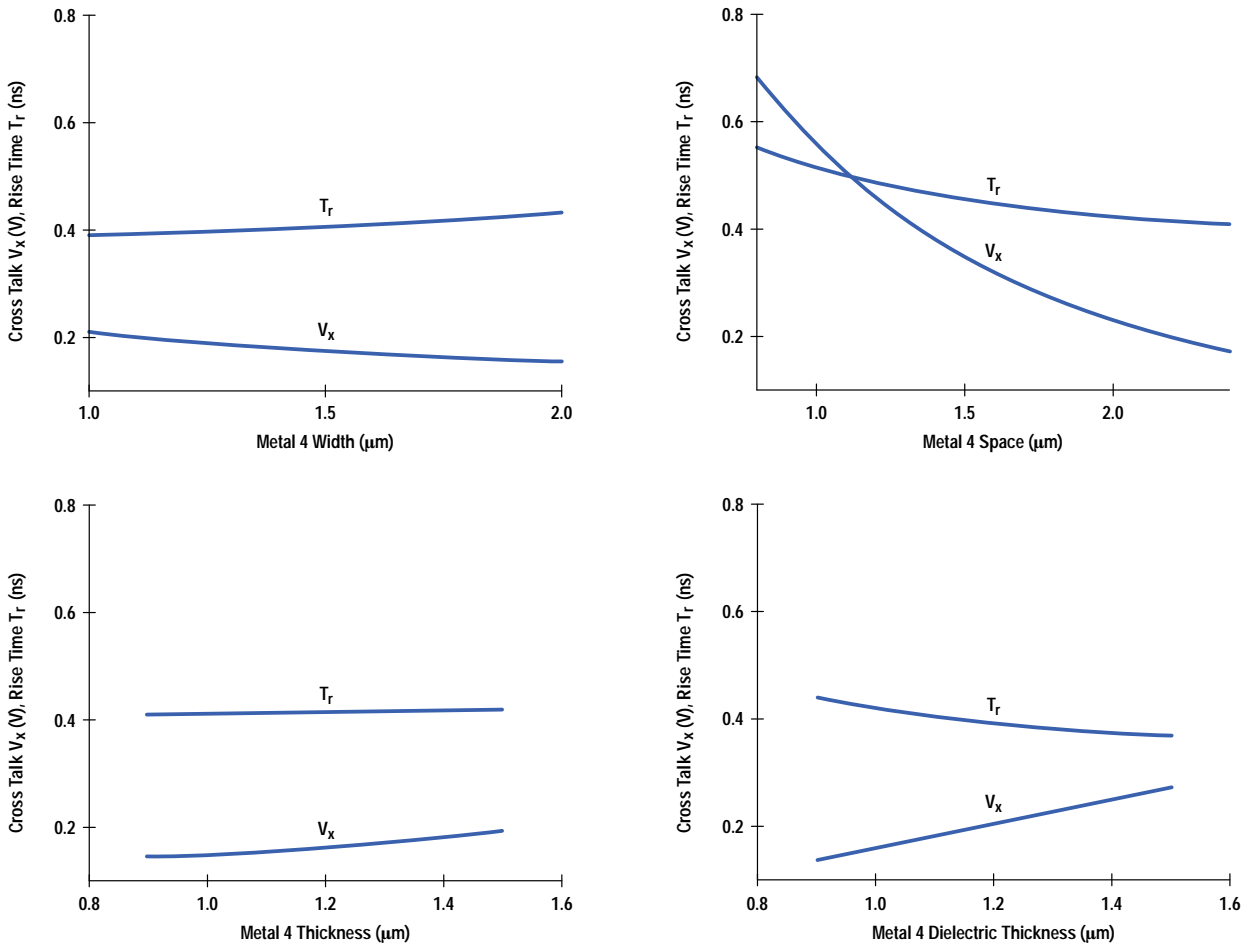


Fig. 7. Sensitivity of net3 rise time T_r and cross talk V_x to metal 4 parameters.

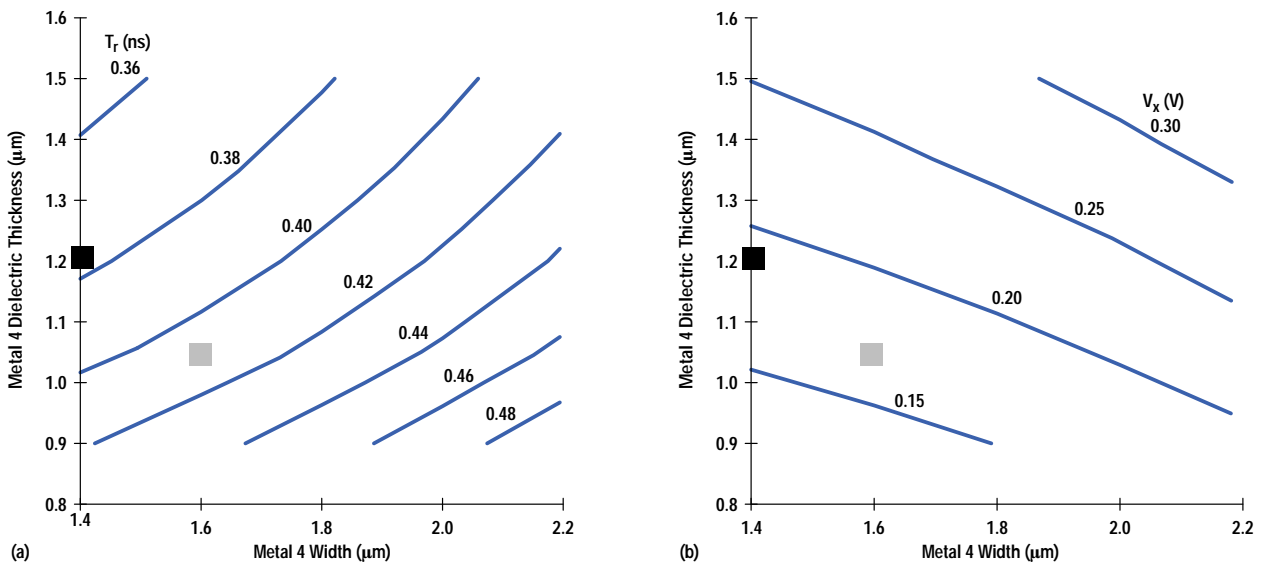


Fig. 8. Contours of constant rise time (a) and cross talk (b) for the metal 4 dominated net3. The nominal design point is indicated by the gray square while the possible optimized point is shown in black. The pitch (width + space) was kept constant at 4.0 μm .

References

1. J. Lotz, G. Lesartre, S. Naffziger, and D. Kopp, "A quad-issue out-of-order RISC CPU," *Technical Digest, International Solid-State Circuits Conference*, 1996, p. 210.
2. M. Tremblay, "UltraSparc-I: A 64-bit superscalar processor with multimedia support," *Technical Digest, Hot Chips Conference*, 1995, p. 207.
3. R.P. Colwell and R.L. Steck, "A 0.6- μm BiCMOS processor with dynamic execution," *Technical Digest, International Solid-State Circuits Conference*, 1995, p. 176.
4. *The National Technology Roadmap for Semiconductors (NTRS) 1994*, Semiconductor Industry Association, 1994.
5. S-Y. Oh and K-J. Chang, "2001 needs for multilevel interconnect technology," *IEEE Circuits and Devices Magazine*, January 1995, p. 16.
6. P. Raje, "A Framework for Insight into the Impact of Interconnect on 0.35- μm VLSI Performance," *Hewlett-Packard Journal*, Vol. 46, no. 1, February 1995, p. 97.
7. K. Rahmat, O.S. Nakagawa, S-Y. Oh, J. Moll, and W.T. Lynch, "A scaling scheme for interconnect in deep-submicron processes," *Technical Digest, International Electron Devices Meeting (IEDM)*, 1995, p. 245.

▶ [Go to Next Article](#)

▶ [Go to Journal Home Page](#)