Optimisation of full-custom logic cells using response surface methodology

M.V. Scotti, Z. Malik, P.Y.K. Cheung and J. Nelder

The authors demonstrate the significant advantages of response surface methodology for the optimisation of pass-transistor logic circuits with a large number of design variables. It is shown how the 'curse of dimensionality' in this type of optimisation problem can be effectively dealt with by using methods from the set of techniques referred to as 'design of experiments'.

Introduction: The design of digital circuits, especially for full-custom cells, is often a time consuming and laborious process relying on the intuition and heuristic knowledge of the designer. Optimising transistor sizes is the main key in meeting the various design specifications such as time-delay, area and power. The problem with this type of design optimisation is the large number of design variables leading to an explosion in the design space. Various solutions to these problems have been proposed which are based on the use of either static timing models [1], dynamic timing models [2] or a combination of the two [3]. Static timing models, while efficient in computation, do not take all transition states into account, resulting in low accuracy. Dynamic timing models, on the other hand, require the consideration of all possible input combinations and therefore suffer from the dimensionality problem.

We propose to use a set of techniques known as design of experiments (DOE) [4] and response surface methods [5] to effectively deal with the optimisation problem while still being able to carry out the analysis using dynamic timing. The DOE techniques, particularly fractional factorial methods [4], can greatly reduce the number of combinations of design variables that need to be tried (thereby ameliorating the dimensionality problem). For this given reduced set DOE, techniques greatly increase the information per simulation run [6]. Response surface models (RSMs) effectively replace the slow circuit-level simulator and allow instant calculation of the simulated responses.

Optimisation problem: We show how a class of optimisation problems in circuit design can be effectively and efficiently solved using the following iterative procedure:

(i) Select an initial region of exploration in the design space.
(ii) Use DOE [4] to reduce the number of simulations and carry out circuit-level simulation (SPICE).
(iii) Build RSMs of the simulated responses.
(iv) Use the RSMs for instantaneous evaluation of a cost function and hence determine the predicted optimal direction.
(v) Follow the predicted direction, checking progress with a simulation and stopping at the lowest point.
(vi) Select a new sampling region around this lowest point and iterate from step (ii) until the termination criteria are met.

As can be seen, the procedure is very general and can be applied to a wide variety of design problems. There are several classes of RSM that may be chosen: ordinary polynomial, inverse polynomial, log-log, exponential transformations (see [7] for further details). The simplest RSMs are just ordinary regression models that assume that responses and performances can be related linearly by, essentially, two terms: a systematic component and an error component. The elements of the error component are assumed to be independent while having the same normal distribution. Generalised linear models (GLMs) [7] are a subset of RSMs and extend the ordinary regression framework to situations where the errors in the data do not follow a normal distribution, and/or where a transformation needs to be applied to the systematic component; this allows the fitting of nonlinear terms in the response variables. In our case, for simulated data, the error term is due entirely to the lack of fit of the model.

In general it is not feasible to hand-craft models for each response at each new region of exploration. We have created a general automated fitting algorithm to fit GLMs.

Example: SRPL circuit: The above techniques are illustrated by applying them to a specific circuit: a 1 bit swing restored pass-transistor logic (SRPL) [8] adder circuit (see Fig. 1). The circuit consists of input and output drivers for each signal, an NMOS transistor logic network and a cross-coupled inverter latch for signal restoration. Pass-transistor circuits are difficult to analyse due to the bi-directional signal flow and complex charge sharing within the NMOS logic network combined with

nonlinear behaviour of the cross-coupled inverter latch. The objective here is to optimise the transistor sizes for the shortest worst-case time-delay.
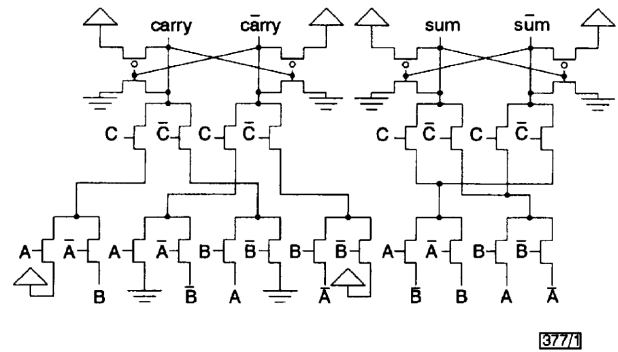

Fig. 1 *Full-adder circuit in pass-transistor logic (SRPL)*

No preference was given to any particular signal path. In total there are 12 independent design variables (transistor sizes). Exploration around a nominal point by trying all combinations of the design variables at just two values each (two-level designs) would require $2^{12} = 4096$ simulations. Using DOE reduces [4] this to just 16 simulations at a time. The results of these runs are used to build RSMs of the time-delays which are used for optimisation (effectively replacing the simulation). From the RSMs a direction for minimising the worst-case time delay is found and a binary search carried out along this direction from the centre to outside the original sampling region. This procedure is then iterated as described in the preceding Section.

We have discovered that a class of models of the log-log form best suit the problem of modelling the time delays. The log-log model requires that the time delays and transistor sizes are always positive, as they are in the real physical system. An ordinary polynomial might extrapolate to non-physical negative values and hence not predict as well as the log-log model. Also we expect the time-delays to behave with some kind of power law. For a two-level sampling design we used bilinear log-log models of the form

$$\log f = \beta_0 + \sum_{i=1}^{12} \beta_i \log x_i + \sum_{i>j}^{12} \beta_{ij} \log x_i \log x_j \qquad (1)$$

The β coefficients were then calculated to best fit the data using the GLM algorithm [7]. Each time-delay, f, was represented by the above equation and the x variables are the 12 different transistor sizes.

To minimise the worst-case time-delay we created a smooth cost function, $F = \Sigma_i f_i^p$, which was equal to the sum of powers of the individual response time-delays. Taking a suitably large value of p effectively gave an L_∞ norm of the fs, enabling a gradient descent vector to be calculated.

Results: The results of the optimisation procedure using response surface methods are shown in Fig. 2. This method took five sets of designs (dots) together with five gradient descent trajectories (solid lines) to reach a point satisfying the termination criteria. This involved a total of 127 simulation runs and gave a minimum time-delay of 0.5148ns, where the initial nominal point was 1.311ns. A simulation run takes 11.79s on a SUN-ULTRA-S workstation. The termination criteria are guided by the silicon process technology: the circuit is realised using a 0.6μm process technology which enables gate-widths from 0.8μm in 0.4μm steps to be obtained. We also restrict the maximum transistor sizes to 16μm for NMOS transistors and 32μm for PMOS transistors. The RSM optimisation terminates when the minimum point reached by steepest descent has not managed to move out of the original sampling region.

For comparison the Nelder-Mead downhill simplex algorithm [9], a robust and well tried method of unconstrained nonlinear optimisation, gave a minimum of 0.5299ns after 305 runs. This minimum was matched in just 100 simulation runs (four iterations of the optimisation procedure) by the RSM technique. Also these results can be contrasted with those from a manual optimisation approach (with transistor sizing based on [8]), which gave a minimum of 0.7431ns.

The results show the superiority of our RSM technique for optimisation over the other methods. The downhill simplex method takes considerably more runs and never reaches the minimum found by the RSM method; the RSM method is very quick at the beginning of the optimisa-

tion. The RSM method basically 'sniffs' around its surroundings and then decides on the best route for minimisation.
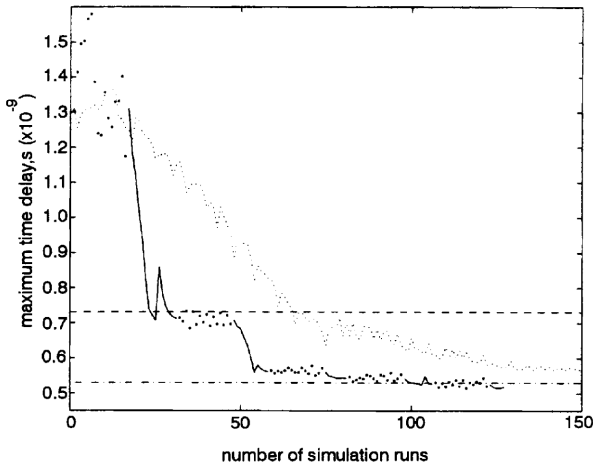

Fig. 2 *RSM against Nelder-Mead simplex optimisation*

———— predicted optimum trajectories
● DOE sample points
·········· Nelder-Mead simplex (NMS)
· – · – NMS minimum reached after 305 simulations
– – – – manual optimisation

Discussion: There are many advantages to using RSM techniques for optimisation besides the obvious one of significant increases in the speed of optimisation. Physical insight can be used to obtain appropriate model classes which would lead to better results. The response surface methods also allow constrained optimisation for multiple objectives (i.e. time-delay, power-dissipation and area requirements). They also allow an insight to be gained into the physical systems being studied rather than just being treated as a black-box for the optimisation routine. The methods themselves are independent of the simulation time and are therefore increasingly effective for more complex systems.

The idea of using DOE combined with GLMs (within the framework of response surface methods) as a way to efficiently exploit expensive simulations for the purpose of optimisation is relatively new and is applicable to many other fields of engineering. These techniques could lead to significant time and cost savings as well as allowing full design exploration rather than just design confirmation.

M.V. Scotti, Z. Malik and P.Y.K. Cheung (*Department of Electrical and Electronic Engineering, Imperial College, London SW7 2BT, United Kingdom*)

J. Nelder (*Department of Mathematics, Imperial College, London SW7 2BT, United Kingdom*)

References

1   FISHBURN, J.P., and DUNLOP, A.E.: 'Tilos: A posynomial programming approach to transistor sizing'. Proc. Int. Conf. Computer Aided Design, November 1985,
2   CONN, A.R., COULMAN, P.K., HARING, R.A., MORRILL, G.L., and VISWESWARIAH, C.: 'Optimisation of custom mos circuits by transistor sizing'. IEEE Int. Conf. Computer Aided Design, November 1996, pp. 174–180
3   SHYU, J.M., SANGIOVANNI-VINCENTELLI, A., FISHBURN, J.P., and DUNLOP, A.E.: 'Optmisation-based transistor sizing', *IEEE J. Solid-State Circuits*, 1988, **23**, (2), pp. 400–409
4   BOX, C.E.P., HUNTER, W.G., HUNTER, J.S.: 'Statistics for experimenters' (John Wiley & Sons, 1987)
5   BOX, C.E.P., and DRAPER, N.R.: 'Empirical model-building and response surfaces' (John Wiley & Sons, 1987)
6   MALIK, Z., SU, H., and NELDER, J.: 'Informative experimental design for electronic circuits', *Quality Reliability Eng. Int.,* 1998, **14**, pp. 177–186
7   NELDER, J.A., and WEDDERBURN, R.W.M.: 'Generalized linear models', *J. R. Statist. Soc. A,* 1972, **135**, pp. 370–384
8   PARAMESWAR, A., HARA, H., and SAKURAI, T.: 'A swing restored pass-transistor logic-based multiply and accumulate circuit for multimedia applications', *IEEE J. Solid-State Circuits,* 1996, **31**, (6), pp. 804–809
9   NELDER, J.A., and MEAD, R.: 'A simplex method for function minimization', *Comput. J.,* 1965, **7**, pp. 308–313