Delivering Global DC Convergence for Large Mixed-Signal Circuits via Homotopy/Continuation Methods

Jaijeet Roychowdhury Robert Melville

ABSTRACT

Homotopy/continuation methods are attractive for finding DC operating points of circuits because they offer theoretical guarantees of global convergence. Existing homotopy approaches for circuits are, however, often ineffective for large mixed-signal applications. In this paper, we describe a robust homotopy technique that is effective in solving large MOS-based mixed-signal circuits. We demonstrate how certain common circuit structures involving turning-point nesting can lead to extreme inefficiency, or failure, of conventional probability-one homotopy methods. We also find that such situations can lead to numerical ill-conditioning and homotopy paths that fold back upon themselves, leading to algorithm failure. Our new homotopy model for MOS devices, dubbed ATANSH, features *decoupled continuation parameters* that are instrumental in avoiding these problems. ATANSH-based homotopy methods in production use have led to the routine solution of large, previously hard-to-solve industrial circuits, several examples of which are presented.

I. INTRODUCTION

Finding DC operating points of nonlinear circuits is a fundamental task in circuit simulation. The operating point is not only essential as a first basic check of circuit operation, but is also a prerequisite for further analyses. Small-signal AC analysis, noise analysis, and transient analysis [14, 23] all rely on a prior DC operating point having been calculated; furthermore, the operating point is also useful for steady-state and envelope analyses. As a result, the problem of finding DC operating points has long attracted the attention of CAD researchers and practitioners.

Mathematically speaking, finding a DC operating point involves the numerical solution of a potentially large system of nonlinear equations. The standard approach for this problem is the venerable Newton-Raphson (NR) algorithm (*e.g.*, [17]), which has been the mainstay of virtually all circuit simulators available today. While NR has been, by any standard, extremely successful as a general-purpose algorithm for the circuit DC operating point problem, it does not offer any *guarantee* of success. Indeed, "DC convergence failure" (*i.e.*, failure of NR to find an operating point) is a common problem faced by circuit designers across a broad swath of circuit types and functionalities. For certain classes of circuits, such as the large MOS circuits involved in ADCs and DSP systems, the problem can become acute enough to constitute a *serious bottleneck* in the overall process of designing mixed-signal and analog chips. In current analog and mixed-signal design methodologies, in fact, it is not uncommon for considerable time and ingenuity to be spent in working around NR failure. Typical workarounds include splitting larger circuits into smaller sub-circuits whose DC operating points are found individually and then combined; 'tweaking' circuits (such as by adding additional resistors); and a variety of 'stepping' methods (see Section II). Usually, the design of each circuit involves many separate operating point calculations, as the circuit is modified and refined. In the event of persistent DC convergence problems, it is not uncommon for designers ignore regions of the design space, because of the time and effort needed to obtain convergence manually [3]. For these reasons, there has been considerable interest in DC algorithms that work much more reliably than NR for 'problem' circuits in practice, even at the expense of greater computational time.

In this paper, we present novel adaptations and applications of a class of techniques, namely *homotopy (or continuation)*, to solve the circuit DC operating point problem far more robustly than NR. Because it can be proven to be globally convergent (*i.e.*, always successful in finding a solution), homotopy has offered the hope of solving for circuit DC operating points reliably, and has been applied previously to this problem (see Section II). However, our experience has shown that existing homotopy-based approaches often fail to converge when applied to large MOS circuits, despite their theoretical guarantees of convergence. Indeed, some existing homotopy approaches appear to perform significantly worse than straightforward Newton-Raphson or stepping methods for large circuits. Some of these issues have been noted previously (*e.g.*, [2, 27, 29]), and have contributed to uncertainty about the usefulness of homotopy for circuit simulation applications.

The approach followed in this work is 1) to first investigate why straightforward applications of homotopy do not deliver in practice the guaranteed convergence promised by theory, and then 2) to devise a homotopy-based method that does in fact deliver global convergence robust enough for general-purpose industrial applications. Towards these ends, we first identify mechanisms responsible for the failure of 'standard' homotopies, and using the understanding thus gained, devise new homotopy methods that circumvent failure. We show that an important failure mechanism is triggered by circuits with features of nested flip-flop-like structures, which can lead to computation that grows exponentially with circuit size. This mechanism exacerbates another failure mechanism, numerical ill-conditioning, which traditional homotopies suffer from when applied to large MOS circuits. To alleviate these problems, we present a new MOS homotopy (named ATANSH) that has proven robust in practice: it has been successful in finding the

Jaijeet Roychowdhury is with the University of Minnesota, Minneapolis, USA.

Robert Melville is with Agere Systems, Allentown, PA, USA.

DC operating points of virtually all 'problem' circuits it has encountered, and has been used in production within AT&T/Lucent Microelectronics since 1995. Existing device models, which represent considerable investment for in-house simulators, can be incorporated into an ATANSH-homotopy-based DC convergence algorithm.

The remainder of the paper is organized as follows. Previous work on finding DC operating points, including existing homotopy methods, is reviewed briefly in Section II. An overview of the basic principles of arc-length continuation is provided in Section III. In Section IV, empirical failure mechanisms of existing homotopy techniques are examined, and a simple nested flip-flop model developed to explain the fundamental mechanism behind these failures. In Section V, the ATANSH homotopy, developed to circumvent these failures, is described. Results on large and conventionally difficult-to-solve industrial circuits are presented in Section VI.

II. Previous Work

Possibly the best-known and most commonly used method for finding an operating point is the locally quadratically convergent Newton-Raphson (NR) algorithm (*e.g.*, [17]), which can have significant convergence problems if the starting guess is not sufficiently close. Modifications of NR that improve convergence, such as damped NR [17], have been been generally successful in make any significant practical difference for circuit problems. In practice, ad-hoc techniques, such as limiting the Newton step [14, 23], are widely used to improve convergence, with moderate success. It should be noted that the locally-quadratic convergence enjoyed by Newton-Raphson makes it valuable as a building-block for other algorithms, such as the pseudo-transient and stepping techniques, and also for the homotopy techniques of this paper.

The so-called pseudo-transient technique [24, 25], often used when Newton-Raphson fails, consists of performing a transient analysis while slowly ramping the circuit's sources from zero to their DC values. It relies on the dynamic elements of the circuit to bring the solution smoothly from zero to a steady-state at the DC solution. The approach fails when the DC solution is not absolutely transient-stable (*i.e.*, the differential equation solutions do not converge asymptotically to a time-invariant solution, as in oscillators); even when a stable DC solution exists, slowly-damped transient responses can make the method very inefficient. A similar approach, but one that does not rely on the dynamics of the circuit, is stepping. Stepping methods change a simple circuit with known solution to the one to be solved for in small increments, while tracking the solution using Newton-Raphson's local convergence properties. These methods fail when they encounter turning points or folds, i.e., points at which small changes to the circuit result in discontinuous changes to the solution, as in circuits with hysteresis.

In contrast to the above techniques, homotopy methods (*e.g.*, [1,22]) can be *proven* to be globally convergent on broad classes of problems. The application of homotopy to circuits is not new. In [11, 18], Melville, Trajković et al. applied natural and artificial-parameter homotopy methods to small bipolar circuits, and investigated conditions for their global convergence. Multi-parameter homotopy methods have been investigated [4], but their global convergence properties remain an open issue. Homotopy methods have also been applied to discovering more than one (potentially all) operating points of a nonlinear circuit [6, 8]. More recently, homotopy methods have been applied by several researchers to solving steady-state problems in circuits [2, 5, 7, 28]. Finally, initial results of the present work were reported in [19].

In spite of a valuable body of existing work, our experience has been that achieving provable global convergence via homotopy, especially on large practical circuits, is fraught with surprising problems and failure mechanisms. It is these aspects that we identify and rectify in this work.

III. Arclength-continuation homotopy methods: background

In this section, we review the basic concepts of homotopy and continuation methods; more detailed expositions can be found in, *e.g.*, [1, 29]. The principle of continuation is similar to that of source or GMIN stepping (also known as *monotonic continuation*), familiar to users of circuit simulators such as SPICE2G6 or SPICE3 [14, 23]. In stepping methods, the circuit equations are first modified by means of a *continuation parameter*. In source stepping, for example, the parameter is a multiplier for the independent sources; in GMIN stepping, it is a conductance GMIN from each node to ground. The parameter is first set to a value at which the circuit becomes easy to solve or its solution becomes known. The parameter is then changed back slowly to a value at which the original circuit is retrieved and simultaneously, the solution of the changing circuit is followed. The underlying hypothesis is that small changes in the parameter cause small changes to the circuit and its solution, hence the new solution is easy to obtain using numerical techniques with local convergence properties (e.g., the Newton-Raphson method). At first sight, it may appear natural to expect this hypothesis to hold for circuits described by equations that are smooth (*i.e.*, continuous and differentiable). However, this hypothesis is not valid for many circuits, especially those characterized by positive feedback leading to multiple operating points. A simple example is the Schmitt trigger circuit [16], where stepping can fail at critical values of the continuation parameter because the state of the circuit can change abruptly from low to high (and vice-versa) for even the slightest monotonic change in the parameter. The phenomenon is illustrated in Figure 1. As the supply voltage V_{cc} is stepped upward from 0V, the output of the circuit changes smoothly; until $V_{cc} \approx 4.5$V, where there is a large discontinuity in the output as V_{cc} is increased. Such points where monotonic increase or decrease of the continuation parameter leads to abrupt jumps in the solution are termed *turning points* or *folds*. Many practical feedback systems composed of smoothly-behaved components exhibit turning points that can cause stepping algorithms to fail.

Homotopy, which also features a continuation parameter (typically denoted λ) is in principle similar to stepping methods, but with important differences that imbue it with much more powerful solution properties. The continuation parameter in homotopy may or may not have a simple physical interpretation. It is in the treatment of turning points that homotopy or continuation differs from stepping. By "detecting" these points and changing

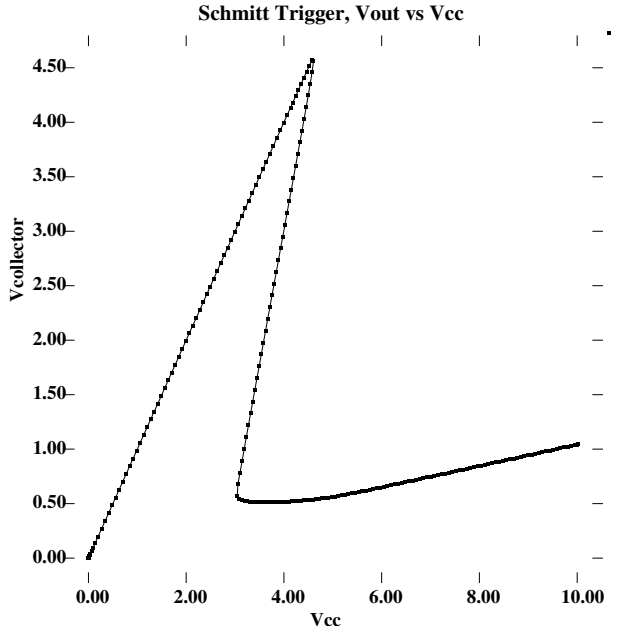


Fig. 1. Schmitt trigger circuit: the V_{out} vs V_{cc} characteristic exhibits hysteresis, leading to failure of simple stepping schemes.

the direction in which the continuation parameter is being incremented, it maintains continuity in the solution path and eventually reaches the desired solution of the original circuit. For the Schmitt trigger characteristic of Figure 1, this corresponds to *reducing* V_{cc} after the turning point at $V_{cc} \approx 4.5$V is reached, taking care to follow the central section of the characteristic and not backtrack onto the initial section already covered. Another turning point is reached at $V_{cc} \approx 3$V, after which V_{cc} is increased again and the lower right section of the characteristic followed.

Different continuation algorithms achieve the negotiation of turning points by different means. Parameter switching [15], a method closely related to stepping, redefines the continuation parameter when it detects a problem with monotonically changing the original continuation parameter. It relies on the fact that *some* member of the circuit's solution vector (e.g., a node voltage) will continue changing in the same direction after the turning point as before, and starts stepping that member as the new continuation parameter. Such switching is performed as many times as required.

Arclength continuation, on the other hand, can negotiate turning points automatically without their explicit detection. Its power stems from that it does not treat the continuation parameter λ differently from the unknowns of the circuit being solved for, but treats it as another unknown whose next value on the curve it determines. For the Schmitt trigger circuit of Figure 1, the process corresponds to 'walking' along the characteristic without paying special heed to curvature and folds or treating V_{cc} as special. This is achieved mathematically by solving a special differential equation that produces as output a sequence of values of λ (in general *not* monotonically increasing) together with solutions of the circuit at these values of λ. Furthermore it can be proved without restrictive assumptions that at some point in this sequence, $\lambda = 1$ will be reached; in other words, the solution at that point will be the desired solution of the original circuit. The remainder of this section contains an outline of the technique, with further details provided in Appendix A.

Any nonlinear circuit's equations can be put in the general form

$$\bar{g}(\bar{x}) = \bar{0}. \tag{1}$$

The first step in arclength continuation is to modify this system to add the continuation parameter λ, which by convention is allowed values in the interval $[0,1]$. The new system is denoted

$$\bar{f}(\bar{x},\lambda) = \bar{0}, \tag{2}$$

and is constructed so as to reduce to the original system at $\lambda = 1$ by convention, i.e., $\bar{f}(\bar{x},1) \equiv \bar{g}(\bar{x})$. Furthermore, the system $\bar{f}(\bar{x},0) = \bar{0}$ is constructed so as be easy to solve by traditional methods.

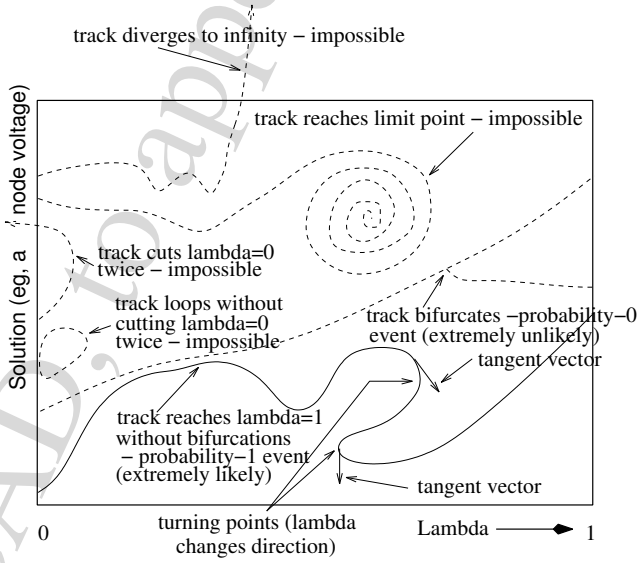


Fig. 2. The different types of track apparently possible in arclength continuation. All except the one that leads to eventual solution of $\bar{g}(\bar{x}) = 0$ are either impossible or of probability 0.

The key innovation of arclength continuation is to transform Equation 2 into a special differential equation, called the *defining ODE*. Note that time does not play any role in this differential equation. The initial-value solution of the defining ODE constitutes a solution of Equation 2 (details are provided in Appendix A). The arclength continuation algorithm then merely solves the defining ODE as an initial value problem, thereby generating a sequence of points $\{(\bar{x}^i,\lambda^i)\}$ which satisfy Equation 2. If $\lambda^k = 1$ at any point, then from definition, Equation 1 is solved by $\bar{x}^k$, *i.e.*, the original nonlinear problem is solved.

The sequence $\{(\bar{x}^i,\lambda^i)\}$ is merely the discrete samples of the continuation track, such as the one shown in Figure 1 for the Schmitt trigger. Several qualitatively distinct types of track are possible, as shown in Figure 2. The horizontal axis depicts the progress of λ, which varies between 0 and 1. The vertical axis represents the solution of the circuit at a given value of λ. The continuation starts at $\lambda = 0$, where the solution of the circuit is easily obtained. The algorithm continues to generate the track until $\lambda = 1$ is reached, *i.e.*, the DC operating has been obtained.

It can be shown [1] that the top four kinds of tracks illustrated in Figure 2 cannot occur with arclength continuation. The topmost track veers off towards a solution of infinity; this is impossible for well-defined circuits because no-gain conditions on the circuit's models dictate that unbounded solutions cannot exist [11]. The next track spirals towards a limit point; this can also be shown to be impossible owing to special properties of the defining ODE solved by arclength continuation. The third track returns to $\lambda = 0$, but not to the same solution that it started from; this is impossible because of an important additional condition imposed on the *start system* $\bar{f}(\bar{x},0) = \bar{0}$, namely that its solution be *unique*. The fourth track returns to the original solution at $\lambda = 0$ but from a different direction, thereby completing a loop; this can be shown to cause the start system $\bar{f}(\bar{x},0) = \bar{0}$ to violate another assumption, that the start system is *structurally stable* [9]. The fifth track from the top, depicted in a bolder dashed line, corresponds to a situation in which the defining ODE of the homotopy ceases to become well-defined mathematically. This corresponds to a bifurcation of the solution curve, as shown. Though technically possible, this is a probability-0 (*i.e.*, extremely unlikely) event, due to Sard's theorem [21]; it can be shown that random perturbations can always replace such a situation with that depicted by the lowermost track. Yet another possibility, not depicted in the figure, exists – that the track wanders between $\lambda = 0$ and $\lambda = 1$ forever without crossing $\lambda = 1$. This can again be shown to be impossible based on properties of the defining ODE. The lowermost track illustrates the normal, extremely likely (or probability-1) case of tracks that reach $\lambda = 1$ without bifurcations.

An important concept in arclength continuation is that of the *tangent vector*, which has a simple interpretation: it is the tangent to the track at any point. Two instances of the tangent vector are shown on the lowermost track. The algorithm proceeds by first calculating the tangent vector, from which it then determines the next point on the curve by extrapolation. Turning points correspond to the λ-component of the tangent vector becoming 0; two turning points are shown on the lowermost curve in the figure.

As mentioned, an important requirement for the start system $\bar{f}(\bar{x},0) = \bar{0}$ is that it have a unique solution $\bar{x}_0$ that is easy to solve for. For a given nonlinear problem $\bar{g}(\bar{x})$, a variety of different constructions for $\bar{f}$ can meet this condition. One construction that can be applied easily to any $\bar{g}$ is

$$\bar{f}(\bar{x},\lambda) \equiv \lambda \bar{g}(\bar{x}) + (1-\lambda)(\bar{x} - \bar{a}), \qquad (3)$$

where $\bar{a}$ is any constant vector. If $\bar{g}(\bar{x})$ is a nodal analysis formulation of the circuit's equations, the above equation has a simple circuit interpretation: the current through each existing device is multiplied by λ and new resistors and current sources are added from each node to ground, of conductance $1-\lambda$ and current $(1-\lambda)a_i$ respectively. A variant is to consider only nonlinear elements (e.g., MOS devices) when adding the conductances, as depicted in Figure 3. The currents through the resistor are multiplied by $1-\lambda$ while the currents through the MOSFET are multiplied by λ, as indicated by Equation 3.

IV. Failure of homotopy methods in circuit applications

A. Empirically observed failure mechanisms

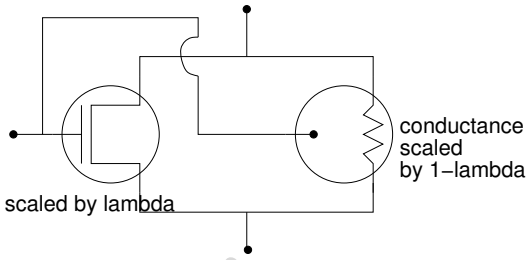


Fig. 3. A simple homotopy embedding for MOSFETs, based on a λ-weighted parallel combination of the MOSFET with a resistor.

Despite the provable convergence properties of arclength continuation, straightforward homotopies, such as the one depicted in Figure 3, were found not to work well in practice, especially on large circuits. Figure 4 depicts the tracks produced by two different homotopies on a medium-sized circuit with 1800 MOS devices. The horizontal axis depicts the arc-length of the track and the vertical axis the value of λ. The simple resistive homotopy of Figure 3 produced the longer track with many turning points, which never reached $\lambda = 1$ but continued to meander indefinitely. The shorter track was produced by an experimental homotopy called SSIM [20]. Though considerably more robust than the resistive homotopy for small circuits, SSIM also failed on a significant number of large circuits. In this example, failure was manifested by the track's returning to zero instead of going to 1. Similar observations have been noted by others [10], possibly accounting for the absence of homotopy-based techniques in commercial simulators.

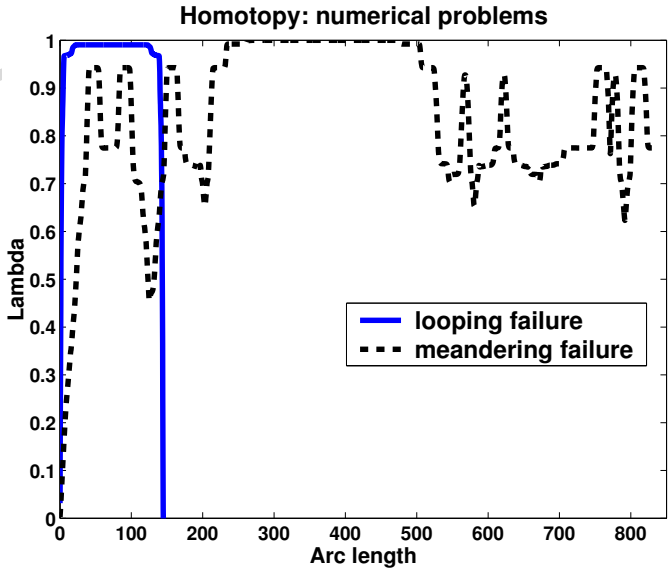


Fig. 4. Continuation tracks corresponding to failures of straightforward homotopies for an 1800-MOSFET circuit..

The ATANSH homotopy described later in Section V resulted from an effort to determine why the above homotopies performed poorly in practice despite theoretical guarantees of suc-

cess. During the course of the investigation, two broad mechanisms of failure were found: 1) ill-conditioned numerics leading to failure of path-following, and 2) homotopy paths that continued forever (i.e., for impractically long) without reaching $\lambda = 1$.

Numerical ill-conditioning manifested itself in the linear systems being solved by the continuation algorithm to find the next point on the track. Condition numbers of 10^{12} or more were common for large MOS circuits with conventional homotopies. Such poor conditioning led to a number of problems: large inaccuracies in tangent vector calculation, leading to poor prediction of Newton starting points, further exacerbated by inaccurate Newton updates, leading to Newton failure. This forced the algorithm to try smaller steps along the track, leading to increased ill-conditioning thus compounding the problem further. Eventual failure resulted either from arclength-steps becoming too small or by the track looping back upon itself and returning to zero (as illustrated in Figure 4). Track looping resulted from such inaccurate tangent or newton update calculation that the algorithm converged on the path already traversed. Examination of several tracks revealed that ill-conditioning was often associated with the track bending back and traversing close to previous sections.

In another mode of failure ("the long path phenomenon"), however, the continuation would proceed with relatively acceptable matrix conditioning for very long arc-lengths without reaching $\lambda = 1$. The algorithm either would not terminate at all or would eventually fail by the ill-conditioning mechanisms described above. Again, examination of the track revealed evidence that the track was tracing a spiral in n dimensions, thereby leading to very slow progress in λ.

It was found empirically that the two failure modes were correlated. Long paths were often the result of the track's looping or spiralling in n-d space, a problem exacerbated as the problematic segments of the track become less well-separated.

B. Understanding the failure of homotopy methods: exponentially long homotopy tracks

In this section, we identify a fundamental cause for the long-path phenomenon. We show that given a homotopy, the path length of the solution track is exponential in the number of *nested sets of turning points*.

Consider the circuit of Figure 5, consisting of two Schmitt trigger circuits with their inputs connected together at x. The outputs of the triggers are at y_1 and y_2 respectively. The input-output characteristics of the Schmitt triggers are shown in Figure 6. It is assumed that the two characteristics are *not* identical, but that the horizontal interval between the two folds of one is entirely contained within that of the other, as shown. The circuit corresponding to the smaller interval is termed the *inner* flip-flop; the larger one the *outer* flip-flop; the structure is termed *nested*. The nested structure can be generalized to more Schmitt triggers, with each succeeding characteristic containing all the previous ones; only two are considered here for simplicity.

For the purpose of exposition, a homotopy technique that is much simpler than ATANSH is applied to the circuit. The object of this homotopy is to sweep the input voltage x from -10V

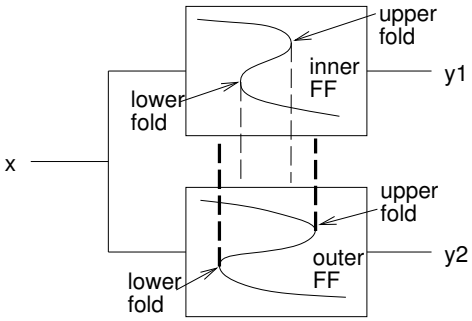


Fig. 5. Two 'nested' flip-flops with their inputs ganged together.

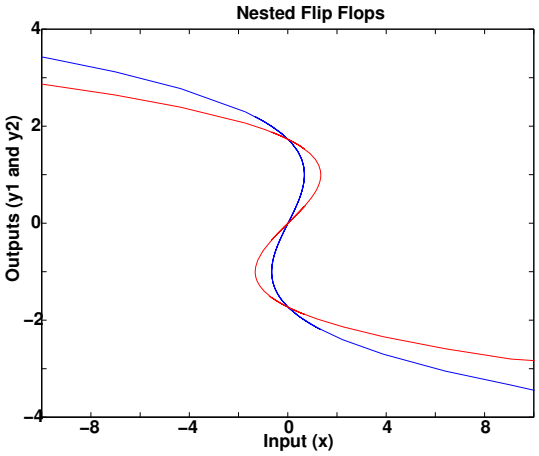


Fig. 6. Characteristics of two flip-flops with the hysteresis region of one nested within that of the other.

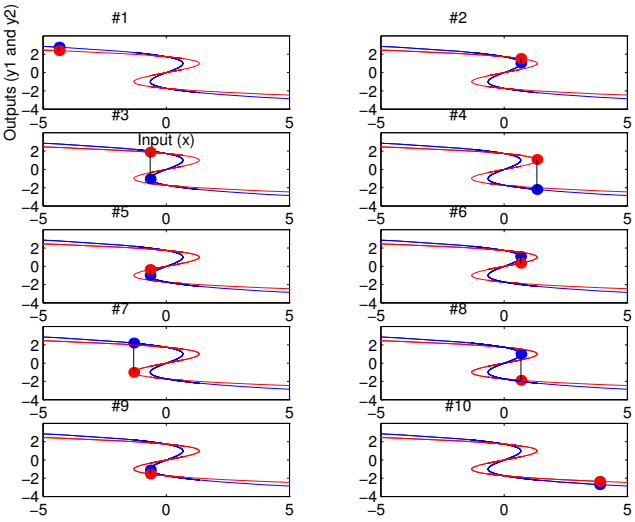


Fig. 7. Stages in the progress of the source homotopy for the nested flip-flops.

(corresponding to $\lambda = 0$) to 10V (corresponding to $\lambda = 1$). This is similar to source stepping of the circuit's input, but using numerical continuation so that turning points (or folds) are negotiated smoothly, *i.e.*, without abrupt jumps. It will be shown shortly that the very fact that jumps are not allowed, a key factor in the global convergence of homotopy, is also responsible for making the track exponentially long in the number of nested flip-flops.

The sequence in Figure 7 illustrates the progress of the source homotopy. The progress of the solution track is represented by the red and blue dots which mark the outputs of the outer and inner flip-flops respectively. Since the inputs to the two flip-flops are connected, the x-intercept of the two dots must be the same – this is emphasized in the figures by the vertical line connecting the dots.

In sub-plot #1, the input x is close to to -5V (a point it has reached by progressing from the start of the continuation at $x = -10$V) and the dots rest on the left edges of the two characteristic curves, as shown. As the continuation progresses, x increases and the outputs y_1 and y_2 change smoothly until the dots reach the position shown in #2, where the blue dot reaches the upper fold of the inner flip-flop. *It is not possible for x to continue increasing* without causing an abrupt jump of the blue dot to the lower segment of the blue curve. Hence *x is forced to decrease*; the blue dot moves smoothly onto the central section of the blue curve, but the red dot is forced to move back over part of the characteristic that it has already covered. x continues decreasing until the blue dot reaches the lower fold of the inner flip-flop, shown in #3; here x starts increasing again and the blue dot continues making progress on its characteristic. The red dot starts moving to the right again on the upper section of the outer flip-flop's characteristic; it is now retracing some parts of this curve for the third time. x continues increasing until it reaches the position shown in #4; by now the blue dot has negotiated both folds of its characteristic, but the red dot has only just arrived at its first fold. Once again, x cannot continue increasing because that would cause a jump, but this time for the red dot. Hence x decreases, the red dot moves onto the central section of its characteristic, and the blue dot traverses sections of its curve for the second time. Before the red dot can reach its lower fold, however, the blue dot arrives at its lower fold, as shown in #5. The red dot cannot continue moving towards its lower fold because the blue dot would have to jump; hence it backtracks (x starts increasing) while the blue dot traverses its central section completely for the second time, but in the reverse direction, until it arrives at its upper fold, shown in #6. At this point, x starts decreasing, the blue dot moves on to the upper section of its curve (still going the "wrong" way), and the red dot starts heading closer to its lower fold again, until it reaches it in #7. At this point, the blue dot is almost back where it started in #1, having made two full traverses of its folds. The red dot has, however, reached its curve's lower segment. x now increases towards +5V, but has to decrease again when the blue dot reaches its top fold in #8; it continues decreasing until #9, where it increases again, having done a third traverse of its folds. Finally, both red and blue dots reach the lower segments of their respective curves, on which they proceed without further setbacks till they reach their goal of +10V, progress towards which is shown in #10.

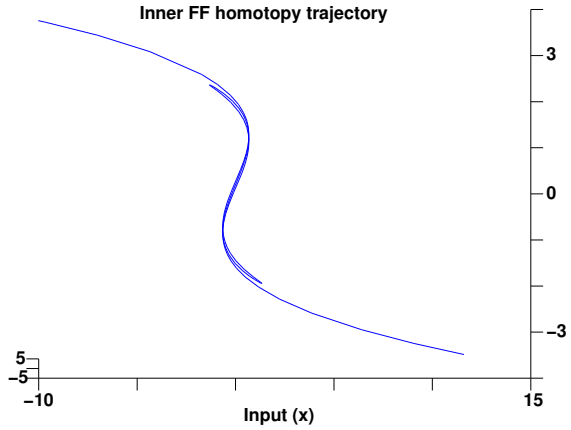


Fig. 8. The trajectory followed by the source homotopy along the characteristic of the 'inner' flip-flop.

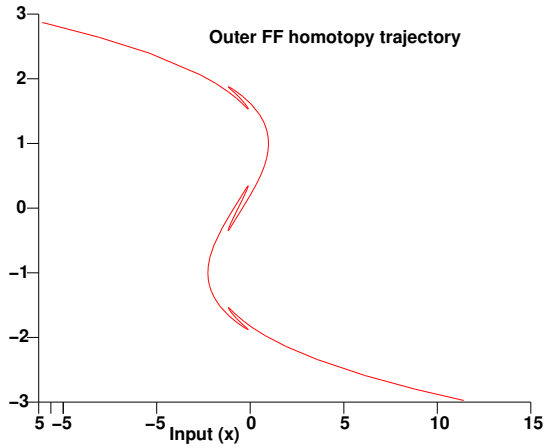


Fig. 9. The trajectory followed by the source homotopy along the characteristic of the 'outer' flip-flop.

Figures 8 and 9 depict the reversals and multiple passes of the trajectories over the characteristics. It may appear from these figures and the above arguments that the trajectories, while continuous, are not smooth because of the pointed corners in these figures. The smoothness of the trajectory is easier to visualize when it is plotted in its totality in three dimensions, as in Figure 10. The input x is plotted on one horizontal axis, while the outputs are plotted on the other two axes. It is apparent now that the pointed corners are an artifice of the projection of the three-dimensional curve onto two dimensions.

The equivalent of the λ vs arclength plots of Section VI is shown in Figure 11. The vertical axis represents x, the equivalent of λ in the ATANSH homotopy; the horizontal axis depicts the arclength s, which is simply the length traversed on the 3-d curve (Figure 10). The many turning points and the qualitative similarity to the trajectory shown in Figure 4 are apparent.

The above demonstrates that the inner characteristic needs to be traversed *thrice* for a single traversal of the outer characteristic. The argument is easily extended to k nested flip-flops, for

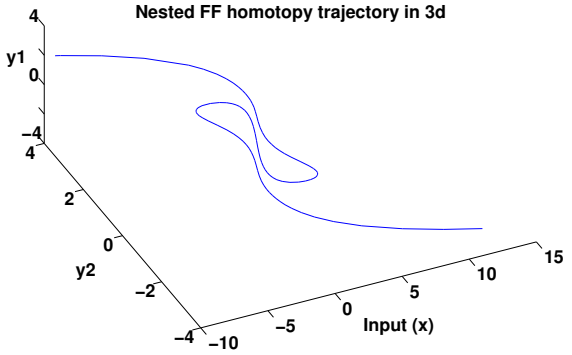


Fig. 10. Three-dimensional view of the trajectory followed by the source homotopy.

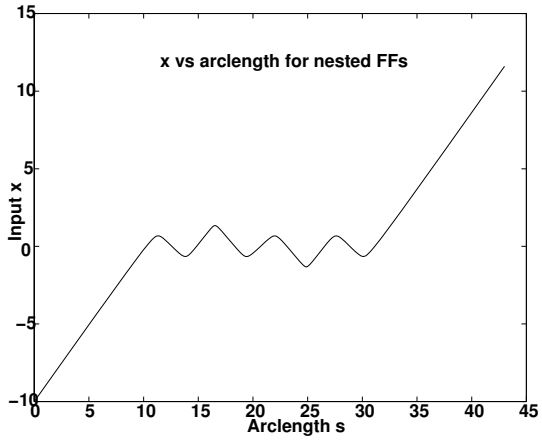


Fig. 11. The track of x vs arclength for the ganged nested flip-flop model.

which the total number of traversals of the innermost characteristic for a single traversal of the outermost is 3^{k-1}. Hence the number of turning points, as well as the total arclength, is *exponential in the number of flip-flops that are nested*. It is interesting to note that the exponential inefficiency vanishes if one characteristic is not completed nested within the other.

It is important to investigate what happens if the flip-flops are identical (or if they are different but the x-intercepts of their folds coincide). This case is likely if a subcircuit representing a flip-flop is instantiated multiple times in the circuit. It can be shown that such a situation will lead to the failure of the arclength continuation algorithm because the Jacobian matrix of the system will lose rank by more than 1 at the folds, arriving at a bifurcation (the fifth track from the top in Figure 2). While the occurrence of several exactly identical flip-flops is very natural because of the use of sub-circuits (certainly not a probability-0 event in hierarchical designs), it does result in several different elements of the circuit being exactly identical, which considering the continuum of possibilities is a probability-0 event from a mathematical standpoint. It is important to note, therefore, that certain mathematically improbable situations are in fact highly likely to arise in practice. It is for this reason that randomization [1, 18] in probability-1 homotopy techniques is of considerable practical importance. The use of randomization in arclength continuation addresses the algorithm failure problem by introducing random changes to all elements so that they do not coincide, but in such a manner as to ensure that the solution reached is not affected by the random changes.

In the case of identical flip-flops, however, small random changes can introduce nesting, thereby replacing outright failure of the algorithm with extreme inefficiency and ill-conditioned numerics. In practice this is possibly even worse because it takes longer to declare failure and wastes design time. Randomization in numerical continuation is usually applied as small perturbations, perhaps because the term 'probability-1' is interpreted to imply that even the slightest change corrects algorithm failure.

Although the example above uses a simple natural-parameter homotopy, the same phenomenon also occurs in artificial-parameter homotopies . Any single-parameter continuation on a large circuit consisting of sub-circuits with independent turning points is potentially susceptible to the long path phenomenon.

The identification of the mechanism leading to long paths immediately suggests a means of circumventing it. Note that the reason for the looping is that the inputs to the flip-flops are constrained to be the same; more generally, *a single continuation parameter* λ *is common to all the sub-circuits with turning points*. It is apparent in the flip-flop example above that if the inputs were decoupled and the path of each flip-flop followed independently, then the long path problem is circumvented. It is this key insight, *decoupling the continuation parameters* of different turning point mechanisms, that leads to the two-phase ATANSH homotopy described in the following section.

V. The ATANSH two-phase MOS homotopy

In this section, a robust and efficient homotopy that is effective for large MOS circuits is described. A key feature of this MOS homotopy model is that it is constructed with *two* λ parameters, λ_1 and λ_2. In Section V-A, the dependence of the MOS model on these parameters is described; in Section V-B, the use of a single-λ-based homotopy solver with a coupling between λ_1 and λ_2 is presented.

A. Dependence on λ_1 and λ_2

The ATANSH MOS homotopy model is symmetric and bulk-referenced [13, 26], taking the electrical inputs $V_{gb} = V_g - V_b$, $V_{sb} = V_s - V_b$ and $V_{db} = V_d - V_b$. V_s, V_b, V_g and V_d represent the voltages at the source, bulk, gate and drain nodes respectively. In addition, the model uses two homotopy parameters λ_1 and λ_2 which take values in $[0,1]$. λ_1 influences the drain-source driving-point characteristic whereas λ_2 controls the transfer characteristic, i.e., the influence of the gate on the drain current.

The form of the drain-source current I_{ds} for the ATANSH homotopy is

$$I_{ds} = \frac{\beta}{2}\left[V'_{gs}(V_{gb}, V_{db}, V_{sb}, \lambda_2, \lambda_1)\right]^2 h(V_{db} - V_{sb}, \lambda_1). \quad (4)$$

Equation 4 is a single-piece model, qualitatively resembling the Schichman-Hodges (SH) model in that it contains a quadratic term in V_{gs} multiplying a term determined by V_{ds}. β is functionally identical to the corresponding parameter in the SH model; $V'_{gs}(V_{gb},V_{db},V_{sb},\lambda_2,\lambda_1)$ is a symmetric, bulk-referenced version of the gate-source voltage that is modulated mainly by λ_2. This latter term has maximum effect at $\lambda_2 = 1$ and minimum effect at $\lambda_2 = 0$. The effect of bulk referencing is to switch V'_{gs} smoothly between the gate-source and the gate-drain voltages depending on whether the drain is at a higher or lower potential than the source. $h(V_{db}-V_{sb},\lambda_1)$ is a λ_1-modulated version of the driving-point characteristic; the effect of λ_1 is to modify the degree of nonlinearity in the driving point characteristic.

Specifically, for an N-type MOS device, V'_{gs} in (4) is given by

$$\begin{aligned} V'_{gs}(V_{gb},V_{db},V_{sb},\lambda_2,\lambda_1) &= \frac{\lim_s(V'_{gb}-\min_s(V'_{sb},V'_{db}))}{V_{gs,nom}}, \\ \text{where } V'_{gb} &= (1-\lambda_2)V_{gb,nom}+\lambda_2 V_{gb}, \\ V'_{sb} &= V_{sb}(0.1(1-\lambda_1)+\lambda_1), \\ \text{and } V'_{db} &= V_{db}(0.1(1-\lambda_1)+\lambda_1). \end{aligned} \tag{5}$$

$V_{gb,nom}$ is a constant representing a nominal value for V_{gs} with the transistor on; for example, $V_{gb,nom} = 3$ is a reasonable value for most current technologies. $\min_s(a,b)$ is any smooth function that approximates $\min(a,b)$, while $\lim_s(a)$ is a smoothed version of the function

$$f(a) = \begin{cases} a & \text{if } a > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Example instantiations of these functions are

$$\begin{aligned} \min_s(a,b) &= a - \lim_s(a-b), \text{ with} \\ \lim_s(x) &= x\frac{1+\tanh(kx)}{2}, \end{aligned} \tag{6}$$

where $k > 0$ represents a smoothness parameter.

$h(\cdots)$ in (4) is given by

$$h(V_{db}-V_{sb},\lambda_1) = \frac{2}{\pi}\tan^{-1}\left(G_{nom}(V'_{db}-V'_{sb})\right), \tag{7}$$

where G_{nom} is a constant representing, loosely, a nominal value of the triode-region drain-source conductance of a MOS device; for example, $G_{nom} = 5$.

An appreciation of how varying λ_1 and λ_2 affects the characteristics of the ATANSH model can be gained from Figure 12. Each small three-dimensional plot represents the variation of the drain-source current (plotted on the vertical axis) as a function of the gate-source and drain-source voltages (represented on the horizontal axes) at fixed values of λ_1 and λ_2. λ_1 and λ_2 vary on the large vertical and horizontal axes. The bottom left corner depicts the $(\lambda_1,\lambda_2) = (0,0)$ case and the top right the $(1,1)$ case. Moving vertically from bottom to top, λ_1 increases from 0 to 1; likewise, λ_2 increases from 0 to 1 horizontally from left to right.

At $(\lambda_1,\lambda_2) = (1,1)$ (the top right), the model characteristics are similar to that of the SH model, exhibiting a quadratic dependence on V_{gs} and linear and saturation regions as a function

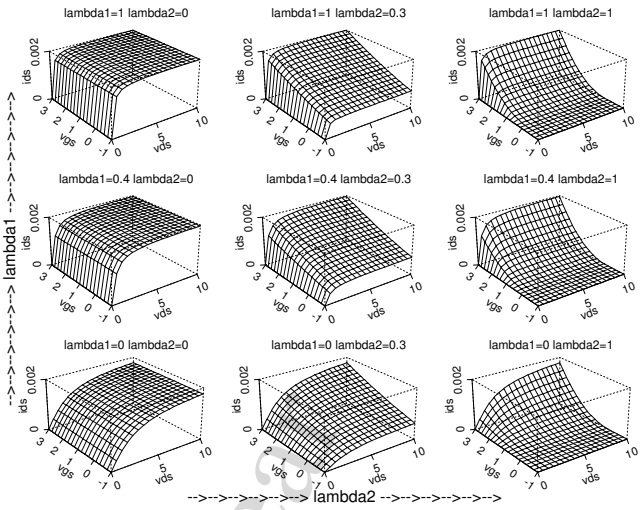


Fig. 12. **ATANSH**: Model characteristics as a function of the decoupled continuation parameters λ_1 and λ_2.

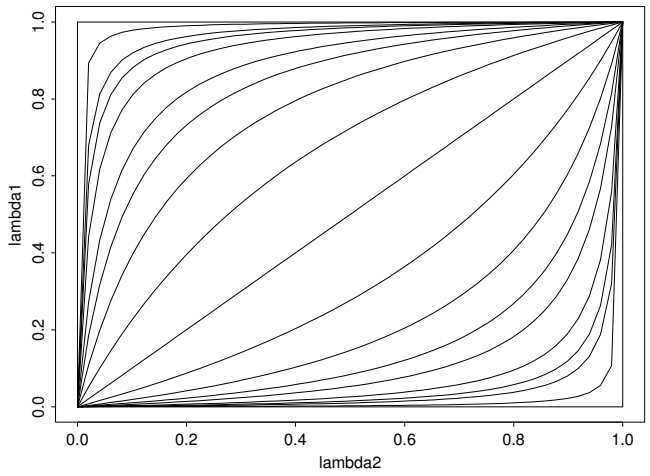
of V_{ds}. At $(\lambda_1,\lambda_2) = (0,0)$ (the bottom left), it can be seen that there is no transfer characteristic (varying V_{gs} does not alter I_{ds}), and that the driving point characteristic is much less sharp than for the original MOSFET. The start system corresponds to $(\lambda_1,\lambda_2) = (0,0)$, at which each MOS device becomes a two-terminal almost-linear resistor; hence the circuit becomes easy to solve using the Newton-Raphson method, typically taking fewer than 10 iterations to solve. The effect of varying λ_1 and λ_2 is also apparent from the figure: λ_1 sharpens the driving point characteristic without affecting the gain whereas λ_2 ramps the gain without sharpening the driving-point characteristic.

B. Homotopy using two λ parameters

Practical arc-length continuation algorithms [12] are based on a *single* continuation parameter λ, leading to a system of n equations in $n+1$ variables. Since ATANSH has two continuation parameters, a system of n equations in $n+2$ variables results. One approach to converting this into a one-parameter homotopy is to add an extra equation to obtain a system of $n+1$ equations in $n+2$ variables to which a conventional homotopy solver can be applied.

It is necessary for the extra equation to be specified such that the solution of the original circuit is respected and that the conditions for arclength continuation continue to hold. Any smooth curve relating only λ_1 and λ_2 and passing through $(\lambda_1,\lambda_2) = (0,0)$ and $(\lambda_1,\lambda_2) = (1,1)$ satisfies the above conditions. An infinite number of such curves is possible; one such family, arbitrarily chosen for purposes of illustration, is shown in Figure 13. This family of curves, parametrized by the real number m, is given by

$$\begin{aligned} \phi_m(\lambda_1,\lambda_2) &\equiv \lambda_1 - \psi_m(\lambda_2) = 0, \\ \psi_m(x) &= \gamma(m) + \frac{1}{m^2}(1-x-\gamma(m)), \text{ and} \\ \gamma(m) &= \frac{1}{2}\left(1 - \frac{m}{|m|}\sqrt{1+\frac{4}{m^2}}\right). \end{aligned} \tag{8}$$



.

Fig. 13. The family of curves $\phi_m(\lambda_1,\lambda_2) = 0$ given by Equation 8.

As $m \to 0$, $\psi_m(x) \to x$; as m increases from 0, $\psi_m(x)$ is shown by the upper curves in the figure; likewise, as m decreases, $\psi_m(x)$ is shown by the lower curves. Of interest are the limiting curves obtained as $m \to \pm\infty$, given by the left and upper boundaries of Figure 13, and by its lower and right boundaries, respectively. Corresponding to these limit curves are the first column and top row of Figure 12, and the bottom row and third column, respectively.

While these limit curves are not smooth (violating smoothness requirements for arclength continuation methods), they do have the property of decoupling the homotopy into two independent parts, one controlled by λ_1 keeping λ_2 fixed, the other by λ_2 keeping λ_1 fixed. For the $m \to \infty$ limit curve, λ_1 is ramped first, whereas for the $m \to -\infty$ limit curve, λ_2 is ramped first. In this work, the lower curve ($m \to -\infty$) is used; the horizontal and vertical segments of this path are referred to as *phase 1* and *phase 2* of the homotopy, respectively. It is to be noted that while using the $m \to -\infty$ limit leads to a robust and efficient DC solution technique, the $m \to \infty$ curve causes failures due to inefficiency and numerical problems. An intuitive understanding of this behaviour is provided by Figure 12, where it can be seen that the latter path is "smoother" than the former, which reaches a highly nonlinear characteristic at $(\lambda_1,\lambda_2) = (1,0)$ before becoming smoother again at $(\lambda_1,\lambda_2) = (1,1)$.

For practical design, it is of course necessary to obtain the operating point of the circuit using existing in-house MOS models that have been characterized to model fabricated devices very accurately [13]. The ATANSH model is not meant to be a substitute for such models; indeed it is inadequate for modelling second-order effects on which circuit performance is often predicated. The utility of ATANSH lies in that the operating point obtained with it is very similar to that with more accurate models – hence this operating point can be used as a starting guess to solve the circuit with standard models using, for example, the Newton-Raphson method, relying on its local convergence properties. This approach works very well for most circuits. It is possible, moreover, to use continuation for smoothly substituting the standard model for the ATANSH model as well. Each MOSFET is replaced by a composite weighted combination of ATANSH and the desired model with the weights depending on a third continuation parameter λ_3. Using continuation of λ_3 (*phase 3*), the composite is changed smoothly from ATANSH at $\lambda_3 = 0$ to the desired model at $\lambda_3 = 1$.

From a theoretical standpoint, it is preferable to perform all three phases (ramping λ_2, λ_1 and λ_3) as part of a single smooth homotopy, since it restores smoothness conditions that are violated by the approach outlined in the previous paragraphs. This can be achieved by the straightforward extension of the construction of Figure 13 to three continuation parameters. Our experience however has been that in practice, very few circuits fail as a result of the sharp corners in the limit curves of Figure 13 and its three-dimensional extension; only one has in fact been identified, out of a conservative estimate of a few thousand conventionally hard-to-solve circuits on which the three-phase technique has been effective. A reason for preferring the three-phase technique over the single unified homotopy is that implementation and coding become significantly simpler due to the decoupling of the λ_2, λ_1 and λ_3 homotopies. Further, a saving in computation is also achieved during the first and second phases because ATANSH is several times less expensive to compute than accurate MOS models such as the BSIM family.

VI. Results

The ATANSH homotopy described in the previous section was in production use within AT&T/Lucent Microelectronics since 1995. It proved to be so reliable in finding DC operating points of large circuits that methodology changes resulted within some design groups [3]. It became possible to relegate worst-case testing of much larger blocks than previously practical to automated scripts operating without user intervention. Tedious manual initialization, splitting circuits into smaller parts and other ad-hoc techniques, often a last resort of frustrated designers who required an operating point to proceed with other analyses, were virtually eliminated. This led to considerable savings in design time; it was not unusual previously for several days to be spent in obtaining operating points of "tough" circuits.

The first and second columns of Table I list the names and types of a sampling of circuits that exhibit problems with conventional methods. The circuits range from active filters (dlopata1, heideh), mixed analog-digital circuits involving sigma-delta ADCs, filters, phase mixers, control and division circuitry (test9, vf_test, rabb-xare, addas, dctl.t) to digital blocks and SRAMs (s1423, goh). All circuits except s1423 were obtained from designers in AT&T/Lucent Microelectronics who were unable to find an operating point using conventional methods; s1423 is an ISCAS benchmark circuit which exhibited convergence difficulties with the most accurate in-house MOS model.

The third column lists the number of MOS devices in the circuits, which range from small (127 MOSFETs) to relatively large (8489 MOSFETs) in size. The fourth column lists the CPU time (on a Sun SPARCstation 2 with 96MB of memory) required by the ATANSH homotopy to obtain an operating point of the circuit. The fifth column lists the CPU time for conventional techniques to announce *failure* – this is helpful as a lower

Circuit name and type	Number of MOS devices	ATANSH CPU secs (for success)	Newton Raphson CPU secs (to declare failure)
dlopata1 (analog)	127	13	4331
heideh (analog)	192	49	244
test9 (mixed A/D)	1380	599	3209
vf_test (mixed A/D)	1621	565	2101
rabb-xare (mixed A/D)	1877	1035	2340
addas.com (mixed A/D)	3413	1195	4395
s1423 (digital)	3736	678	4207
dctl.t (mixed A/D)	7199	10385	10150
goh (digital)	8489	3339	11700

TABLE I
ATANSH HOMOTOPY VS CONVENTIONAL ALGORITHMS.

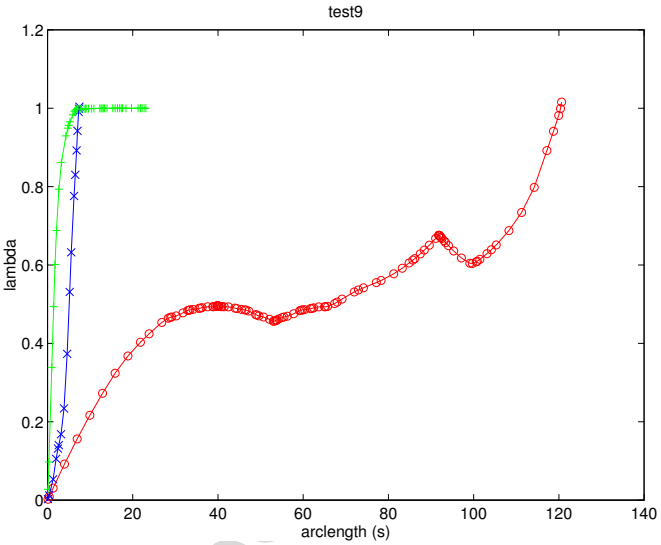


Fig. 15. ATANSH homotopy track (λ vs arclength s) for the circuit test9.

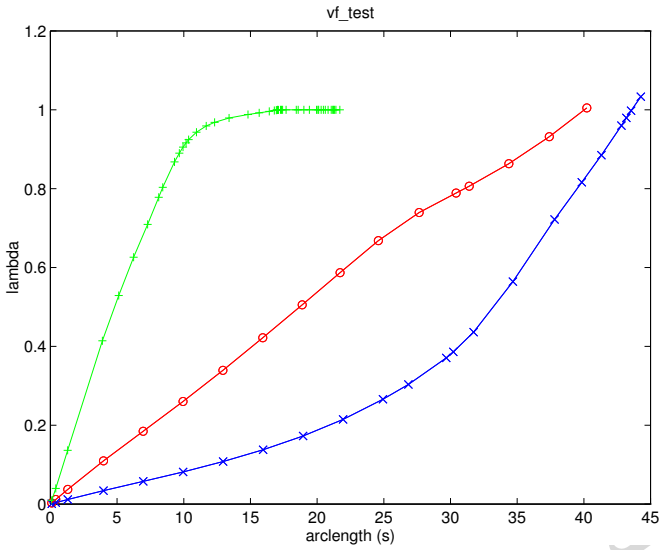


Fig. 14. ATANSH homotopy track (λ vs arclength s) for the circuit vf_test.

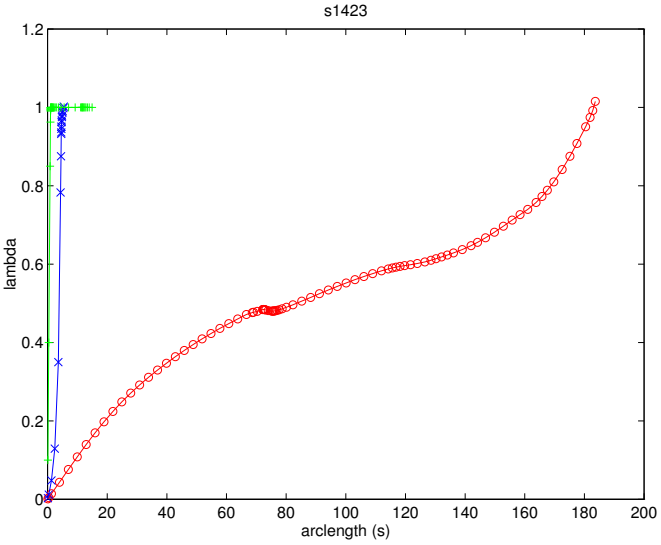


Fig. 16. ATANSH homotopy track (λ vs arclength s) for the ISCAS benchmark circuit s1423.

bound on the time wasted by a designer trying to obtain a solution of the circuit.

It can be seen that in most cases, the ATANSH homotopy took considerably less time to obtain the DC operating point of the circuit than conventional methods took to give up. It should be noted, however, that when the Newton-Raphson method does succeed, it is a factor of 2–3 faster than the ATANSH homotopy on the average. It must be emphasized, though, that CPU time is not the most important criterion for designers trying to obtain an operating point (especially for extracted circuits), so long as it remains within reasonable limits – far more important is success as opposed to failure. For big circuits, most designers are happy to trade several extra hours of unattended clock time in return for guaranteed convergence. The impact of homotopy stems from its ability to deliver on this guarantee for large, 'hard', circuits.

Figures 14 – 16 provide a graphical representation of the progress of the ATANSH homotopy for three of the above circuits. The horizontal axis represents the arclength s (see Appendix A for an explanation of the significance of s) of the $n+1$-dimensional solution curve generated by the homotopy solver. Roughly speaking, it is a measure of computation time for a given circuit (note that similar values of s do not correspond to similar computation times across different circuits). On the vertical axis, the value of the continuation parameter λ is plotted. This is a measure of the progress the algorithm has made; success is indicated by the track's reaching $\lambda = 1$. Each figure has three plots in the colours red, blue and green. The λ axis represents λ_2, λ_1 or λ_3 depending on the colour of the plot. The red plot corresponds to the first phase of ATANSH where (λ_1, λ_2) changes from $(0,0)$ to $(0,1)$; in other words, $\lambda = \lambda_2$ is varied by the continuation algorithm while λ_1 is kept constant at 0. The

second phase is depicted by the blue plot; $\lambda = \lambda_1$ is varied while λ_2 is kept constant at 1. The green plot depicts the final phase, the transition from the ATANSH model to the desired accurate in-house MOS model controlled by $\lambda = \lambda_3$. The solution of the circuit with the accurate in-house MOS model is found when the green track reaches 1 on the λ axis.

The three tracks in Figure 14 are for the vf_test circuit. Both red and blue tracks (phases 1 and 2) proceed monotonically and with relatively few points from $\lambda = 0$ and $\lambda = 1$, indicating that the circuit is not particularly challenging for the ATANSH homotopy. The green track (phase 3) shows fast progress initially, indicating very little change from the solution obtained with the ATANSH model; the progress slows as it approaches $\lambda_3 = 1$, indicating that the solution is changing at the last stages of the substitution of ATANSH by the accurate in-house model. This is typical of circuits in which some node voltages depend strongly on the second-order details of the MOS model being used – for example near-floating nodes whose voltages are primarily determined by the g_{ds} of MOS devices connected to them.

More interesting behaviour is observed in Figure 15 for the test9 circuit. The red track of the first phase is seen to be non-monotonic; it displays two pairs of turning points at which λ_2 changes from increasing to decreasing or vice-versa. Circuits that display such turning points often fail with conventional methods. The blue and green tracks are straightforward by contrast.

The tracks of the s1423 circuit are shown in Figure 16. It can be seen that the red track of the first phase also has a "small" pair of turning points at λ_2 of about 0.5.

It should be noted that the two-parameter ATANSH homotopy presented in this paper is eventually a heuristic, though a remarkably effective one – it has been seen to be extremely effective for a vast array of practical problems. While the nested turning-point phenomenon may not be the only cause of homotopy failure, our results indicate that it is the predominant mechanism, whether or not actual flip-flop structures are involved. It is emphasized that with artificial-parameter homotopies, nesting of turning points can occur without any designer-obvious nesting of multi-stable circuits; abstract nesting can occurs between different devices without any need for real flip-flop-like circuit structures. The key property of ATANSH is that it decouples homotopy-induced changes of individual transistors in large circuits from each other, thus mitigating the long-path phenomenon examined in Section IV-B. Intuitively, signal transfer from transistor to transistor is first reduced by setting $\lambda_1 = 0$; this makes it easier to solve all parts of the circuit independently first, by ramping λ_2, before the circuit is 're-connected' by turning on driving point characteristics (important for, *e.g.*, active loads) with λ_1.

VII. Conclusion

A new homotopy-based technique for finding DC operating points of large-scale MOS circuits has been presented. A generic means by which traditional homotopies often fail, characterized by interacting turning points and exponentially long path lengths, has been identified. The failure mechanism has been illustrated using nested bistable structures. A key feature of the new MOS homotopy ATANSH, its two-phase nature, is based on the insight that decoupling the continuation parameter λ over the circuit can circumvent the inefficiency.

ATANSH was in production use inside AT&T/Lucent Microelectronics since 1995. It succeeded in finding operating points for the overwhelming majority of 'difficult' circuits (*i.e.*, those that failed Newton-Raphson) that it encountered.

Acknowledgments

The first author would like to thank Bruce McNeill and David Rich for detailed feedback on initial versions of ATANSH and their encouragement of this work. He would like to acknowledge discussions with Ljiljana Trajković on many aspects of this and previous work on homotopy. This work benefitted considerably from initial work by Sani Nassif and Colin McAndrew on the SSIM homotopy. Doug Lopata, Bob Walden, Scott Fetterman, Jeff Sonntag, Bernie Morris and Goh Komoriya contributed difficult circuits that were instrumental in testing and refining the new homotopy. The first author would like to acknowledge many helpful interactions with members of the development team of Celerity, the simulator in which this work was implemented: Sani Nassif, Kishore and Kumud Singhal, Ken Haruta, Colin McAndrew, Peter Lloyd, Jeff Hantgan, Bijan Bhattacharyya, Karti Mayaram, Larry Nagel and David Lee. Partial support of the first author during preparation of this paper by the National Science Foundation (awards CCR-0312079 and CCR-0204278), the Semiconductor Research Corporation, and DARPA, is acknowledged. Computational facilities utilized included those provided by the Minnesota Supercomputing Institute. Last but not least, the first author gratefully acknowledges a number of insightful suggestions for improvement of this paper by the anonymous reviewers, particularly reviewer #3.

References

[1] E.L. Allgower and K. Georg. *Numerical Continuation Methods*. Springer-Verlag, New York, 1990.

[2] A.Ushida, Y.Yamagami, Y.Nishio, I.Kinouchi, and Y.Inoue. An efficient algorithm for finding multiple DC solutions based on SPICE-oriented Newton Homotopy method. *IEEE Trans. CAD*, 21:337–348, March 2002.

[3] B. McNeill, D. Lopata, D. Rich. Personal communications, 1995.

[4] D. Wolf and S. Sanders. Multiparameter homotopy methods for finding DC operating points of nonlinear circuits. *IEEE Trans. Ckts. Syst. – I: Fund. Th. Appl.*, 43(10):824–838, October 1996.

[5] M. Gourary, S. Ulyanov, M. Zharov, S. Rusakov, K.K. Gullapalli, and B.J. Mulvaney. Simulation of High-Q oscillators. In *Proc. ICCAD*, November 1998.

[6] M.M. Green and R.C. Melville. Sufficient conditions for finding multiple operating points of dc circuits using continuation methods. *Proc. IEEE Int. Symp. Ckts.r Sys.*, pages 117–120, May 1995.

[7] H.G. Brachtendorf, S. Lampe, R. Laur, R. Melville, and P. Feldmann. Steady state calculation of oscillators using continuation methods. In *Proc. IEEE DATE Conference*, page 1139, 2002.

[8] J. Lagarias and L. Trajkovic. Bounds for the number of DC operating points of transistor circuits. *IEEE Trans. Ckts. Syst. – I: Fund. Th. Appl.*, 46(10):1216–1221, October 1999.

[9] J. Kevorkian and J.D. Cole. *Perturbation methods in Applied Mathematics*. Springer-Verlag, 1981.

[10] K.S. Kundert. Personal communication, April 1992.

[11] R.C. Melville L. Trajković and S.-C. Fang. Passivity and no-gain properties establish global convergence of a homotopy method for DC operating points. *Proc. IEEE Int. Symp. Ckts. Sys.*, pages 914–917, May 1990.

[12] L.T. Watson, S.C. Billups and A.P. Morgan. Algorithm 652: HOMPACK: a suite of codes for globally convergent homotopy algorithms. *ACM Trans. Math. Software*, 13:281–310, 1987.

[13] C.C. McAndrew and B.K. Bhattacharyya. A single-piece C^∞-continuous MOSFET model including subthreshold conduction. Technical Report 52864-900901-01TM, AT&T Bell Laboratories.
[14] L.W. Nagel. *SPICE2: a computer program to simulate semiconductor circuits*. PhD thesis, EECS Dept., Univ. Calif. Berkeley, Elec. Res. Lab., 1975. Memorandum no. ERL-M520.
[15] J.M. Ortega and W.C. Rheinboldt. *Iterative solution of nonlinear equations in several variables*. Academic Press, 1970.
[16] P. Horowitz and W. Hill. *The Art of Electronics*. Cambridge University Press, Cambridge, UK, 1995.
[17] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes – The Art of Scientific Computing*. Cambridge University Press, 1989.
[18] S.-C. Fang R.C. Melville, L. Trajković and L.T. Watson. Artificial Parameter Homotopy Methods for the DC Operating Point Problem. *IEEE Trans. CAD*, 12(6):861–877, June 1993.
[19] J. Roychowdhury and R.C. Melville. Homotopy Techniques for Obtaining a DC Solution of Large-Scale MOS Circuits. In *Proc. IEEE DAC*, June 1996.
[20] S. Nassif and C. McAndrew. Personal communications, 1993–1994.
[21] A. Sard. The measure of the critical values of differential maps. *Bull. Amer. Math. Soc.*, 48:883–890, 1942.
[22] R. Seydel. *From Equilibrium to Chaos – Practical Bifurcation and Stability Analysis*. Elsevier, New York, 1988.
[23] T.L. Quarles. *SPICE 3C.1 User's Guide*. University of California, Berkeley, EECS Industrial Liaison Program, University of California, Berkeley California, 94720, April 1989.
[24] T.S. Coffey, C.T. Kelley, and D.E. Keyes. Pseudotransient continuation and differential-algebraic equations. *SIAM J. Scientific Computing*, 25(2):553–569.
[25] T.S. Coffey and D.E. Keyes. Convergence analysis of pseudo-transient continuation. *SIAM J. Numerical Analysis*, 35(2):508–523.
[26] Y.P. Tsividis. *Operation and modeling of the MOS transistor*. McGraw-Hill, New York, 1987.
[27] A. Ushida and L.O. Chua. Tracing solution curves of nonlinear equations with sharp turning points. *International Journal of Circuit Theory and Applications*, 12:1–21, 1984.
[28] W. Ma, L. Trajkovic, and K. Mayaram. HomSSPICE: a homotopy-based circuit simulator for periodic steady-state analysis of oscillators. In *Proc ISCAS*, pages I–645–I–648, June 2002.
[29] W.I. Zangwill and C.B. Garcia. *Pathways to Solutions, Fixed Points and Equilibria*. Prentice-Hall, 1981.

APPENDIX

A. Arclength continuation and requirements for global convergence

The problem of obtaining a DC operating point is equivalent to that of finding a solution of the system of circuit equations, denoted here by

$$\bar{g}(\bar{x}) = \bar{0}. \tag{9}$$

Here $\bar{x}$ is the vector of circuit unknowns (node voltages, branch currents, etc). The system of equations for the circuit may be in any formulation; standard ones are the modified nodal analysis formulation (MNA) and the tableau formulation. In this work, the MNA formulation is used. We denote by n the size of the system of equations and unknowns.

The use of numerical continuation requires an additional scalar parameter, denoted λ, to be embedded into the above system of equations, *i.e.*,

$$\bar{f}(\bar{x},\lambda) = \bar{0}. \tag{10}$$

Note that $\bar{f}(\bar{x},\lambda) = \bar{0}$ is a system of n equations in $n+1$ unknowns. The object of the continuation algorithm is to find a solution to $\bar{f}(\bar{x},1) \equiv \bar{g}(\bar{x}) = \bar{0}$. In doing so, the algorithm will use the fact that $\bar{f}(\bar{x},0) = \bar{0}$ is easy to solve and exploit smoothness properties of $\bar{f}$ to be stated at the end of the section.

The embedding of λ into $\bar{g}$ must be done such that certain conditions are satisfied. Two key constraints are that the solution of $\bar{f}(\bar{x},0) = \bar{0}$ be easy to obtain and unique; and that $\bar{f}(\bar{x},1)$ reduce to the original, presumably difficult-to-solve, system in Equation 9, $\bar{g}(\bar{x})$. The nature of the dependence of $\bar{f}$ on λ is the subject of Section V, where the MOS homotopy is described.

Following [1], we introduce some notation. Denote by J the Jacobian matrix of $\bar{f}(\bar{x},\lambda)$, i.e.,

$$J(\bar{x},\lambda) = \left.\frac{\partial \bar{f}}{\partial(\bar{x},\lambda)}\right|_{(\bar{x},\lambda)}. \tag{11}$$

Note that J is an $n \times (n+1)$ matrix.

The rank of J is important in numerical continuation. J is said to be of *full rank* if its rank is the maximum possible for a matrix of its dimensions, i.e., n. If $\text{rank}(J) < n$, then J is said to be *rank deficient*. It will be seen at the end of the section that J will be of full rank with probability 1.

If the $n \times n+1$ matrix J is of full rank, then its null-space is of dimension one. It follows that there are exactly two vectors in this null-space that have unit norm, denoted by $\bar{t}(J)$ and $-\bar{t}(J)$ and termed *tangent vectors* of J. We emphasize that $\bar{t}(J)$ is not defined if J is rank deficient.

At $\lambda = 0$, the equation $\bar{f}(\bar{x},\lambda) = \bar{0}$ is called the *start system*. As mentioned already, the start system should be designed to be easy to solve and have a unique solution. There are further conditions, stated later, that this solution must also satisfy. Denote the unique solution of the start system by $\bar{x}_0$ and denote $J(\bar{x}_0,0)$ by J_0. Of the two tangent vectors $\bar{t}(J_0)$ and $-\bar{t}(J_0)$ of the start system, select the one that has a positive λ-component (i.e., the last entry is positive) . Denote this as $\bar{t}_0$. We will refer to $\bar{t}_0$ as *the* tangent vector of the start system; in other words, from two tangent vectors for J_0, a unique one has been defined.

The unique specification of the tangent vector is extended to other $J \neq J_0$ as follows. The $(n+1)\times(n+1)$ matrix $J_{0a} = \begin{bmatrix} J_0 \\ \bar{t}_0^T \end{bmatrix}$ is formed. The sign of the determinant of J_{0a} is computed (note that $\det J_{0a} \neq 0$ because $\bar{t}_0$ is orthogonal to and linearly independent of all the rows of J_0). Given any J with $\bar{t}$ and $-\bar{t}$ its tangent vectors, form $J_a = \begin{bmatrix} J \\ \bar{t}^T \end{bmatrix}$. If $\text{sign}(\det J_a) = \text{sign}(\det J_{0a})$, $\bar{t}$ is defined to be the unique tangent vector of J; otherwise, $-\bar{t}$ is defined to be the unique tangent vector. In this manner, a unique tangent vector $\bar{t}(J)$ has been defined for any J.

Some useful properties of $\bar{t}(J)$ are now stated. Proofs may be found in [1].

Property 1: $\bar{t}(J)$ is a smooth function of J, and hence of $(\bar{x},\lambda)$.

Property 2: $\bar{t}(J)$ is orthogonal to (and linearly independent of) all the rows of J, i.e., $J\bar{t}(J) = \bar{0}$ (this follows immediately from the definition of $\bar{t}(J)$).

The unique tangent vector defined above is of central importance in the arclength continuation algorithm, which is based on the following *defining ODE (ordinary differential equation)*:

$$\frac{\partial(\bar{x},\lambda)}{\partial s} = \bar{t}(J((\bar{x},\lambda))) \quad \text{with} \quad (\bar{x},\lambda)|_{s=0} = (\bar{x}_0,0). \tag{12}$$

That the above equation is well-defined is guaranteed by the unique definition of $\bar{t}(J((\bar{x},\lambda)))$. That its solution exists and is unique is a consequence of the smoothness of $\bar{t}$. The scalar

parameter s with respect to which differentiation is being performed does not have any direct physical significance (unlike time in ODEs governing time-evolution of systems). It can however be shown that its value is always equal to the total length of the solution track from the start of the homotopy $(\bar{x}_0, 0)$ to any point $(\bar{x}(s), \lambda(s))$ on it – hence the term *arclength* continuation.

The utility of Equation 12 stems from the following result:

Theorem: The solution $(\bar{x}(s), \lambda(s))$ of Equation 12 satisfies $\bar{f}(\bar{x}(s), \lambda(s)) = \bar{0}$ for all s.

Proof:

$$\frac{\partial \bar{f}}{\partial s} = \frac{\partial \bar{f}}{\partial (\bar{x}, \lambda)} \frac{\partial (\bar{x}, \lambda)}{\partial s} = J(\bar{x}(s))\bar{t}(J(\bar{x}(s))) = \bar{0}. \qquad (13)$$

Hence $\bar{f}(s) = \text{constant}$. Since $\bar{f}(s=0) = \bar{f}(\bar{x}_0, 0) = \bar{0}$ from the initial condition, the theorem is proved. ■

The above theorem shows that the solution of the initial-value problem in Equation 12 is also a solution of the augmented nonlinear system in Equation 10. Numerical integration methods for initial value problems (*e.g.*, [17]) can thus be applied to finding this solution. An important question that remains, however, is whether the solution track $(\bar{x}(s), \lambda(s))$ ever reaches a point where $\lambda(s) = 1$, *i.e.*, a point where $\bar{x}(s)$ is the solution of the original problem $\bar{g}(\bar{x}) = \bar{0}$.

That this is indeed the case is based on the fact that the defining ODE of Equation 12 has a unique solution for *all* values of s, a fact that can be shown to follow from the RHS $\bar{t}$ being of norm 1. As already explained in Section III and Figure 2, several types of solution can be eliminated. That the solution cannot diverge to infinity is guaranteed by the assumption that all solutions of $\bar{f}(\bar{x}, \lambda) = 0$, for all $\lambda \in [0, 1]$ are bounded (this can be shown to be a consequence of the no-gain property [11]); that it cannot reach a limit point by the fact that the RHS of Equation 12 always has norm 1; that it cannot reach $\lambda < 0$ by the uniqueness of the solution $\bar{x}_0$ of the start system; that the solution cannot loop as in the fourth track from the top in Figure 2 is a consequence of the assumed structural stability of the start system. That the solution must cross $\lambda = 1$ is a consequence of the fact that no limit points of the solution can exist and that all solutions of $\bar{f}$ have been assumed to be bounded. Provided the defining ODE is well-defined, the only remaining possibility is the lowermost track of Figure 2, which must cross the $\lambda = 1$ axis (otherwise the track would necessarily have a limit point because it would be confined forever in a bounded region).

Finally, we collect together all the assumptions required for arclength continuation to provide theoretically guaranteed global convergence:

• *$\bar{x}_0$, the solution of the start system, is easy to solve for and is unique*: This requirement is easily fulfilled because it can be ensured while constructing $\bar{f}$. It is met for the simple homotopy of Section III as well as for the ATANSH homotopy of Section V.

• *The start system $\bar{f}(\bar{x}, 0) = 0$ is structurally stable*: The precise requirement is that there is no turning point at $\lambda = 0$, equivalent to the $n \times n$ Jacobian $\left.\frac{\partial \bar{f}}{\partial \bar{x}}\right|_{(\bar{x}_0, 0)}$ being of full rank. This is needed in order to avoid loops in the solution track. This condition can be ensured easily by the proper construction of $\bar{f}$.

• *$\bar{f}(\bar{x}_0, \lambda)$ is a smooth function of $(\bar{x}_0, \lambda)$*: The precise requirement is that $\bar{f}$ be C^2. That it needs to be at least C^1 is dictated by the smoothness requirement of $\bar{t}(\bar{x}_0, \lambda)$, important for global convergence. The C^2 requirement is required by Sard's theorem which is used in the next assumption.

• *0 is a* regular value *of $\bar{f}(\bar{x}, \lambda)$*: $\bar{b}$ is defined to be a *regular value* of $\bar{f}(\bar{x}, \lambda)$ if at *all* solutions of $\bar{f}(\bar{x}, \lambda) = \bar{b}$, the Jacobian $J(\bar{x}, \lambda)$ has full rank. Sard's theorem [1, 21] states that if $\bar{f}$ is C^2, *almost all* values $\bar{b}$ of $\bar{f}$ are regular values. This provides the justification for the assumption that J is full-rank. In the unlikely event of 0 not being a regular value of $\bar{f}$, this theorem provides the basis for expecting a random perturbation to restore 0 as a regular value of the system.

• *All solutions of $\bar{f}(\bar{x}, \lambda) = 0$ are bounded for $\lambda \in [0, 1]$*: This is required to ensure that the solution track remains within a bounded region, key to the establishment of global convergence. It is met for well-defined circuits and realistic models that satisfy no-gain properties [11].