A Connection Between the Convergence Properties of Waveform Relaxation and the A-stability of Multirate Integration Methods

J. White
Research Laboratory of Electronics
Dept. of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139

F. Odeh
I.B.M. T. J. Watson Research Center
Yorktown Heights, N. Y.

Abstract

Application of waveform relaxation to semiconductor device transient simulation has demonstrated encouraging results. In particular, empirical observations suggest that the relaxation is a contraction with respect to the *sup*-norm in time. In addition, the underlying multirate integration method has not exhibited instabilities. In this paper we prove that the two properties are connected, and use the result to show that the first and second-order backward-difference based multirate methods are A-stable.

1 Introduction

When standard spatial discretization techniques are applied to solving the classical drift-diffusion based differential equation system used to model semiconductors, the result, in the time-dependent case, is a large sparsely-connected system of algebraic and differential equations. The so-generated system can be solved by waveform relaxation(WR), an iterative method whose eventual convergence is guaranteed in this case. Experiments with WR indicate that it is efficient for two-dimensional simulation of transients in MOS devices, and can be as much as an order of magnitude faster than more commonly used direct methods[1][5].

As the WR algorithm is an iterative technique which directly decomposes systems of differential equations into subsystems which are solved independently, when WR is used to simulate a semiconductor device, different sections of the device can be integrated in time with different discretization timesteps. This *multirate integration* property of WR is one of its main advantages over direct methods or "point"-relaxation schemes.

Examination of the results from numerical experiments using WR to simulate semiconductor devices yeilds two surprising observations: the error waveforms produced by iterations of the WR algorithm contract in a *sup*-norm in time, and there does not seem to be evidence of instabilities introduced by the multirate integration. There have been separate theoretical investigations into the reasons for *sup*-norm convergence for both linear and nonlinear model problems[2][4], and some investigation into criteria for multirate stability[6]. In this short paper, we show that the two properties are connected, and use that result to prove that the second-order backward difference method is multirate A-stable for diagonally-dominant problems.

[0]This work was supported by the Defense Advanced Research Projects Agency contract N00014-87-K-825, the National Science Foundation contract (MIP-8858764 A02) and grants from I.B.M..

We start below with basic definitions. In section 3, we prove that under certain conditions convergence on the infinite interval of both the exact and discrete WR algorithm imply multirate A-stability. In section 4 we present a discretized WR convergence theorem for the multirate case, and finally in section 5 we present our conclusions.

2 Definitions

For purposes of analysis, we consider solving the linear time-invariant problem

$$\dot{x}(t) = Ax(t) \quad x(0) = x_0 \tag{1}$$

where $t \in [0,\infty)$, $x(t) \in \Re^n$, and $A \in \Re^{n\times n}$ is assumed strictly diagonally dominant with negative diagonal entries. That is, there exists some $\epsilon > 0$ such that

$$-a_{ii} \geq \epsilon + \sum_{j\neq i} |a_{ij}| \tag{2}$$

for all $i \in \{1,...,n\}$, where a_{ij} is the ij^{th} element in the matrix A.

When Gauss-Jacobi WR is applied to solving (1), the iteration update equation for x^{k+1} given x^k is

$$\dot{x}_i^{k+1}(t) = a_{ii}x_i^{k+1}(t) + \sum_{j\neq i} a_{ij}x_j^k(t). \tag{3}$$

where the superscript k is the WR iteration index, $i \in \{1,...,n\}$ is the component index, and $x^{k+1}(0) = x_0$. With the assumptions on A above, it is relatively simple to show that WR is a contraction in a *sup*-norm[7]. That is, given any two arbitrary waveforms x^k and y^k which match the initial conditions of (1), when (3) is applied to compute x^{k+1} and y^{k+1} the following inequality holds:

$$\sup_t \|x^{k+1}(t) - y^{k+1}(t)\| < \gamma \sup_t \|x^k(t) - y^k(t)\| \tag{4}$$

where $\gamma < 1$ and is independent of x^k and y^k. For a general A, it is possible to show that the WR algorithm is a contraction in an exponentially weighted norm, but this only implies that the WR iterates converge to the solution of (1) on finite time intervals.

Assuming (3) is solved using a fixed-timestep backward-difference formula, the so-generated discretized WR iteration update equation becomes

$$\sum_{l=0}^{p} \alpha_l x_i^{k+1}[m-l] = h_i a_{ii} x_i^{k+1}[m] + h_i \sum_{j\neq i} a_{ij} I_{mh_i}\left(\{x_j^k\}\right) \tag{5}$$

where p is the order of the integration method, the α_l's are the integration method coefficients, h_i is the timestep used to compute the component x_i, $x_i^{k+1}[m]$ is the discrete approximation to $x_i^{k+1}(mh_i)$ which is assumed to be identically zero for $m < 0$, and I_t is an interpolation operator that maps $t \in \Re^+$ and a sequence, denoted with $\{\cdot\}$, into $\Re$. The discretized WR algorithm represented by (5) allows different timesteps to be used to compute different components of x, but for simplicity we have assumed that for the integration of a given component x_i, the timestep does not change.

If the iteration in (5) converges, the resulting sequence $\{x\}$ will satisfy

$$\sum_{l=0}^{p} \alpha_l x_i[m-l] = h_i a_{ii} x_i[m] + h_i \sum_{j \neq i} a_{ij} I_{mh_i}(\{x_j\}) \qquad (6)$$

for $i \in \{1, ..., n\}$, and this discrete system is said to be a multirate integration method for solving (1). Note that the interpolation operator in (6) must be chosen carefully, as the choice can effect both the stability and accuracy of the multirate method. Also, as the result of interpolation is used only in the calculation of approximations to $\dot{x}_i$, and is effectively always multiplied by the timestep, the asymptotic error of the interpolation operator can be one order lower than the desired order for the multirate local truncation error.

Following from the usual definition of A-stability for an integration method, we say a multirate method of the form of (6) is component-by-component A-stable for a given $A \in \Re^{n \times n}$ if the solution to

$$\sum_{l=0}^{p} \alpha_l x_i[m-l] = h_i a_{ii} x_i[m] \qquad (7)$$

is such that $\lim_{m \to \infty} x_i[m] = 0$ for any initial condition and any positive h_i. We say the multirate integration method is A-stable for a given $A \in \Re^{n \times n}$ if for any set of positive timesteps $\{h_1, ..., h_n\}$ and any initial conditions

$$\lim_{m \to \infty} x_i[m] = 0 \qquad (8)$$

for all $i \in \{1, ..., n\}$. Clearly, these definitions are only appropriate when the matrix A is stable (has eigenvalues in the open left-half of the complex plane).

3 Main Theorem

The following theorem, which connects WR convergence properties to multirate A-stability properties, is the main result of this paper.

Theorem 1 *A multirate integration method of the form of (6) is A-stable for any stable $A \in \Re^{n \times n}$ which has negative diagonals if the multirate method is component-by-component A-stable, and if the iterations defined by (3) and (5) converge on the infinite interval to the exact solution of (1) and (6) respectively, uniformly in t and m respectively.*

Proof of Theorem 1: The assumptions of Theorem 1 imply that both the WR and the discretized WR algorithms converge to the exact solutions of (1) and (6) respectively, given any initial guess which matches the initial conditions. Therefore, the theorem can be proved by selecting a particular initial guess waveform and showing that starting from this initial guess the discretized WR iterates converge to a solution in which

$$\lim_{m \to \infty} x_i[m] = 0. \qquad (9)$$

To begin, let $x^0(t) = x_0$ for all $t \in [0, \infty)$ and $x^0[m] = x_0$ for all $m \in \{1, 2, ..., \infty\}$. Trivially,

$$\lim_{t \to \infty} x_i^0(t) = x_0. \qquad (10)$$

Using (10) as the starting point for an inductive argument, it is possible to show that the iteration update equation (3) combined the assumption that the diagonals of A are negative imply that

$$x_i^k(\infty) \equiv \lim_{t \to \infty} x_i^k(t) \qquad (11)$$

exists for all k. If A is stable, then the exact solution to (1) goes to zero as $t \to \infty$, and therefore

$$\lim_{k \to \infty} x_i^k(\infty) = 0 \qquad (12)$$

as WR is assumed to converge on the infinite interval.

It is not hard to show that if the multirate integration method is component-by-component A-stable,

$$\lim_{m \to \infty} x_i^k[m] = x_i^k(\infty). \qquad (13)$$

We omit the detailed proof of (13) for brevity, but roughly the argument involves showing that when $x_i^k(t)$ in (3) approaches a limit, then the solution to (5) computed with an A-stable integration method eventually approaches exactly that same limit, regardless of the timesteps used.

Taking the limits of both sides of (13) yields

$$\lim_{k \to \infty} \lim_{m \to \infty} x_i^k[m] = 0, \qquad (14)$$

as the discretized WR iterates converge uniformly on the infinite interval. From this assumption it also follows that the limits in (14) can be interchanged resulting in

$$\lim_{m \to \infty} \lim_{k \to \infty} x_i^k[m] = \lim_{m \to \infty} x_i[m] = 0 \qquad (15)$$

which proves the theorem ■.

4 Discretized Convergence and A-stability Theorems

In this section we show that a multirate method which combines linear interpolation with either the first- or second-order backward difference formula is multirate A-stable for all $A \in \Re^{n \times n}$ which satisfy the diagonal dominance condition in (2). To accomplish this, we prove below that our assumptions imply the discretized WR algorithm converges uniformly on the infinite interval. Given such a result, Theorem 1 above can then be applied to prove multirate A-stability.

We start the discretized WR convergence proof with the definition of a specialized norm, followed by a precise statement of this section's main theorem.

Definition 1 *Given an infinite sequence $\{x\}$ on $\Re$, the quantity*

$$|||\{x\}|||_h \equiv \sqrt{\frac{1}{h} \sum_{n=1}^{\infty} (x[m] - x[m-1])^2}. \qquad (16)$$

is a norm for any positive h if the sequences considered are restricted to those for which $x[0] = 0$. In addition, if

$$|||\{x\}|||_1 < \infty, \tag{17}$$

we say $\{x\}$ has bounded variation.

Theorem 2 *Let $A \in \Re^{n\times n}$ in (1) satisfy the condition in (2). If linear interpolation is used in (5), and the α_l's in (5) correspond to the first or second-order backward difference formula, then there exists a $\gamma < 1$ such that for any two sets of real sequences with bounded variation, $\{x_i^k\}$, $\{y_i^k\}$, $i \in \{1,...,n\}$, such that $x^k[0] = y^k[0] = x_0$,*

$$\max_{i\in\{1,\ldots,n\}} |||\{x_i^{k+1} - y_i^{k+1}\}|||_{h_i} \leq \gamma \max_{i\in\{1,\ldots,n\}} |||\{x_i^k - y_i^k\}|||_{h_i}. \tag{18}$$

where x_i^{k+1} and y_i^{k+1} are derived from (5).

Proof of Theorem 2: Given $\{x^k\}$ and $\{y^k\}$, the difference between $\{x^{k+1}\}$ and $\{y^{k+1}\}$ can be derived from (5) and is

$$\sum_{l=0}^{p} \alpha_l \delta_i^{k+1}[m-l] = h_i a_{ii} \delta_i^{k+1}[m] + h_i \sum_{j\neq i} a_{ij} I_{mh_i}\left(\{\delta_j^k\}\right) \tag{19}$$

where $\delta_i^k[m] \equiv x_i^k[m] - y_i^k[m]$. Note that the linearity of the interpolation operator has been exploited.

Subtracting (19) at $m-1$ from (19) at m yields

$$\sum_{l=0}^{p} \alpha_l(\delta_i^{k+1}[m-l] - \delta_i^{k+1}[m-l-1]) = h_i a_{ii}(\delta_i^{k+1}[m] - \delta_i^{k+1}[m-l-1]) + h_i \sum_{j\neq i} a_{ij}\left(I_{mh_i}\left(\{\delta_j^k\}\right) - I_{(m-1)h_i}\left(\{\delta_j^k\}\right)\right) \tag{20}$$

The z-transform of (20) is

$$\left(\alpha_0 - h_i a_{ii} + \sum_{l=1}^{p} \alpha_l z^{-l}\right) \mathcal{Z}(\delta_i^{k+1}[m] - \delta_i^{k+1}[m-1]) = h_i \sum_{j\neq i} a_{ij} \mathcal{Z}\left(I_{mh_i}\left(\{\delta_j^k\}\right) - I_{(m-1)h_i}\left(\{\delta_j^k\}\right)\right) \tag{21}$$

which can be reorganized as

$$\mathcal{Z}(\delta_i^{k+1}[m] - \delta_i^{k+1}[m-1]) = \left(\alpha_0 - h_i a_{ii} + \sum_{l=1}^{p} \alpha_l z^{-l}\right)^{-1} h_i \sum_{j\neq i} a_{ij} \mathcal{Z}\left(I_{mh_i}\left(\{\delta_j^k\}\right) - I_{(m-1)h_i}\left(\{\delta_j^k\}\right)\right). \tag{22}$$

Taking the l_2 norm of both sides and applying the Cauchy-Schwarz inequality leads to

$$\|\mathcal{Z}(\delta_i^{k+1}[m] - \delta_i^{k+1}[m-1])\|_2 \leq \left\|\left(\alpha_0 - h_i a_{ii} + \sum_{l=1}^{p} \alpha_l z^{-l}\right)^{-1}\right\|_2 + h_i \sum_{j\neq i} a_{ij} \|\mathcal{Z}\left(I_{mh_i}\left(\{\delta_j^k\}\right) - I_{(m-1)h_i}\left(\{\delta_j^k\}\right)\right)\|_2. \tag{23}$$

Parseval's relation then implies

$$\sqrt{\sum_{m=0}^{\infty} |\delta_i^{k+1}[m] - \delta_i^{k+1}[m-1]|^2} \leq M h_i \sum_{j\neq i} |a_{ij}| \sqrt{\sum_{m=0}^{\infty} \left|I_{mh_i}\left(\{\delta_j^k\}\right) - I_{(m-1)h_i}\left(\{\delta_j^k\}\right)\right|^2} \tag{24}$$

where

$$M = \max_{|z|=1} \left|\left(\alpha_0 - h_i a_{ii} + \sum_{l=1}^{p} \alpha_l z^{-l}\right)^{-1}\right| \tag{25}$$

By definition, $\delta_i^{k+1}[0] = 0$, and therefore can be measured using the norm defined in (16). Equation (24) can be expressed in terms of that norm,

$$|||\{\delta_i^{k+1}\}|||_{h_i} \leq M h_i \sum_{j\neq i} |a_{ij}| \; |||\{I_{mh_i}\left(\{\delta_j^k\}\right)\}|||_{h_i}. \tag{26}$$

where here the term $\{I_{mh_i}(\{\delta_j^k\})\}$ is used, somewhat cryptically, to denote the sequence whose m^{th} element is given by $I_{mh_i}(\{\delta_j^k\})$.

From Lemma 1 in the appendix, in particular from equation (38), it follows that if I_t is linear interpolation then

$$|||\{I_{mh_i}\left(\{\delta_j^k\}\right)|||_{h_i} \leq |||\{\delta_j^k\}|||_{h_j}. \tag{27}$$

Therefore, it is sufficient to prove the theorem to show that

$$M = \max_{|z|=1} \frac{h_i \sum_{j\neq i} |a_{ij}|}{|\alpha_0 - h_i a_{ii} + \sum_{l=1}^{p} \alpha_l z^{-l}|} < 1. \tag{28}$$

for all $i \in \{1,...,n\}$. To see this, consider that (27) and (28) imply

$$|||\{\delta_i^{k+1}\}|||_{h_i} \leq \max_{j\in\{1,\ldots,n\}} |||\delta_j^k|||_{h_j}. \tag{29}$$

from which (18) follows directly.

To finally prove the theorem, we need demonstrate only that the inequality in (28) holds for the α_l's corresponding to first- and second-order backward difference integration methods. To show (28) for the the first-order method, also know as backward-Euler, we must determine that

$$\max_{|z|=1} \frac{h_i \sum_{j\neq i} |a_{ij}|}{\left|1 - h_i a_{ii} + \frac{1}{z}\right|} \tag{30}$$

is less than one. As a_{ii} is negative and real by assumption,

$$\min_{|z|=1} \left|1 - h_i a_{ii} + \frac{1}{z}\right| = h_i |a_{ii}|. \tag{31}$$

Then

$$\max_{|z|=1} \frac{h_i \sum_{j\neq i} |a_{ij}|}{\left|1 - h_i a_{ii} + \frac{1}{z}\right|} \leq \frac{\sum_{j\neq i} |a_{ij}|}{|a_{ii}|} < 1 \tag{32}$$

as A was assumed to be diagonally dominant.

To prove the theorem for the second-order backward-difference method, we need to show

$$\max_{|z|=1} \frac{h_i \sum_{j\neq i} |a_{ij}|}{\left|\frac{3}{2} - h_i a_{ii} - \frac{2}{z} + \frac{1}{2z^2}\right|} \tag{33}$$

is less than one. This follows as before from the fact that

$$\min_{|z|=1} \left|\frac{3}{2} - h_i a_{ii} - \frac{2}{z} + \frac{1}{2z^2}\right| = h_i |a_{ii}| \tag{34}$$

which, finally, proves the theorem ■.

5 Conclusions

In this paper we demonstrate that the WR algorithm is not just an efficient numerical technique for device simulation, but is also a useful analytic tool for analyzing multirate integration methods. The authors wish to thank M. Reichelt for the experimental results that led to this paper, A. Elfadel for a valuable discussion on Lemma 1, and M. Silveira and K. Nabors for proof-reading the manuscript under duress.

References

[1] R.E. Bank, W.C. Coughran Jr. W. Fichtner, E. H. Grosse, D.J. Rose, and R.K. Smith, "Transient Simulation of Silicon Devices and Circuits", IEEE Transactions on Computer-Aided Design, Vol. CAD-4, No. 4, October 1985, pp. 436-451.

[2] U. Miekkala and O. Nevanlinna, "Convergence of Dynamic Iteration Methods for Initial Value Problems," SIAM J. Sci. Stat. Comput. 8(1987), 459-482.

[3] E. Lelarasmee, A. Ruehli, A. Sangiovanni-Vincentelli, "The Waveform Relaxation Method for the Time Domain Analysis of Large Scale Integrated Circuits." *IEEE Trans. on CAD*, Vol. 1, No. 3, July 1982.

[4] M. Reichelt, J. White, J. Allen and F. Odeh, "Waveform Relaxation Applied to Transient Device Simulation," *1988 Int'l. Symp. on Circuits and Systems*, Espoo, Finland.

[5] M. Reichelt, J. White, J. Allen, "Waveform Relaxation for Transient Simulation of Two-Dimensional MOS Devices," *Proc. Int. Conf. on Computer-Aided Design*, Santa Clara, California, October 1989, p412-415.

[6] S. Skelboe and P. U. Andersen, "Stability Properties of Backward-Euler Multirate Formulas," *SIAM J. Sci. Stat. Comput.*, Vol. 10, No. 5, September 1989.

[7] J. White and A.S. Vincentelli, **Relaxation Techniques for the Simulation of VLSI Circuits**, Kluwer Academic Publishers, 1986.

6 Appendix I - Linear Interpolation Lemma

Using a multirate integration method implies that different partitions of a system are integrated with different timesteps, and therefore variables computed in one partition and used in another must be somehow translated. The translation used in the algorithms presented above can be thought of as first interpolating the discrete representation of the variable into a continuous function of time, and then evaluating that continuous function at times associated with the timesteps in other partitions. In this way, the discrete sequence computed in one partition can be converted into a discrete sequence usable by another partition.

In this appendix, we prove that if linear interpolation is used, then the translation process described above doesn't magnify a particular norm on the discrete sequence, given a scaling. In particular, we will show that if $\{x\}$ is the sequence computed using a fixed timestep h_x, and $\{y\}$ is the sequence produced by translating $\{x\}$ with linear interpolation so as to use the fixed timestep h_y, then

$$\frac{1}{h_y}\sum_{n=1}^{\infty}(y[n]-y[n-1])^2 \leq \frac{1}{h_x}\sum_{m=1}^{\infty}(x[m]-x[m-1])^2 \quad (35)$$

The formal statement of the Lemma follows.

Lemma 1 *Given a sequence* $\{x\} : 0,1,2,\ldots \rightarrow \Re$ *for which*

$$\sum_{m=1}^{\infty}(x[m]-x[m-1])^2 < \infty, \quad (36)$$

let the sequence $\{y\}$ *be derived from* $\{x\}$ *by*

$$y[n] = \gamma x[m] + (1-\gamma)x[m-1], \quad (37)$$

where $m = ceil(nr)$, $\gamma = nr - (m-1)$, *and* r *is an aribtrary positive number (Note that in (35),* r *would be given by* $\frac{h_y}{h_x}$*). For any* $r \in \Re^+$, $\{y\}$ *satisfies the following inequality:*

$$\sum_{n=1}^{\infty}(y[n]-y[n-1])^2 \leq r\sum_{m=1}^{\infty}(x[m]-x[m-1])^2 \quad (38)$$

Proof: Let $x : [0,\infty) \rightarrow \Re$ be a a continuous and differentiable function defined by

$$x(t) = \gamma x[m] + (1-\gamma)x[m-1] \quad (39)$$

where $m = ceil(t)$, $\gamma = t-(m-1)$. Then by definition,

$$\int_0^{\infty}\left(\frac{d}{dt}x(t)\right)^2 = \sum_{m=1}^{\infty}(x[m]-x[m-1])^2 \quad (40)$$

and

$$y[n] = x(nr) \quad (41)$$

Given (41) and that x is differentiable,

$$(y[n]-y[n-1])^2 = \left[\int_{(n-1)r}^{nr}\left(\frac{d}{dt}x(t)\right)\right]^2. \quad (42)$$

It can be derived from Schwarz's inequality that

$$\left[\int_{(n-1)r}^{nr}\left(\frac{d}{dt}x(t)\right)\right]^2 \leq r\left[\int_{(n-1)r}^{nr}\left(\frac{d}{dt}x(t)\right)^2\right], \quad (43)$$

and therefore

$$(y[n]-y[n-1])^2 =\leq r\left[\int_{(n-1)r}^{nr}\left(\frac{d}{dt}x(t)\right)^2\right]. \quad (44)$$

Summing (44) over n leads to

$$\sum_{n=1}^{\infty}(y[n]-y[n-1])^2 \leq r\sum_{n=1}^{\infty}\left[\int_{(n-1)r}^{nr}\left(\frac{d}{dt}x(t)\right)^2\right] \quad (45)$$

and as

$$r\sum_{n=1}^{\infty}\left[\int_{(n-1)r}^{nr}\left(\frac{d}{dt}x(t)\right)^2\right] = r\sum_{m=1}^{\infty}(x[m]-x[m-1])^2, \quad (46)$$

this proves the lemma ■.