A New Tightly-Coupled Transient Electro-Thermal Simulation Method for Power Electronics

Quan Chen
Department of Electrical and Electronic Engineering
The University of Hong Kong
Pokfulam, Hong Kong
quanchen@eee.hku.hk

Wim Schoenmaker
Magwel N.V.
Vital Decosterstraat 44 bus 27
3000 Leuven, Belgium
wim.schoenmaker@magwel.com

ABSTRACT

This paper presents a new transient electro-thermal (ET) simulation method for fast 3D chip-level analysis of power electronics with field solver accuracy. The metallization stacks are meshed and solved with 3D field solver using nonlinear temperature-dependent parameters, and the active devices are modeled with nonlinear tabular compact models to avoid time-consuming TCAD simulation. The main contributions include: 1) A tightly-coupled formulation that solves the electrical and thermal systems simultaneously to enable a more natural characterization of ET interaction and faster convergence for strong ET coupling; 2) A specialized exponential-integrator-Newton-Krylov (EI-NK) transient solver to address the numerical challenges arising from the tightly-coupled ET simulation. Numerical experiments show that the proposed method outperforms the existing implicit time stepping with the Newton's method and the exponential integrator with an explicit nonlinear approximation.

1. INTRODUCTION

Bipolar-CMOS-DMOS (BCD) integration is a key technology for power integrated circuits (ICs) offering many advantages by integrating three distinct types of devices on a single die. New challenges, however, have also been triggered to the thermal managements in the BCD technology due to the closer proximity of high-power DMOS transistors to other temperature-sensitive components and the more complicated geometry and material configurations. Accurate prediction in temperature profile is needed to guide heat removal design and avoid potential reliability issues such as electromigration and negative bias temperature instability (NBTI). To this end, the strong coupling between electrical and thermal dynamics, e.g., the nonlinear temperature dependencies of electrical parameters and device characteristics [9], must be appropriately accounted. In addition, DMOS is often used as switches in BCD and operates under pulse inputs. The peak temperature of devices can go up and down in one pulse period and exceed the maximum threshold momentarily causing permanent damage, which cannot be detected by steady-state analysis [2]. Therefore, an accurate transient electro-thermal (ET) co-simulation is highly desired to detect and avoid thermal failures in the early stage of BCD designs.

Most existing transient ET, and in a broader sense multi-physics, simulation methods fall into the "loosely-coupled" category. The electrical and thermal dynamics are treated as two uni-physics problems, solved separately by specialized methods, such as field solution or equivalent network analysis, and coupled via communication between the two solvers through an appropriate interface at software level [4, 7]. Despite the methodological simplicity, the loosely-coupled strategy requires a careful bookkeeping of the interaction between electrical and thermal variables, and a complicated software communication scheme to ensure efficient information exchange [10]. The deliberate split of electrical and thermal dynamics may also lead to slow convergence when the their coupling is strong. More importantly, modern high-performance computing architecture has entered an era that data motion is "penalized" while flops are offered nearly "for free". The frequent data exchange and relatively low arithmetic intensity required in loosely-coupled schemes are actually opposite to this hardware evolution direction [5]. Alternatively, a tightly-coupled ET simulation assembles the electrical and thermal components in one single numerical system and solves them simultaneously. The concurrent treatment to multiple physics generally leads to more natural and consistent characterization of physical interactions, and subsequently to more automated analysis, less data communication, and faster convergence for strong coupling.

In the context of ET analysis for power ICs, a tightly-coupled simulation method was proposed in [10] which solves the electrical and thermal equations simultaneously in the time-domain. Being one step closer to the "ab initio" simulation, the ET solver in [10] can determine the voltage drop in the metal structures and the device temperatures with high accuracy and without limiting the applicability to special cases. The limitation of [10] is that the capacitive effects in the electrical system are neglected, i.e., only the thermal system is actually evolved in time with the electrical response determined solely by instantaneous boundary condition. This limitation was removed in [8] in which the interconnect and device gate capacitance is included in the formulation of electrical system, and an exponential integrator (EI) method was developed to speed up the transient electrical simulation. Yet [8] still used a loosely-coupled scheme to handle the ET coupling.

While tightly-coupled approaches are promising for ET problems and, in a wider perspective, multi-physics problems, some numerical challenges need to be properly addressed for the tightly-coupled approaches to be a better strategy in practice. These challenges include presence of multiple time scales (stiff systems), strong nonlinearity and considerably increased problem size. The large

ICCAD '16 November 07-10, 2016, Austin, TX, USA

DOI: http://dx.doi.org/10.1145/2966986.2966993.

time scale difference demands implicit time stepping methods such as the backward differentiation formula (BDF), the high nonlinearity favors the Newton's method over the fixed-point iteration underlying loosely-coupled approaches to achieve better convergence, and the elevated system size renders direct matrix factorization less feasible. As a consequence, BDF combined with the Newton-Krylov (NK) method, which uses iterative methods based on the Krylov subspace to solve the Jacobian matrix, is currently the preferred choice for large-scale evolutionary multi-physics problems [6]. However, the ET coupling considered in this work exhibits a particularly strong stiffness, with the thermal time constant larger than the electrical one by 3 to 6 orders of magnitude. Temporal integration using BDF usually results in unnecessarily small time steps, and ill-conditioned Jacobian matrices in the Newton's method that make Krylov-subspace-based iterative methods difficult to converge.

In this paper, we aim to develop a tightly-coupled transient ET simulation method with accelerated convergence for strong coupling and sufficient numerical scalability for large-scale power IC analysis. A nonlinear exponential-integrator-Newton-Krylov (EI-NK) method is developed to enhance numerical scalability with a combination of three components: 1) EI-based time integration to combat stiffness and allow larger time steps than BDF; 2) Newton's method as core nonlinear solver for faster nonlinear convergence; and 3) Faster convergence for the Krylov subspace iterative solvers within the Newton's method. Numerical experiments with a realistic power IC are conducted to demonstrate the desired performance of the proposed ET solver.

2. TIGHTLY-COUPLED ET FORMULATION

2.1 Basic Electrical and Thermal Models

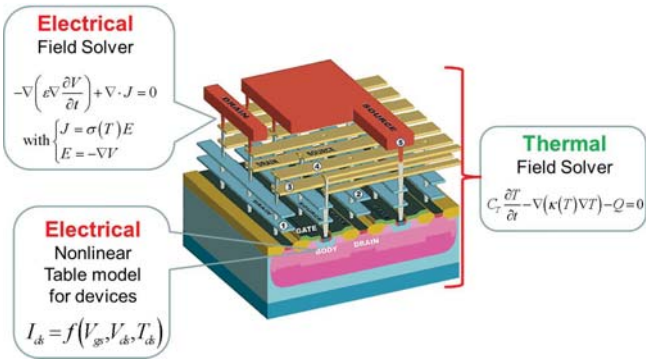

Figure 1: Multi-domain ET modeling (electrical field solver for metallization + compact models for devices + thermal field solver. the substrate is included in the thermal solver but not in the electrical solver)

The transient ET co-simulator developed in this work is based on the multi-domain modeling strategy illustrated in Fig. 1 [8,10]. The electrical system is characterized by the current continuity equation with account for the capacitive effects

$$\left(-\nabla\cdot\left(\varepsilon\nabla\right)+C_d\right)\frac{\partial V}{\partial t}+\nabla\cdot\left(J_c+J_d\right)=0, \tag{1}$$

Here the first term represents the displacement currents and the second term the conduction currents. The displacement currents consist of the contribution from interconnect capacitive parasitics, and that from the device gate capacitances (C_d). For the conduction currents, J_c refers to the conduction currents of the on-chip metallic structures including interconnects, vias and contacts, and J_d the drain-source currents of power devices

$$\begin{aligned}J_c &= -\sigma\left(T\right)\nabla V, \quad \text{conductors}\\ J_d &= f\left(V_{gs},V_{ds},T_{ds}\right), \quad \text{power devices}\end{aligned} \tag{2}$$

In (2), J_c is obtained from the electrical potential V (primary electrical variable) with a field solver using temperature-dependent conductivity $\sigma\left(T\right)$. The device current J_d, on the other hand, is looked up from a temperature-aware table model to avoid expensive TCAD simulation. V_{gs}, V_{ds}, and T_{ds} denote the gate-source voltage, the drain-source voltage and the device temperature averaged at the source and drain terminals, respectively. In this work the device capacitance is assumed temperature independent. For more details of the electrical model we refer the readers to [8] and the references therein.

To model the thermal dynamics, the heat conduction equation is solved on the entire domain with a 3D thermal field solver

$$C_T\frac{\partial T}{\partial t}-\nabla\cdot\left(\kappa(T)\nabla T\right)-Q=0, \tag{3}$$

where T is the temperature (primary thermal variable), $\kappa(T)$ the temperature-dependent thermal conductivity, C_T the volumetric heat capacity and Q the heat sources or sinks. The Q term in (3) has several contributors: 1) the Joule self-heating of the metal structures; 2) the self-heating of the active devices and 3) heat injected or extracted at the boundary of the simulation domain. Similar to [10], the first two contributions are calculated by

$$Q=\begin{cases}E\cdot J_c=\sigma(T)|\nabla V|^2, & \text{conductors}\\ J_d\cdot V_{ds}, & \text{power devices}\end{cases} \tag{4}$$

from the information provided by the electrical solver. The third one is modeled by thermal contacts placed on the domain boundary enforcing fixed temperatures or fixed heat flux.

The temperature-dependent models of the electrical and the thermal conductivities are given by the exponential models

$$\sigma(T)=\sigma_0\left(\frac{T}{T_0}\right)^{\alpha_\sigma}, \tag{5}$$

$$\kappa(T)=\kappa_0\left(\frac{T}{T_0}\right)^{\alpha_\kappa}, \tag{6}$$

where the exponents α_σ and α_κ are material-dependent parameters. Other material parameters are assumed to be temperature-independent.

2.2 Loosely-Coupled and Tightly-Coupled Formulations

In the loosely-coupled formulation, the electrical equation (1) and the thermal equation (3) are solved in a Gauss-Seidel manner [8]: in each thermal time step, the temperature is assumed constant to determine the electrical parameters and the transient electrical response is solved with multiple steps. At the end of the thermal step, the power dissipation of all electrical steps are summed and passed to the thermal solver, which restarts the same thermal step with the updated heat generation and temperature-dependent thermal conductivities to provide a new temperature. The two solvers iterate until self-consistency is achieved.

For weak to medium ET coupling the loosely-coupled scheme usually converges reasonably fast. However, the convergence rate

$$\begin{bmatrix} -\nabla\cdot(\varepsilon\nabla)+C_d & 0 \\ 0 & C_T I \end{bmatrix}\begin{bmatrix}\dot{V}\\ \dot{T}\end{bmatrix} = -\begin{bmatrix} -\nabla\cdot[\sigma(T)\nabla] & 0 \\ 0 & -\nabla\cdot[\kappa(T)\nabla]\end{bmatrix}\begin{bmatrix}V\\T\end{bmatrix} - \begin{bmatrix}J_d(V,T)\\ Q(V,T)\end{bmatrix} - \begin{bmatrix}b^{(E)}\\ b^{(T)}\end{bmatrix} \tag{7}$$

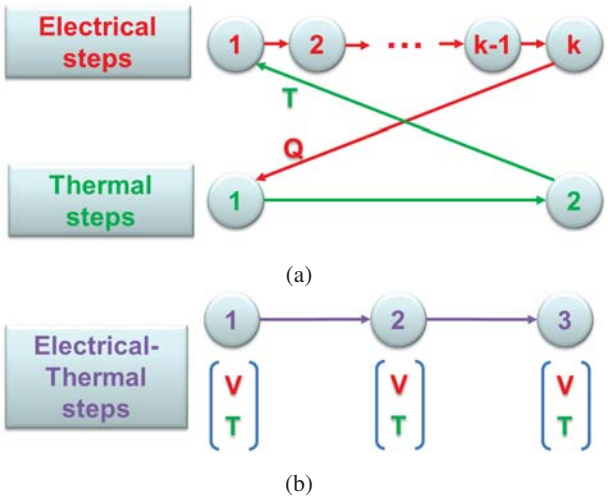

Figure 2: (a) Loosely-coupled scheme. (b) Tightly-coupled scheme.

of loosely-coupled schemes deteriorates rapidly when the ET systems have strong mutual interactions, and the tightly-coupled scheme is preferred in this scenario to accelerate the convergence. The tightly-coupled scheme is formulated by combining the electrical system (1) and the thermal system (3) in one single system as shown in (7) and updating the potential V and temperature T simultaneously in each ET step. Here $b^{(E)}, b^{(T)}$ denote respectively the electrical and thermal boundary conditions. The nodal potential vector V of N_V length and the nodal temperature vector T of N_T length are the primary unknowns to be determined. Fig. 2 compares the existing loosely-coupled and the proposed tightly-coupled schemes in the time evolution.

With a given spatial discretization, (7) can be represented by a general nonlinear differential algebraic equation (DAE)

$$\mathbf{C}\frac{dx}{dt} = -F(x) - b, \tag{8}$$

where

$$\mathbf{C} = \begin{bmatrix}\mathbf{C}_E & \\ & \mathbf{C}_T\end{bmatrix}, F = \begin{bmatrix}\nabla\cdot(\sigma(T)\nabla V) + J_d(V,T)\\ \nabla\cdot(\kappa(T)\nabla T) + Q(V,T)\end{bmatrix}, x = \begin{bmatrix}V\\T\end{bmatrix}. \tag{9}$$

The standard solution approach for the nonlinear DAE is the BDF-Newton-Krylov combo (BDF-NK), which starts with reducing (8) to a nonlinear algebraic equation via a polynomial approximation of the time derivative term. For instance, the first-order BDF (backward Euler) results in

$$\mathbf{C}x_{n+1} + hF(x_{n+1}) - \mathbf{C}x_n + hb_{n+1} = 0, \tag{10}$$

where h is the time step size between the nth and the $(n+1)$th steps. Then the Newton's method is applied to solve (10) iteratively

$$\left(\mathbf{C} + h\mathbf{J}(x_{n+1}^k)\right)\Delta x_{n+1}^{k+1} = -\left(hF(x_{n+1}^k) + \mathbf{C}\left(x_{n+1}^k - x_n\right) + hb_{n+1}\right), \tag{11}$$

where Δx_{n+1}^{k+1} denotes the update of x at the $(k+1)$th Newton iteration in the $(n+1)$th step. The Jacobian of $F(x)$ is given by

$$\mathbf{J}(x) = \begin{bmatrix}\nabla\cdot(\sigma(T)\nabla) + \frac{\partial J_d}{\partial V} & \nabla\cdot\left(\frac{\partial\sigma(T)}{\partial T}\nabla V\right) + \frac{\partial J_d}{\partial T}\\ \frac{\partial Q}{\partial V} & \nabla\cdot\left(\kappa(T)\nabla + \frac{\partial\kappa(T)}{\partial T}\nabla T\right) + \frac{\partial Q}{\partial T}\end{bmatrix} \tag{12}$$

When it comes to whether a loosely-coupled or a tightly-coupled formulation should be selected for a particular ET problem, it depends critically on the ET coupling strength, which can be estimated by looking at the Jacobian (12). The upper right block represents the temperature sensitivity of the metal conductivities and the power device currents, and the lower left block indicates how sensitive the total heat generation is to the voltage variations. It is shown in [12] that the product of the norm of the two off-diagonal blocks in (12) largely determines the convergence rate of the nonlinear Gauss-Seidel iteration underlying the loosely-coupled approach. Specifically, one may partition the matrix $\mathbf{A} = \mathbf{C}^{-1}\mathbf{J}$ into a block form $[\mathbf{A}_{11}\mathbf{A}_{12}; \mathbf{A}_{21}, \mathbf{A}_{22}]^T$ and calculate the norm of the square matrix [12]

$$\|\mathbf{S}\| = \|\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\| \tag{13}$$

The larger is $\|\mathbf{S}\|$, the stronger is the coupling and the slower convergence one may expect from the loosely-coupled scheme.

The last step in BDF-NK is solving the linear system in (11) by a Krylov subspace iterative method such as the GMRES (generalized minimal residual method), which is more scalable compared to the direct method for large-scale problems. Each GMRES iteration only requires a matrix-vector product

$$\left(\mathbf{C} + h\mathbf{J}(x_{n+1}^k)\right)v \tag{14}$$

The BDF-NK approach, when applied in tightly-coupled ET simulation, has two major weaknesses. Firstly, the order of accuracy of BDF cannot go beyond two without compromising the numerical stability. To meet the accuracy requirement, the minimal step size is typically limited to be comparable with the fastest electrical transients in the metal interconnects. Due to that fact the electrical and thermal systems evolve at the same pace in a tightly-coupled scheme, a time stepping dictated mostly by the electrical characteristics is less efficient for modeling the much slower thermal dynamics. Second, the linear system in (11) is often ill-conditioned due to the stiffness of the ET coupled system, and the Krylov subspace iterative solvers are known to converge slowly with such systems.

3. EXPONENTIAL-INTEGRATOR-NEWTON-KRYLOV METHOD FOR TIME-DOMAIN SOLUTION

3.1 Formulation of EI-NK

An exponential integrator (EI) approach was devised in [8] to tackle the stiffness in the ET coupled system[1]. The DAE (8) is firstly transformed to an ordinary differential equation (ODE) as-

[1]In [8] the EI was applied to the electrical subsystem only in the loosely-coupled scheme. However, extension to the entire tightly-coupled system is not difficult.

suming $\mathbf{C}$ is invertible [11][2]

$$\frac{dx}{dt} = -\mathbf{C}^{-1}F(x) - \mathbf{C}^{-1}b. \tag{15}$$

Then a linearisation is applied at the beginning of the simulation with respect to the initial guess x_0

$$\begin{aligned}\frac{dx}{dt} &= -\mathbf{C}^{-1}\mathbf{J}_0 x - \mathbf{C}^{-1}\left(F(x) - \mathbf{J}_0 x\right) - \mathbf{C}^{-1}b \\ &= \mathbf{A}x - D(x) - \tilde{b},\end{aligned} \tag{16}$$

where $\mathbf{J}_0$ is the Jacobian (12) evaluated at x_0 and $D(x)$ is a nonlinear difference function. The nonlinear ODE (16) has an analytical solution given by

$$x_{n+1} = e^{\mathbf{A}h}x_n + \int_0^h e^{\mathbf{A}(h-\tau)}\left[-D(t_n+\tau) - \tilde{b}(t_n+\tau)\right]d\tau. \tag{17}$$

One way to differentiate different EI variants is to look at the approximation of the integral of the nonlinear function in (17). [8] applies the second-order Runge-Kutta (RK) approximation [1] to $D(x)$ resulting in the EI-RK2 method

$$\begin{aligned}a_n &= x_n + h\varphi_1(\mathbf{A}h)\left[\mathbf{A}x_n - D(x_n) - \tilde{b}_{n+1}\right] \\ x_{n+1} &= a_n - h^2\varphi_2(\mathbf{A}h)\left[D(a_n) - D(x_n)\right]\end{aligned} \tag{18}$$

where the φ-functions up to order-2 are defined as

$$\varphi_0(x) = e^x, \quad \varphi_1(x) = \frac{e^x - 1}{x}, \quad \varphi_2(x) = \frac{e^x - x - 1}{x^2}$$

Since matrix A is stiff with large gap between the largest and smallest eigenvalues, the matrix φ-functions applied to vectors are computed using the shift-and-invert (SAI) Krylov subspace method, an "inverse" version of the standard Krylov subspace methods, to provide a better approximation to slow manifold of the waveform and to enable large h to be used [8, 13, 14]. To keep the paper focused the details of SAI-Krylov method is given in Appendix A.

The EI approaches are attractive in transient ET analysis as they solve the linearized subpart of (16), i.e., $\frac{dx}{dt} = \mathbf{A}x$ exactly and thus avoid the error from the finite-difference approximation of time derivative underlying BDF methods in the first place [3]. The main source of error in EI comes only from the numerical approximation of $\varphi_k(A)v$, and thus the order of accuracy can be made fairly high without losing numerical stability by increasing the subspace dimension m [11].

The drawback of EI-RK2 lies in the relatively poor nonlinear modeling capability due to the explicit (RK), non-iterative approximation of $D(x)$ in (17). Such approximation is acceptable in [8] while weak to medium ET nonlinearity is targeted (it is meanwhile the reason for selecting a loosely-coupled framework in their work). The nonlinearity considered in our work, however, has to be strong enough to favor the tight coupling scheme. The explicit and one-off nonlinear solver thus becomes insufficient and would limit the use of large steps. Implicit and iterative nonlinear modeling is needed which motivates the combination of EI and NK as detailed in the following.

Starting from (17), we approximate $D(x)$ by a second-order implicit formulation

$$D(x) = \frac{1}{2}\left[D(x_{n+1}) + D(x_n)\right] \tag{19}$$

[2]the non-singularity of $\mathbf{C}$ can be warranted for the ET problem concerned by differentiating the Gauss law in non-conducting domain

Substituting (19) into (17) and exploiting the φ-function expression, one obtains

$$\begin{aligned}&x_{n+1} \\ &= e^{\mathbf{A}h}x_n - \int_0^h e^{\mathbf{A}(h-\tau)}\left(\frac{D(x_{n+1}) + D(x_n)}{2} + \tilde{b}(t+\tau)\right)d\tau \\ &= -\frac{h}{2}\varphi_1(\mathbf{A}h)D(x_{n+1}) + e^{\mathbf{A}h}x_n - \frac{h}{2}\varphi_1(\mathbf{A}h)\left(D(x_n) + 2\tilde{b}_{n+1}\right) \\ &= x_n - \frac{h}{2}\varphi_1(\mathbf{A}h)\left[D(x_{n+1}) - 2\mathbf{A}x_n + D(x_n) + 2\tilde{b}_{n+1}\right]\end{aligned} \tag{20}$$

Note that now the right-hand-side (RHS) of (20) also contains x_{n+1}. Solving (20) with the Newton's method requires solving the a composite linear system in each iteration

$$\begin{aligned}&\left(I + \frac{h}{2}\varphi_1(Ah)\left(A\left(x_{n+1}^{(k)}\right) - A\right)\right)\Delta x_{n+1}^{(k+1)} = \\ &-\left\{x_{n+1}^{(k)} + \frac{h}{2}\varphi_1(Ah)\left[D\left(x_{n+1}^{(k)}\right) + D(x_n) - 2Ax_n + 2\tilde{b}_{n+1}\right] - x_n\right\}\end{aligned} \tag{21}$$

Note that in (21), $\mathbf{A}\left(x_{n+1}^{(k)}\right) = -\mathbf{C}^{-1}J\left(x_{n+1}^{(k)}\right)$ uses the Jacobian evaluated at $x_{n+1}^{(k)}$. The RHS involves one evaluation of φ-function-vector product. As in the BDF case, the linear system in (21) is solved iteratively by GMRES. In each iteration the matrix-vector product

$$\left(I + \frac{h}{2}\varphi_1(Ah)\left(A\left(x_{n+1}^{(k)}\right) - A\right)\right)v \tag{22}$$

is evaluated using the SAI-Krylov subspace solver. The algorithmic flow of EI-NK is given in Algorithm 1.

Algorithm 1: EI-NK

begin
- Given initial values $[V_0, T_0]^T$
- Compute J_0 via (12) and pre-factorize $(C + \gamma J_0)$
- **for** $n = 1, 2, ...$ **to** n_{ET} **do**
  - Obtain $\tilde{b}_{n+1}$
  - **for** $k \leftarrow 1$ **to** n_{NT} **do**
    - Compute RHS of (21) with one SAI-Krylov solution
    - Solve (21) for $\Delta x_{n+1}^{(k+1)}$ by GMRES, with each matrix-vector product (22) evaluated by one SAI-Krylov solution
    - If $\Delta x_{n+1}^{(k+1)} < tol_{NT}$, $x_{n+1} = x_{n+1}^{(k+1)}$

One difference between BDF-NK and EI-NK lies in the linear systems they solve by iterative solvers. The linear system (22) of EI-NK is intrinsically better conditioned than the one (14) of BDF-NK. This is because the φ-function performs in effect a contractive spectral mapping of Ah, i.e., if the eigenvalues of Ah, $\lambda(Ah) \in (-\infty, 0]$, then $\lambda(\varphi_1(Ah)) \in [0, 1]$. As a consequence, no preconditioner is needed for (22) in the GMRES solution. On the other hand, preconditioners must be applied to (14) to ensure practical convergence. For a fair comparison, in this work we apply $(C + hJ_0)^{-1}$ as the preconditioner for (14). As will be shown in the numerical experiment section, the GMRES convergence of the unpreconditioned system in EI-NK is still better than the preconditioned system in BDF-NK.

3.2 Complexity Analysis

In this subsection we will briefly analyze the computational complexity of transient ET analysis using the loosely-coupled scheme with BDF-NK (applied individually to the electrical and the thermal subsystems), the tightly-coupled scheme with BDF-NK and the tightly-coupled scheme with EI-NK. Without causing confusion we drop the "NK" in this subsection and denote the three approaches by LC-BDF, TC-BDF and TC-EI. TC-EI-RK2 is not considered since we are aiming at relatively strong ET interactions that make the RK-type of nonlinear approximation inefficient. Fixed time steps are applied in all time integration. Since the major cost of these methods lies in the back-forward-substitutions (one back-substitution plus one forward-substitution), as used in preconditioned GMRES iterations for BDF methods and the SAI Krylov subspace approximation of the φ-function for the EI method, we count the numbers and the sizes of back-forward-substitutions to measure the computational expense, as listed below (all numbers are averaged over the entire simulation)

- LC-BDF: $n_{loop}\times n_E^{LC}\times n_{NT-E}^{LC}\times n_{GMRES-E}^{LC}$ back-forward solves of size $N_V \times N_V$, and $n_{loop} \times n_T^{LC} \times n_{NT-T}^{LC} \times n_{GMRES-T}^{LC}$ back-forward solves of size $N_T \times N_T$. n_{loop} is the loosely-coupled ET loops to converge, n_E^{LC}, n_{NT-E}^{LC} and $n_{GMRES-E}^{LC}$ are the number of time steps, the number of Newton iterations per time step, and the number of GMRES iterations per Newton iteration for the electrical system. Similar notations apply to the thermal system.
- TC-BDF: $n_{ET}^{TC-BDF}\times n_{NT}^{TC-BDF}\times n_{GMRES}^{TC-BDF}$ back-forward solves of size $(N_V+N_T)\times(N_V+N_T)$, where n_{ET}^{TC-BDF}, n_{NT}^{TC-BDF} and n_{GMRES}^{TC-BDF} are the number of time steps, the number of Newton iterations per time step, and the number of GMRES iterations per Newton iteration for TC-BDF.
- TC-EI: $n_{ET}^{TC-EI}\times n_{NT}^{TC-EI}\times\left(n_{GMRES}^{TC-EI}+1\right)\times n_{SAI}$ back-forward solves of size $(N_V+N_T)\times(N_V+N_T)$. The extra factor n_{SAI}^{TC-EI} refers to the dimension of the SAI Krylov subspace for approximating the φ-function-vector products (cf. Appendix A).

In general, the loosely-coupled schemes need more back-forward substitutions of smaller sizes, whereas the tightly-coupled schemes require fewer substitutions with larger matrices. For strongly coupled ET interactions, the large n_{loop} needed in the loosely-coupled scheme will render it less efficient compared to the tightly-coupled one. For the two tightly-coupled schemes, TC-EI requires more back-forward solves per GMRES iteration due to the need of approximating the φ-function in (22), but is able to use fewer time steps (lager step size) and fewer GMRES iterations per Newton iteration than TC-BDF. As will be shown in the next section, the saving of TC-EI in n_{ET} and n_{GMRES} will outweigh the extra cost due to n_{SAI}^{TC-EI}.

4. NUMERICAL RESULTS

All the prototypes are implemented and tested in Matlab. A practical LDMOS ((laterally diffused metal oxide semiconductor) device is used to verify the proposed simulator. The 3D view and the specifications of the test structure are shown in Fig. 3. A 100KHz trapezoidal pulse ($10\mu s$ period, $0.5\mu s$ rising/falling edge, 50% duty cycle) and 5V amplitude is applied as the gate driving signal. A spatial discretization by the 3D finite-volume method (FVM) results in a mesh with $N_V = 217,202$ potential unknowns and $N_T = 337,869$ temperature unknowns. We apply the second-order BDF (BDF2) in all BDF-based approaches and use GMRES as the default Krylov subspace iterative method. Some convergence criteria need to be defined for numerical comparisons. The loosely-coupled ET iteration is considered convergent if

$$\max\left(\frac{\|\Delta V\|}{\|V\|},\frac{\|\Delta T\|}{\|T\|}\right) < tol_{ET}$$

The same criterion is applied to determine the convergence of the Newton's method in the tightly-coupled approaches. $tol_{ET} = 10^{-3}$ is used throughout the testing. The relative GMRES tolerance is $tol_{GMRES} = 10^{-6}$, and the accuracy of the SAI-Krylov approximation of the φ-functions (25) is controlled by $tol_{SAI} = 10^{-8}$.

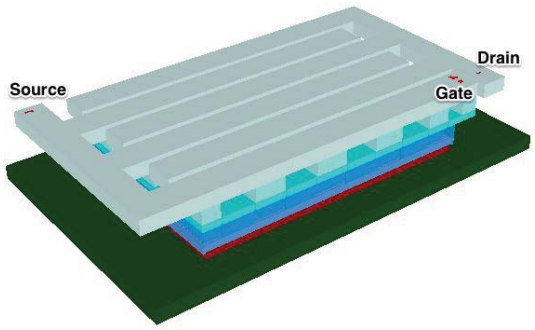

Figure 3: 3D view of the power LDMOS structure (the die part is truncated for better visualization). The back-end structure consists of 6 metal layers. The die has an area of $750\times 450\mu m^2$ and a thickness of $400\mu m$. The LDMOS has 94 transistor fingers with an $18.45mm$ total gate length.

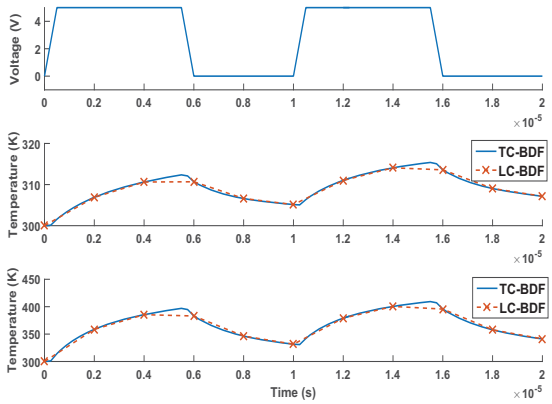

Figure 4: Top: gate driving voltage. Middle: Device 1 drain temperature with low ET coupling ($\|\mathbf{S}\| = 0.8$). Bottom: Device 1 drain temperature with high ET coupling ($\|\mathbf{S}\| = 9.1$).

We start with showing the benefits of using a tightly-coupled scheme when the ET coupling is strong. We produce a weak ET coupling case by modifying the device current table model to reduce the $\frac{\partial J_d}{\partial T}$ and $\frac{\partial Q}{\partial V}$ terms in (12), and a strong ET coupling case by increasing the two terms. The corresponding $\|\mathbf{S}\|$ values evaluated by (13) at the second time step are 0.8 and 9.1 respectively. To simplify the comparisons, only LC-BDF and TC-BDF are applied to simulate the two scenarios. The LC-BDF scheme uses 5 thermal steps and each contains 20 electrical steps, rendering 100 electrical steps in total. TC-BDF needs the same number of ET steps to resolve the electrical transients as it is based on the same BDF2 time

Table 1: Performance comparison of loosely-coupled and tightly-coupled schemes for weak and strong ET couplings. The ET loop numbers are averaged over thermal steps in LC-BDF. The amounts of Newton iterations and back-forward (B-F) solves are total numbers of the whole simulation.

Method	# of steps $n_T \times n_E$ or n_{ET}	Low ET coupling ($\|\|\mathbf{S}\|\| = 0.8$))				High ET coupling ($\|\|\mathbf{S}\|\| = 9.1$			
		ET loop (avg)	Newton	B-F solves	Runtime (s)	ET loop (avg)	Newton	B-F solves	Runtime (s)
LC-BDF	5×20	3	665+39	2793+167	358	9	3840+191	26897+1267	3669
TC-BDF	100	1	335	1672	759	1	571	4768	2155

integration. The transient temperatures at the drain terminal of the 1st device are shown in Fig. 4. As expected, the temperature starts to increase when the MOSFETs are turned on and decrease after the devices are turned off. For strong ET coupling the temperature increase is much more dramatic. Both LC-BDF and TC-BDF produce accurate solutions. The loosely-coupled scheme computes temperature at the end of each thermal step, thus the sampling interval is larger. The tightly-coupled scheme, on the other hand, updates the temperature at each ET step, and thus the sampling is denser. As can be seen in the second period of the high ET coupling case, the temperature difference between the peak and the end point can reach $66K$, which would otherwise be omitted if one only simulates the temperature at the end point of each period and may cause damage to the devices. This indicates the necessity of a sufficient resolution to the transient temperature variation for high-power applications.

The computational performance of LC-BDF and TC-BDF is compared in Table 1 for one period. For the weak ET coupling, the LC method converges rapidly in three loops and thus is more efficient than the TC scheme owing to the smaller matrix size in back-forward (B-F) solves. When the ET coupling becomes strong, the convergence of LC-BDF slows down significantly and requires 9 loops to converge within 10^{-3}. The number of Newton iterations increases by $6X$, and that of back-forward substitutions as well as the runtime by $10X$. In contrast, the performance degradation of tightly-coupled scheme is less significant, in which the increases of Newton iteration, back-forward solves and runtime are only $0.7X$, $2.7X$ and $2.7X$, respectively. The back-forward substitutions for the electrical, thermal and combined systems are respectively $0.12s$,$0.19s$ and $0.45s$, thus the lower efficiency of LC-BDF for the strongly coupled case is due to that the increase in the number of back-forward solves in LC-BDF outweighs the saving from solving smaller systems. This also demonstrates the advantages of tightly-coupled approaches when being used in strongly coupled ET problems.

Next, we compare the performance of three different tightly-coupled methods discussed in this paper, namely, BDF-NK, EI-RK2 and EI-NK (the TC prefix is omitted). Fig. 5 shows the temperature waveforms of the strong ET coupling case computed by the three methods using different step sizes. The reference solution is obtained by BDF-NK with small steps ($h = 0.01\mu s$). For $h = 0.1\mu s$ the accuracy of BDF-NK remains acceptable, but for $h = 0.25\mu s$ the error becomes noticeable. On the other hand, the EI-NK can provide accurate results up to $h = 1\mu s$, $10X$ larger than the step size allowed in BDF-NK owing to the higher order of accuracy in time integration. Finally, the accuracy of EI-RK2 is unsatisfactory due to the explicit approximation of nonlinearity. Since the tightly-coupled formulation is more commonly employed to handle strong ET interactions, the Newton's method is generally the preferred nonlinear solver to accelerate convergence. The performance data of the tightly-coupled BDF-NK and EI-NK are tabulated in Table 2. Here the step sizes of BDF-NK and EI-NK are selected to be $0.1\mu s$ and $1\mu s$ that produce comparable accuracy, and hence EI-NK can use $10X$ fewer time steps to finish the simulation. Another difference lies in the convergence rate of the GMRES solver within the Newton's method. BDF-NK, after the preconditioner is applied, still needs on average $4568/571 = 8.4$ GMRES iterations per Newton step. In contrast, EI-NK requires only $192/45 = 4.3$ GMRES iterations. Although EI-NK needs extra multiple back-forward solves in computing each φ-function-vector product, the saving from fewer time steps and faster GMRES convergence outweigh the extra cost, and render EI-NK 45% faster than BDF-NK.

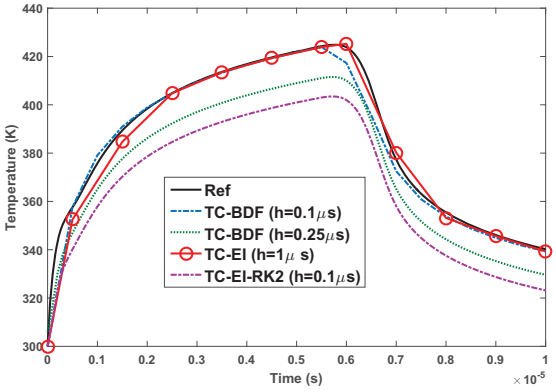

Figure 5: Temperature (strong ET coupling, 1st device drain terminal) computed by BDF-NK, EI-NK and EI-RK2 with different step sizes h.

5. CONCLUSION

We have proposed a new transient ET simulation framework for accurate chip-level ET analysis for BCD integration technology. The framework features a tightly-coupled formulation to provide better handling for the strong ET interactions as a consequence of the ever-increasing functionality integration density. To address the computational challenges arising from the tightly-coupled formulation, a specialized nonlinear EI-NK approach, is developed to enable large time steps and faster convergence in Krylov subspace iterative solution. The numerical results have demonstrated the advantages of the proposed framework.

6. REFERENCES

[1] M. Caliari and A. Ostermann. Implementation of exponential rosenbrock-type integrators. *Applied Numerical Mathematics*, 59(3-4):568–581, 2009.

[2] R. Chandra. Transient temperature analysis. In *EDSSERC*, Sep 2006.

[3] Q. Chen, W. Zhao, and N. Wong. Efficient matrix exponential method based on extended Krylov subspace for transient simulation of large-scale linear circuits. In *Design*

Table 2: Performance comparison of BDF-NK and EI-NK for the high EI coupling case. The GMRES count for EI-NK includes the pure GMRES iterations plus the extra one φ-function evaluation for generating the RHS of each Newton iteration. The SAI-Krylov dimension is unique for EI-NK and averaged over all GMRES iterations

Method	# of steps	Newton	GMRES	SAI-Krylov (avg)	B-F solves	Runtime (s)
TC-BDF-NK	100	571	4568	NA	4768	2155
TC-EI-NK	11	45	147+45	18	3276	1481

Automation Conference (ASP-DAC), 2014 19th Asia and South Pacific, pages 262–266, Jan 2014.

[4] R. Gillon, P. Joris, H. Oprins, B. Vandevelde, A. Srinivasan, and R. Chandra. Practical chip-centric electro-thermal simulations. In *14th International Workshop on Thermal Inveatigation of ICs and Systems (THERMINIC)*, pages 220–223, Sep 2008.

[5] D. E. Keyes et al. Multiphysics simulations: Challenges and opportunities. *International Journal of High Performance Computing Applications*, 27(1):4–83, 2013.

[6] D. Knoll and D. Keyes. Jacobian-free newton-krylov methods: a survey of approaches and applications. *Journal of Computational Physics*, 193(2):357 – 397, 2004.

[7] B. Krabbenborg, A. Bosma, H. de Graaff, and A. Mouthaan. Layout to circuit extraction for three-dimensional thermal-electrical circuit simulation of device structures. 15(7):765–774, Jul 1996.

[8] Q. Mei, W. Schoenmaker, S. H. Weng, H. Zhuang, C. K. Cheng, and Q. Chen. An efficient transient electro-thermal simulation framework for power integrated circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, PP(99):1–1, 2015.

[9] M. Pfost, D. Costachescu, A. Mayerhofer, M. Stecher, S. Bychikhin, D. Pogany, and E. Gornik. Accurate temperature measurements of DMOS power transistors up to thermal runaway by small embedded sensors. *IEEE Transactions on Semiconductor Manufacturing*, 25(3):294–302, Aug 2012.

[10] W. Schoenmaker, O. Dupuis, B. De Smedt, P. Meuns, J. Ocenasek, W. Verhaegen, D. Dumlugol, and M. Pfost. Fully-coupled 3D electro-thermal field simulator for chip-level analysis of power devices. In *19th International Workshop on Thermal Investigations of ICs and Systems (THERMINIC)*, pages 210–215, Sep 2013.

[11] S.-H. Weng, Q. Chen, and C.-K. Cheng. Time-domain analysis of large-scale circuits by matrix exponential method with adaptive control. 31(8):1180–1193, 2012.

[12] J. P. Whiteley, K. Gillow, S. J. Tavener, and A. C. Walter. Error bounds on block gausścseidel solutions of coupled multiphysics problems. *International Journal for Numerical Methods in Engineering*, 88(12):1219–1237, 2011.

[13] H. Zhuang, S.-H. Weng, and C.-K. Cheng. Power grid simulation using matrix exponential method with rational Krylov subspaces. In *IEEE 9th International Conference on ASIC (ASICON)*, pages 369–372, Oct 2013.

[14] H. Zhuang, S.-H. Weng, J.-H. Lin, and C.-K. Cheng. MATEX: A distributed framework for transient simulation of power distribution networks. In *IEEE/ACM Design Automation Conference (DAC)*, pages 81:1–81:6, 2014.

APPENDIX

A. COMPUTATION OF φ-FUNCTION-VECTOR PRODUCTS WITH THE SAI-KRYLOV SUBSPACE METHOD

The main step of SAI-Krylov for computing the φ-function-vector products is an m-step Arnoldi process applied to $(\mathbf{I}-\gamma\mathbf{A})^{-1}$

$$(\mathbf{I}-\gamma\mathbf{A})^{-1}\mathbf{V}_m = \mathbf{V}_m\mathbf{H}_m + \mathbf{H}(m+1,m)v_{m+1}e_m^T, \quad (23)$$

in which $\mathbf{V}_m$ is an orthonormal basis of the m-dimensional Krylov subspace $\mathcal{K}_m\left((\mathbf{I}-\gamma\mathbf{A})^{-1}, v\right)$ with γ being a shift parameter. $\mathbf{H}$ is the upper Hessenberg coefficient matrix, $\mathbf{H}_m$ the leading $m\times m$ submatrix and e_m the m-th column of an identity matrix. Note that in implementation $(\mathbf{I}-\gamma\mathbf{A})^{-1}v$ is computed as $(\mathbf{C}+\gamma\mathbf{J})^{-1}\mathbf{C}v$, for which the LU factors of $(\mathbf{C}+\gamma\mathbf{J})$ are firstly computed, followed by two triangular solves with $\mathbf{C}v$ to obtain the result. Then the φ-function-vector product is approximated as its orthogonal projection onto the SAI-Krylov subspace

$$\varphi_k\left(\mathbf{A}h\right)v \approx \mathbf{V}_m\mathbf{V}_m^T e^{\mathbf{A}h}v = \beta\mathbf{V}_m\varphi_k\left(\widetilde{\mathbf{H}}_m h/\gamma\right)e_1, \quad (24)$$

with

$$\widetilde{\mathbf{H}}_m = \left(\mathbf{I}-\mathbf{H}_m^{-1}\right), \ \ \beta = \|v\|_2$$

A posteriori residual estimate is given by

$$res_{Krylov} = \frac{\beta}{\gamma}\mathbf{H}(m+1,m)\|(\mathbf{I}-\gamma\mathbf{A})v_{m+1}e_m^T\mathbf{H}_m^{-1}\varphi_k\left(\widetilde{\mathbf{H}}_m h/\gamma\right)e_1\|. \quad (25)$$