Generating Current Constraints to Guarantee RLC Power Grid Safety

ZAHI MOUDALLAL and FARID N. NAJM, University of Toronto

A critical task during early chip design is the efficient verification of the chip power distribution network. Vectorless verification, developed since the mid-2000s as an alternative to traditional simulation-based methods, requires the user to specify current constraints (budgets) for the underlying circuitry and checks if the corresponding voltage variations on all grid nodes are within a user-specified margin. This framework is extremely powerful, as it allows for efficient and early verification, but specifying/obtaining current constraints remains a burdensome task for users and a hurdle to adoption of this framework by the industry. Recently, the *inverse* problem has been introduced: Generate circuit current constraints that, if satisfied by the underlying logic circuitry, would guarantee grid safety from excessive voltage variations. This approach has many potential applications, including various grid quality metrics, as well as voltage drop-aware placement and floorplanning. So far, this framework has been developed assuming only resistive and capacitive (RC) elements in the power grid model. Inductive effects are becoming a significant component of the power supply noise and can no longer be ignored. In this article, we extend the constraints generation approach to allow for inductance. We give a rigorous problem definition and develop some key theoretical results related to maximality of the current space defined by the constraints. Based on this, we then develop three constraints generation algorithms that target the peak total chip power that is allowed by the grid, the uniformity of current distribution across the die area, and a combination of both metrics.

CCS Concepts: • **General and reference** → **Verification**; • **Hardware** → **Metallic interconnect**; **Package-level interconnect**; **Interconnect power issues**; **Electronic design automation**; **Power and thermal analysis**; **Signal integrity and noise analysis**; *Partitioning and floorplanning*; *Placement*; *Power grid design*; • **Mathematics of computing** → Discretization; Linear programming;

Additional Key Words and Phrases: Current constraints generation, current budgets, power grid, voltage integrity

ACM Reference Format:
Zahi Moudallal and Farid N. Najm. 2017. Generating current constraints to guarantee RLC power grid safety. ACM Trans. Des. Autom. Electron. Syst. 22, 4, Article 66 (June 2017), 39 pages.
DOI: http://dx.doi.org/10.1145/3054746

1. INTRODUCTION

Successive technology generations of integrated circuits have continuously driven towards reduced feature size, larger operating frequency, and lower voltage supply. The large operating frequency of modern chips often leads to large switching currents that flow in the power and ground networks, causing power supply noise. Furthermore, the lower voltage supply implies that noise margins are reduced and susceptibility to supply noise is increased. As a result, the power and ground networks experience excessive voltage variations that put both circuit performance and reliability at risk. A well-

This work was supported by the Natural Sciences and Engineering Research Council of Canada.
Authors' addresses: Z. Moudallal and F. N. Najm, Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, Ontario, Canada M5S 3G4; emails: zahi.moudallal@mail.utoronto.ca, f.najm@utoronto.ca.

DOI: http://dx.doi.org/10.1145/3054746

designed chip power distribution network should deliver well-regulated voltages at all grid nodes to guarantee correct logic functionality at the intended design frequency. It is clear that analysis and verification of the power distribution network are necessary for reliable high-performance designs. We will use the term "power grid" to refer to either the power or ground distribution networks. Furthermore, a power grid consisting of only resistors and capacitors will be referred to as an RC model or RC grid, while a power grid consisting of resistors, capacitors, and inductors will be referred to as an RLC model or RLC grid.

Voltage variations on the power grid result from two major factors. The metal lines of the gird are resistive in nature that, due to the large number of metal branches required for grid routing, makes a significant power noise, commonly referred to as resistive *IR* drop. This drop is becoming increasingly significant from one technology node to the next as the metal lines widths are shrinking. The fast switching currents in the power grid generate inductive effects, also referred to as Ldi/dt noise, due to the significant inductance of the package leads resulting in power noise at the pad locations. This inductive noise is becoming a significant component of the total power supply noise [Lee et al. 2004; Muramatsu et al. 2005; Srivastava et al. 2005].

A common technique to mitigate the effects of the resistive and inductive parasitics on the power distribution network is to insert *decoupling* capacitances by filling on-die white spaces at strategic locations. On the other hand, these capacitances along with the resistive and inductive parasitics form a complex *RLC* circuit that has a specific resonance frequency. If the chip operating frequency is close to this resonance frequency, then the grid might experience large voltage variations that can be problematic. Therefore, inductive effects on the power grid must be included when verifying circuits operating at high frequencies.

Power grid verification techniques often used in the industry are based on simulation. Such methods assume full knowledge of the current waveforms drawn by every logic block tied to the grid. These waveforms would then be used to simulate the grid and determine the voltage variation at every node. Verifying the grid in this manner is computationally prohibitive, as it requires an exhaustive set of current traces to cover all possible voltage variations exhibited on the grid. Several non-exhaustive methods have been introduced in the recent past to implement some sort of search in current space. For example, there are search techniques that find vectors to maximize the current drawn from the power network [Krstic and Cheng 1997], as well as methods that use voltage drop analysis based on current statistics [Pant et al. 2004]. Another notable limitation of simulation-based methods is that they cannot be applied at an early stage of the design flow, when detailed information on the circuit currents is not yet available. This signifies the need for a verification approach that does not require simulation and only assumes information about the circuit currents that may be available early in the design process, that is, *vectorless* methods.

Vectorless power grid verification, first proposed in Kouroussis and Najm [2003], does not require full knowledge of the circuit currents. Contrary to simulation-based approaches, this method relies on information that may be available at an early stage of the design in the form of *current constraints*. Essentially, vectorless verification consists of finding the worst-case voltage fluctuations achievable at all nodes of the grid under all possible transient current waveforms that satisfy user-specified current constraints. The grid is said to be *safe* if these fluctuations are below certain user-specified thresholds. These methods are often formulated as linear programs (LPs).

Vectorless verification methods require the user to obtain/specify the current constraints, which can be done by an "offline" process of simulation of a logic block, if the block is available and small enough to simulate, or heuristically based on design expertise and engineering judgement (how big it is, what its power needs were in a previous technology, how scaling would affect those needs, etc.). The lack of a *systematic*

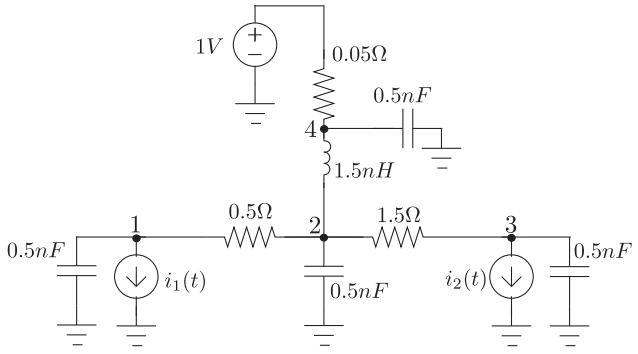
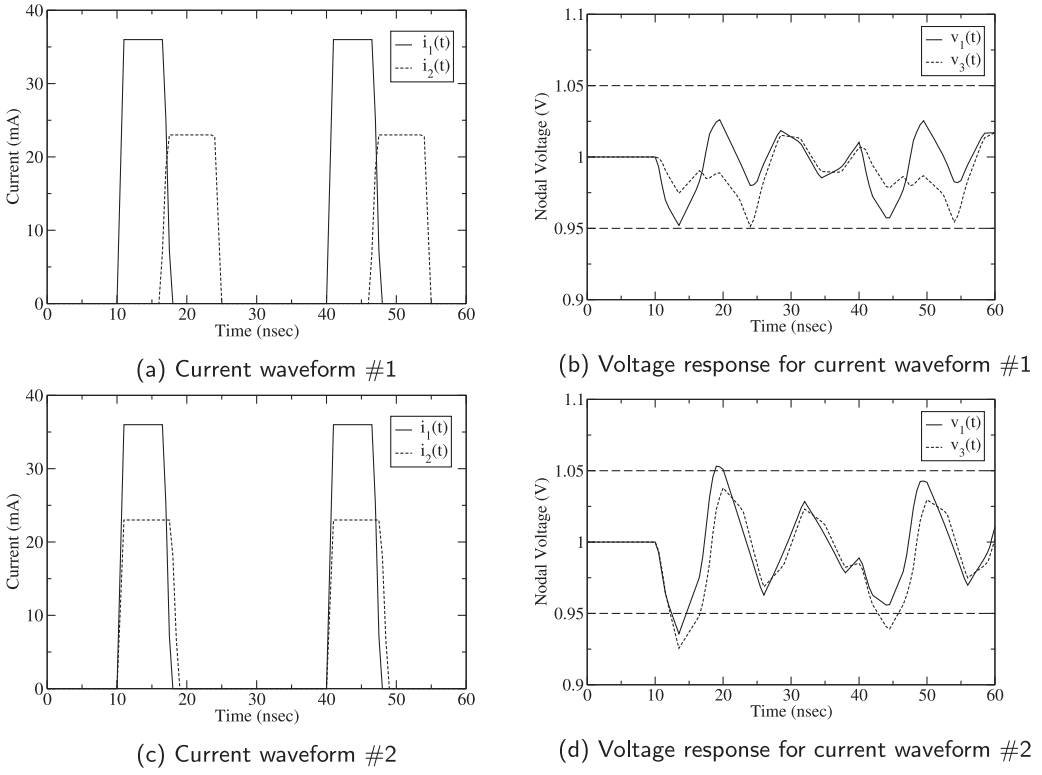


Fig. 1. Simple example of an RLC power grid.

approach to obtain/specify current constraints is a key limitation of vectorless verification methods that remains a burdensome task and a hurdle to adoption of these methods by industry practices. This particular problem of obtaining/specifying the current constraints was first addressed in Moudallal and Najm [2015] by proposing the following framework, referred to as the *inverse* vectorless verification: Given a grid and the allowed voltage drop thresholds at all grid nodes, we aim to *generate* circuit current constraints that, if satisfied by the underlying circuitry, would guarantee grid safety. These current constraints encapsulate much useful information about the grid, because these are essentially power budgets for the logic blocks under the grid. If all design teams respect these budgets throughout the design flow, then the grid is *safe by construction* at the end of the design. If the constraints impose too severe a budget on a certain block in some corner of the die, for example, then one can address the problem early on by modifying the grid, while it is still easy to do so, and generating a fresh set of constraints. Alternatively, if the budgets are too severe for a candidate layout location of a high level block, then perhaps the floorplan needs to be reconsidered. Indeed, the constraints can be used to drive automated floorplanning as well as placement, so grid-aware placement may become feasible, something that has never been done before.

The authors in Moudallal and Najm [2015] laid down the theoretical foundation for the current constraints generation framework. They show that there is an infinite number of sets of current constraints, some of that allow more "flexibility" than others for the underlying logic circuitry. Thus, constraints generation algorithms could target key grid quality metrics such as the peak power dissipation that the grid can safely support and the uniformity of current distribution across the die. These methods have been further improved in Moudallal and Najm [2016] in addition to introducing a combination of those quality metrics in Moudallal and Najm [2015].

In Moudallal and Najm [2015, 2016], an RC model of the power grid is used, ignoring the parasitic inductance. As the inductive noise is becoming a significant component of the power supply noise, the parasitic inductance can no longer be ignored. To demonstrate this, we present a simple example in Figures 1, 2, and 3. Figure 1 shows a simple power grid consisting of a four-node RLC circuit. The two current sources $i_1(t)$ and $i_2(t)$ represent circuit blocks whose switching activity constitutes the load on the grid, inducing voltage swings on the grid nodes. Figure 2 show the voltage variations experienced on nodes 1 and 3 due to two current configurations of the same waveforms $i_1(t)$ and $i_2(t)$ but with different temporal alignment. It is clear from Figure 2 that the second current configuration leads to large power supply noise and should be prohibited. Thus, current constraints provided to the design team should avoid such a configuration. Figure 3 shows a current container (represented as empty polygon)



(a) Current waveform #1

(b) Voltage response for current waveform #1

(c) Current waveform #2

(d) Voltage response for current waveform #2

Fig. 2. An example of a current waveform that leads to voltage violations.

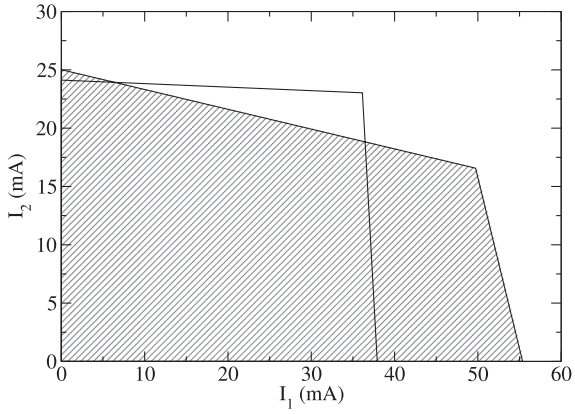


Fig. 3. A current container (represented as an empty polygon) generated for the RC circuit resulting from ignoring the inductance of Figure 1 that includes the problematic current waveform of Figure 2. Also, a current container (represented as striped polygon) generated using one of the proposed algorithms presented below that avoid the problematic current stimulus presented in Figure 2.

resulting from the algorithms in Moudallal and Najm [2015, 2016] on the RC circuit in Figure 1 after ignoring (short-circuiting) the inductance and another current container (represented as striped polygon) resulting from one of the algorithms presented in this article applied to the RLC circuit in Figure 1. Only the latter excludes the current trace of the problematic current configuration.

In this work, we extend the current constraints generation framework to allow for inductance. We use the same systematic way of defining the problem as in our previous work, in the sense that we are interested to discover a safe container that allows as much flexibility as possible to the underlying logic circuitry and targets specific design objectives. The major difference from the work in Moudallal and Najm [2016] is that, because inductive elements introduce voltage overshoots, the bound $\overline{v}(\cdot)$ that was appropriate for the RC case is no longer useful and a new bound $\overline{x}(\cdot)$ is required. Thus, most of the theoretical results in Moudallal and Najm [2016] require new proofs for the RLC context of this article. Furthermore, the constraints generation algorithms described in Section 5, even though they target the same design objectives as in Moudallal and Najm [2016], turn out to significantly differ because of the new theoretical bound $\overline{x}(\cdot)$.

The rest of the article is organized as follows. The next section describes the power grid model. In Section 3, we present some important notations and definitions and give the problem definition. In Section 4, we present the bulk of our theoretical contribution. In Section 5, we give three algorithms for generating circuit current constraints that are provably maximal. Section 6 describes the implementation details of one of the algorithms. In Section 7, we present some test results along with an analysis and comparison of the three algorithms and describe the tradeoffs among them. Finally, in Section 8, we give concluding remarks.

2. GRID MODEL

Throughout the rest of the article, we will use the notation $x \leq y$ (or $x < y$), for any two vectors x and y, to denote that $x_i \leq y_i$ (or $x_i < y_i$), $\forall i$, respectively. Similarly, we will use the notation $X \geq 0$ (or $X > 0$), for any matrix X, to denote that $X_{ij} \geq 0$ (or $X_{ij} > 0$), $\forall i, j$, respectively. We will also use the notation $\mathbb{R}^m_+$ to denote the non-negative part of the real space, that is, $\mathbb{R}^m_+ = \{x \in \mathbb{R}^m : x \geq 0\}$.

The power grid is a large full-chip structure of connected metal lines across multiple layers interconnected through vias and connected by C4 bumps to wiring in the package and on the board. Typically, a power grid is modelled as a linear circuit composed of a large number of lumped linear (RLC) elements. At its metal-1 or metal-2 terminals, the grid is loaded by the circuit blocks, where nonlinearities are encountered due to the circuit metal-oxide-semiconductor field-effect transistors (MOSFETs). It is practically impossible to jointly simulate or analyze both the full nonlinear circuit and the large grid all at once, and common practice is to decouple the two. This typically means that the circuit blocks are represented by some suitable model, consisting of a current source along with some parasitic network to ground. However, for grid verification, these parasitics are often neglected because of the larger impact that uncertainty of currents has on the grid response, and so the circuit current sources are often assumed ideal—and this is what will be assumed in this article.

Consider an RLC model of the power grid in which there are three types of nodes: (1) Some nodes are connected to ideal current sources to ground, in parallel with capacitors to ground, (2) some (most) nodes are connected to resistors or inductors to other grid nodes and capacitors to ground, and (3) some nodes are connected to resistors or inductors to other grid nodes and ideal voltage sources to ground. Note that, in this work, mutual inductances and branch capacitances are ignored. That is, only self-inductances are considered, and all capacitances are assumed to be connected to ground. The current sources (with their parallel capacitors) represent the currents

drawn by the logic circuits tied to the grid at these nodes. The ideal voltage sources represent the external voltage supply, V_{dd}.

Excluding the ground node, let the power grid consist of $n_v + p$ nodes, where nodes $\{1, \ldots, n_v\}$ are the nodes not connected to a voltage source, while the remaining nodes $(n_v + 1), (n_v + 2), \ldots, (n_v + p)$ are the nodes where the p voltage sources are connected. Let m be the number of current sources connected to the grid, whose positive (reference) direction of current is from node to ground, and assumed to be connected at nodes $1, 2, \ldots, m \leq n_v$, and let $i(t) \geq 0$ be the $m \times 1$ vector of all source currents. Also, let H be an $n_v \times m$ matrix of 0 and 1 entries that identifies (with a 1) which node is connected to which current source. Finally, let n_l be the number of inductors in the grid.

Let $\vartheta(t)$ be the $n_v \times 1$ vector of node voltages, relative to ground. By superposition, $\vartheta(t)$ may be found in three steps: (1) open-circuit all the current sources and find the response, which would be $\vartheta^{(1)}(t) = V_{dd}$; (2) short-circuit all the voltage sources and find the response $\vartheta^{(2)}(t)$; and (3) find $\vartheta(t) = \vartheta^{(1)}(t) + \vartheta^{(2)}(t)$. To find $\vartheta^{(2)}(t)$, Kirchhoff's current law (KCL) at every node $k \in \{1, \ldots, n_v\}$ provides the following:

$$G\vartheta^{(2)}(t) + C\dot{\vartheta}^{(2)}(t) + Mi_l(t) + Hi(t) = 0, \tag{1}$$

where $i_l(t)$ is the $n_l \times 1$ vector of inductor branch currents; G is the $n_v \times n_v$ conductance matrix, which is a sparse, symmetric positive semidefinite matrix with positive diagonal entries and non-positive off-diagonal entries; C is an $n_v \times n_v$ non-singular diagonal matrix of the node capacitances; and M is an $n_v \times n_l$ incidence matrix consisting of ± 1 or 0 elements only. If the graph consisting of all grid nodes $1, 2, \ldots, n_v$ and all grid resistances in between these nodes is a connected graph, then G is said to be *irreducible* (see the appendix). If this graph is not connected or does not cover all n_v nodes, then there is an easy and practical "fix" that maintains this useful property of G, which is to attach a large resistance in parallel with every inductor. These large resistors have a negligible effect on the circuit solution, but they have the effect that G becomes irreducible. We are mainly interested in the voltage drop $v(t) \triangleq V_{dd} - \vartheta(t) = -\vartheta^{(2)}(t)$, so

$$Gv(t) + Cv'(t) - Mi_l(t) = Hi(t). \tag{2}$$

To take into account the relationship between the inductor branch currents and the inductor voltages, we have the familiar inductor branch equation $M^T\vartheta^{(2)}(t) = Li_l'(t)$, from which

$$M^T v(t) + Li_l'(t) = 0, \tag{3}$$

where L is an $n_l \times n_l$ non-singular diagonal matrix consisting of the inductance values of the n_l inductors in the circuit. The dynamics of the power grid are governed by the combined set of Equations (2) and (3). Backward Euler discretization, applied to this system, leads to

$$Av(t) - Mi_l(t) = Bv(t - \Delta t) + Hi(t), \tag{4}$$

$$M^T v(t) + Ei_l(t) = Ei_l(t - \Delta t), \tag{5}$$

where $B = C/\Delta t$, $E = L/\Delta t$, and $A = G + B$. Multiplying Equation (5) by E^{-1} to get an expression for $i_l(t)$,

$$i_l(t) = -E^{-1}M^T v(t) + i_l(t - \Delta t), \tag{6}$$

and substituting that into Equation (4) gives

$$Dv(t) = Bv(t - \Delta t) + Mi_l(t - \Delta t) + Hi(t), \tag{7}$$

where

$$D = G + \frac{C}{\Delta t} + M\left(\frac{L}{\Delta t}\right)^{-1} M^T. \tag{8}$$

It can be shown that D is a symmetric positive-definite $\mathcal{M}$-matrix [Fawaz and Najm 2016], so D is non-singular and has non-positive off-diagonal entries [Varga 1962]. Furthermore, notice that the matrix D is real, because G, C, and L, are real matrices, and D is irreducible due to Lemma E.4 (in the appendix), so $D^{-1} > 0$ [Varga 1962]. Multiplying Equation (7) by D^{-1} gives

$$v(t) = D^{-1}Bv(t - \Delta t) + D^{-1}Mi_l(t - \Delta t) + D^{-1}Hi(t), \tag{9}$$

and then substituting this for $v(t)$ into Equation (6), we get

$$i_l(t) = -E^{-1}M^T D^{-1}Bv(t - \Delta t) + (I_{n_l} - E^{-1}M^T D^{-1}M)i_l(t - \Delta t) - E^{-1}M^T D^{-1}Hi(t), \tag{10}$$

where I_{n_l} is the $n_l \times n_l$ identity matrix. Combining Equations (9) and (10) gives the system

$$\begin{bmatrix} v(t) \\ i_l(t) \end{bmatrix} = \begin{bmatrix} D^{-1}B & D^{-1}M \\ -E^{-1}M^T D^{-1}B & (I_{n_l} - E^{-1}M^T D^{-1}M) \end{bmatrix} \begin{bmatrix} v(t - \Delta t) \\ i_l(t - \Delta t) \end{bmatrix} + \begin{bmatrix} D^{-1}H \\ -E^{-1}M^T D^{-1}H \end{bmatrix} i(t). \tag{11}$$

Let

$$x(t) = \begin{bmatrix} v(t) \\ i_l(t) \end{bmatrix}, \quad F = \begin{bmatrix} D^{-1}B & D^{-1}M \\ -E^{-1}M^T D^{-1}B & (I_{n_l} - E^{-1}M^T D^{-1}M) \end{bmatrix}, \quad \text{and} \quad R = \begin{bmatrix} D^{-1}H \\ -E^{-1}M^T D^{-1}H \end{bmatrix}, \tag{12}$$

so the system is governed by the recurrence equation

$$x(t) = Fx(t - \Delta t) + Ri(t), \tag{13}$$

where $x(t)$ denotes the *response* vector of the power grid. For reference throughout the article, we recall the dimensions of all vectors and matrices, where $n = n_v + n_l$: $v(\cdot)$ is an $n_v \times 1$ vector; $i_l(\cdot)$ is an $n_l \times 1$ vector; $i(\cdot)$ is an $m \times 1$ vector; $x(\cdot)$ is an $n \times 1$ vector; G, C, B, A, and D are $n_v \times n_v$ matrices; L and E are $n_l \times n_l$ matrices; M is an $n_v \times n_l$ matrix; H is an $n_v \times m$ matrix; F is an $n \times n$ matrix; and R is an $n \times m$ matrix.

Because the RLC grid is a stable linear system, and because the backward difference approximation used in Equations (4) and (5) is absolutely stable [Lambert 1991], it follows that for $i(t) = 0$, $\forall t$, and any bounded initial condition $x(0)$, $x(t)$ converges to 0 as $t \to \infty$, so $\lim_{p\to\infty} F^p = 0$, which is known to be true if and only if $\rho(F) < 1$, where $\rho(F)$ is the spectral radius of F [Saad 2003]. This allows us to use results from Abdul Ghani and Najm [2011] and Fawaz and Najm [2016] that are replicated below in Equations (16) and (31).

Finally, we assume that a certain number of grid nodes $d \leq n_v$ are required to satisfy some user-provided *voltage safety specifications*, captured in the $2d \times 1$ vector $x_{th} = \left[\begin{smallmatrix} x_{ub} \\ x_{lb} \end{smallmatrix}\right]$, where $x_{ub} \geq 0$ and $x_{lb} \leq 0$ are $d \times 1$ vectors. These would typically be nodes at the lower metal layers, where the chip circuitry is connected. Suppose that nodes with voltage safety specification are labeled $\iota_1, \ldots, \iota_d$, so for any $k \in \{1, \ldots, d\}$, $x_{ub,k}$ and $x_{lb,k}$ are the voltage safety bounds on node ι_k. Let Π be a $d \times n$ matrix consisting of 0 and 1 elements only, specifying (with a 1 entry) the nodes that are subject to a voltage threshold specification, that is, $\Pi_{k,\iota_j} = 0$, $\forall j \neq k$, and $\Pi_{k,\iota_k} = 1$. Note that $\Pi \geq 0$ and has exactly one 1 in every row and at most one 1 per column, otherwise 0s. With this, the voltage safety specifications translate to $x_{lb} \leq \Pi x(t) \leq x_{ub}$, $\forall t$.

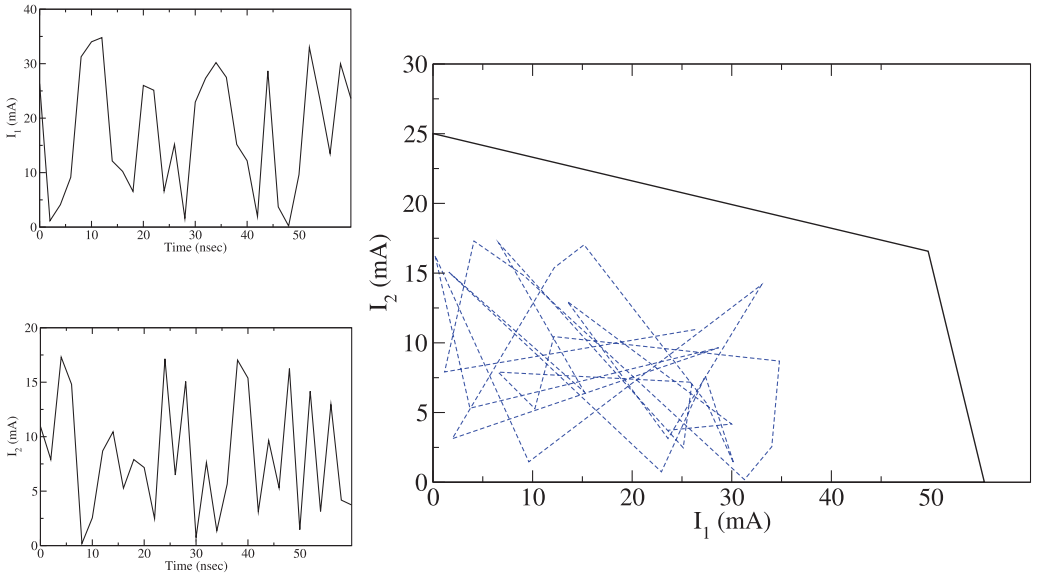


Fig. 4. Example of a container $\mathcal{F}$ for $i_1(t)$ and $i_2(t)$.

3. PROBLEM DEFINITION

Note that Definitions 3.1, 4.1, 4.3, 4.6, 4.7, and 4.9 were given in Moudallal and Najm [2016] and are repeated here for clarity. The statements of Lemmas 4.8, 5.1, D.3, D.4, and Theorem 4.13 have been presented in Moudallal and Najm [2016] for the RC case but require new proofs for the RLC context of this article.

First, we will introduce the notion of a *container* for a vector of current waveforms, which will help us express constraints that guarantee grid safety.

Definition 3.1 (*Container*). Let $t \in \mathbb{R}$, let $i(t) \in \mathbb{R}^m$ be a bounded function of time, and let $\mathcal{F} \subset \mathbb{R}^m$ be a closed subset of $\mathbb{R}^m$. If $i(t) \in \mathcal{F}$, $\forall t \in \mathbb{R}$, then we say that $\mathcal{F}$ contains $i(t)$, represented by the shorthand $i(\cdot) \subset \mathcal{F}$, and we refer to $\mathcal{F}$ as a container of $i(\cdot)$.

Figure 4 illustrates the idea of a container for a simple case of two current waveforms. Because $i(t) = [i_1(t)\ i_2(t)]^T \in \mathcal{F}$ for all time instants, we say that $\mathcal{F}$ contains $i(t)$.

Definition 3.2. If u is an $n \times 1$ vector and w is a $2n \times 1$ vector, then we say that $u \in w$ if, $\forall j \in \{1, \ldots, n\}$,

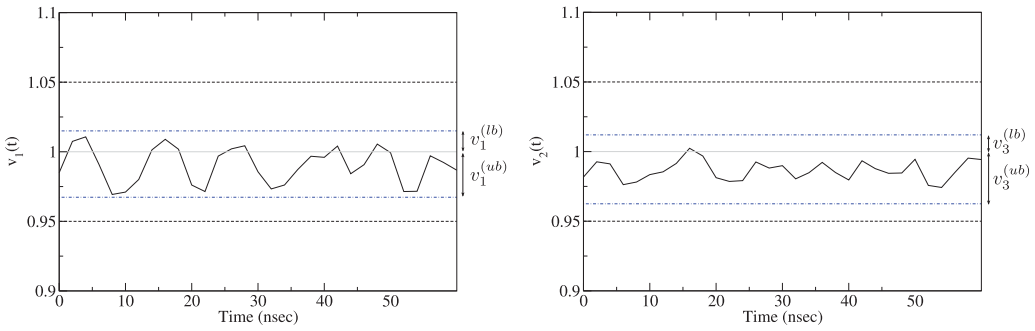
$$u_j \leq w_j \quad \text{and} \quad u_j \geq w_{j+n}. \tag{14}$$

Thus, u is upper bounded by the top half of w and lower bounded by the bottom half of w. We say that w is *empty* if there does not exist any x for which $x \in w$. Note that w is non-empty if and only if $w_j \geq w_{j+n}$, $\forall j \in \{1, 2, \ldots, n\}$. Notice that $0 \in x_{th}$, because $x_{lb} \leq 0 \leq x_{ub}$, so x_{th} is non-empty.

Lemma 3.3. *Let $u, u' \in \mathbb{R}^n$ and $w, w' \in \mathbb{R}^{2n}$. If $u \in w$ and $u' \in w'$, then $u + u' \in w + w'$.*

Proof. For any $j \in \{1, \ldots, n\}$, we have $u_j \leq w_j$, $u_j \geq w_{j+n}$, $u'_j \leq w'_j$, and $u'_j \geq w'_{j+n}$. It follows that $u_j + u'_j \leq w_j + w'_j$ and $u_j + u'_j \geq w_{j+n} + w'_{j+n}$, so $u + u' \in w + w'$. □

Definition 3.4 (*Safe Grid*). A grid is said to be safe for a given $i(t)$, defined $\forall t \in \mathbb{R}$, if $\Pi x(t) \in x_{th}$, $\forall t \in \mathbb{R}$.



(a) Node 1 (b) Node 3

Fig. 5. Nodal voltage at nodes 1 and 3 of Figure 1 due to current waveform in Figure 4. The gray line represents the V_{dd} value. The dashed lines represent the voltage overshoot/undershoot thresholds. The blue dotted lines represent the values of $P\overline{x}(\mathcal{F})$, where $\mathcal{F}$ is the current container represented in Figure 4.

Going back to the example of Figure 1, the nodes of interest are nodes 1 and 3 with voltage overshoot/undershoot thresholds of 50 mV, so $\Pi x(t) = [v_1(t)\, v_3(t)]^T$ and $x_{th} = [50\, 50\ -50\ -50]^T$. In Figure 5, we show the voltage response at nodes 1 and 3 due to the current waveform $i(t) = [i_1(t)\, i_2(t)]^T$ shown in Figure 4. Notice that the voltage response is within the specified thresholds so $\Pi x(t) \in x_{th}$ and the grid is safe under $i(t)$.

To check if a power grid is safe, one would typically be interested in the worst-case voltage variation at some grid node $k \in \{1, \ldots, n_v\}$, at some time point $\tau \in \mathbb{R}$, over a wide range of possible current waveforms. Using the above notation, and given a container $\mathcal{F}$ that contains a wide range of current waveforms that are of interest, we can express this as $\max_{i(t)\subset\mathcal{F}}(x_k(\tau))$ and $\min_{i(t)\subset\mathcal{F}}(x_k(\tau))$. Clearly, because $\mathcal{F}$ is the same irrespective of time, and applies at all time points $t \in \mathbb{R}$, then these worst-case voltage variations must be time invariant, independent of the chosen time point τ. We now introduce the eopt$(\cdot)$ notation, which is used to capture in a single vector all the separate worst-case voltage variations, as follows.

Definition 3.5 (eopt *Operator*). Let $\mathcal{Y}$ be a bounded and closed subset of $\mathbb{R}^m$ and let $f(\cdot)$ be a vector-valued function $f(\cdot) : \mathcal{Y} \to \mathbb{R}^n$. If $z \in \mathbb{R}^{2n}$ is a $2n \times 1$ vector such that, for every $i \in \{1, \ldots, n\}$

$$\begin{aligned} z_i &= \max_{y\in\mathcal{Y}}[f_i(y)] \\ z_{n+i} &= \min_{y\in\mathcal{Y}}[f_i(y)], \end{aligned}$$

then we capture this with the shorthand notation

$$z = \underset{y\in\mathcal{Y}}{\text{eopt}}[f(y)], \tag{15}$$

with the convention that $\text{eopt}_{y\in\mathcal{Y}}[f(y)] = 0$, if $\mathcal{Y} = \phi$.

With this, we can now define $x^{(opt)}(\mathcal{F}) \triangleq \text{eopt}_{i\subset\mathcal{F}}[x(t)]$ to be the worst-case response vector of the power grid. It should be clear that if $\mathcal{F}$ is not an empty set, then $x^{(opt)}(\mathcal{F})$ is not empty (as a $2n \times 1$ vector) and $x(t) \in x^{(opt)}(\mathcal{F})$ for all $i(t) \subset \mathcal{F}$. The exact expression of the worst-case response vector $x^{(opt)}(\mathcal{F})$ was derived in Abdul Ghani and Najm [2011] to be

$$x^{(opt)}(\mathcal{F}) = \sum_{q=0}^{\infty} \operatorname*{eopt}_{I \in \mathcal{F}}(F^q RI), \tag{16}$$

where I is an $m \times 1$ vector dummy variable in units of current. One way to check grid safety is to compute Equation (16) and then check if $[\Pi\, 0]x^{(opt)}(\mathcal{F}) \leq x_{ub}$ and $[0\, \Pi]x^{(opt)}(\mathcal{F}) \geq x_{lb}$, because, clearly, $[0\, \Pi]x^{(opt)}(\mathcal{F}) \leq \Pi x(t) \leq [\Pi\, 0]x^{(opt)}(\mathcal{F})$, $\forall t$. However, this is obviously too expensive to compute directly using Equation (16), although it is possible to get an approximate value of the solution by solving for only a few terms of the summation. Instead, we will use some bounds on $x^{(opt)}$ based on the following notation:

Definition 3.6. If $v = \left[\begin{smallmatrix} v_t \\ v_b \end{smallmatrix}\right]$ and $w = \left[\begin{smallmatrix} w_t \\ w_b \end{smallmatrix}\right]$ are $2n \times 1$ vectors and $v_t,\ v_b,\ w_t$, and w_b are $n \times 1$ vectors, then we say that $v \subseteq w$ if

$$v_t \leq w_t \qquad \text{and} \qquad v_b \geq w_b. \tag{17}$$

Definition 3.7. If $v = \left[\begin{smallmatrix} v_t \\ v_b \end{smallmatrix}\right]$ and $w = \left[\begin{smallmatrix} w_t \\ w_b \end{smallmatrix}\right]$ are $2n \times 1$ vectors and $v_t,\ v_b,\ w_t$, and w_b are $n \times 1$ vectors, then we say that $v \subset w$ if

$$v_t < w_t \qquad \text{and} \qquad v_b > w_b. \tag{18}$$

A few simple properties can be stated without proof.

1) The subset relation among vectors is *transitive*:

$$u \subseteq v \quad \text{and} \quad v \subseteq w \quad \Longrightarrow \quad u \subseteq w, \tag{19}$$
$$u \subset v \quad \text{and} \quad v \subset w \quad \Longrightarrow \quad u \subset w, \tag{20}$$

2) The subset relation may be combined by *summation*:

$$u \subseteq v \quad \text{and} \quad w \subseteq z \quad \Longrightarrow \quad u + w \subseteq v + z, \tag{21}$$
$$u \subset v \quad \text{and} \quad w \subset z \quad \Longrightarrow \quad u + w \subset v + z. \tag{22}$$

3) For any two vectors u and v, we have

$$u \subseteq v \quad \Longleftrightarrow \quad -v \subseteq -u, \tag{23}$$
$$u \subset v \quad \Longleftrightarrow \quad -v \subset -u. \tag{24}$$

4) For any two vectors u and v, we have

$$u \subseteq v \quad \Longleftrightarrow \quad 0 \subseteq v - u, \tag{25}$$
$$u \subset v \quad \Longleftrightarrow \quad 0 \subset v - u. \tag{26}$$

Definition 3.8 (*Matrix Extension*). Let W be an $n \times n$ matrix, and let $W^+ = \frac{1}{2}(W + |W|)$ and $W^- = \frac{1}{2}(W - |W|)$, where $|W|$ is the $n \times n$ matrix consisting of the absolute values of the elements of W. We define the *extension* of W as the $2n \times 2n$ matrix $\tilde{W}$, given by

$$\tilde{W} = \begin{bmatrix} W^+ & W^- \\ W^- & W^+ \end{bmatrix}. \tag{27}$$

Notice that $W^+ \geq 0$ consists of only the non-negative elements of W while $W^- \leq 0$ consists of only the non-positive elements of W, so, with w_{ij} denoting the (i, j)th entry of W, we have for any (i, j):

$$w_{ij}^+ = \begin{cases} w_{ij}, & \text{if } w_{ij} \geq 0; \\ 0, & \text{otherwise.} \end{cases} \qquad w_{ij}^- = \begin{cases} w_{ij}, & \text{if } w_{ij} \leq 0; \\ 0, & \text{otherwise.} \end{cases} \tag{28}$$

Definition 3.9 (*Subset-Preserving*). A $2n \times 2n$ matrix X is said to be *subset-preserving* (SP) if, for any two $2n \times 1$ vectors u, v, we have that $u \subseteq v \Longrightarrow Xu \subseteq Xv$.

From Lemma A.1 in the appendix, we have that, for any matrix W, its extension $\tilde{W}$ is SP.

Because $x^{(opt)}$ is expensive to compute, the authors in Fawaz and Najm [2016] derive a bound $\overline{x}$ on $x^{(opt)}$, which is stated in Definition 3.10. The authors show that for a certain range of the discretization timestep Δt we have $\rho(\tilde{F}) < 1$, and $(I_{2n} - \tilde{F})$ is non-singular, where I_{2n} is the $2n \times 2n$ identity matrix. Furthermore, they show that it is always possible to choose a Δt in that range, it is easy to find such a Δt, and that the choice of Δt is good for the accuracy of $\overline{x}$. Throughout the rest of this document, we will assume that Δt is in such a range, so

$$\rho(\tilde{F}) < 1. \tag{29}$$

With this, define Q to be the $2n \times 2n$ matrix:

$$Q \triangleq (I_{2n} - \tilde{F})^{-1}. \tag{30}$$

Definition 3.10. For any $\mathcal{F} \subset \mathbb{R}^m$, define

$$\overline{x}(\mathcal{F}) \triangleq Q \operatorname*{eopt}_{I \in \mathcal{F}}(RI), \tag{31}$$

where $I \in \mathbb{R}^m$ is a vector of artificial variables, with units of current, that is used to carry out the eopt$(\cdot)$ operation.

In Fawaz and Najm [2016], the authors have shown that $\overline{x}$ is a bound on $x^{(opt)}$ for any container $\mathcal{F}$:

$$x^{(opt)}(\mathcal{F}) \subseteq \overline{x}(\mathcal{F}). \tag{32}$$

Let P be a $2d \times 2n$ matrix defined as follows:

$$P = \begin{bmatrix} \Pi & 0_{d \times n} \\ 0_{d \times n} & \Pi \end{bmatrix}. \tag{33}$$

Using Lemma A.1, and because $\Pi \geq 0$, we have that P is SP and, of course, $P \geq 0$.

LEMMA 3.11. *For any $u, u' \in \mathbb{R}^{2n}$, if $u \subset u'$, then $Pu \subset Pu'$, where P is the $2d \times 2n$ matrix defined in Equation* (33).

PROOF. Let $u = \left[\begin{smallmatrix} u_t \\ u_b \end{smallmatrix}\right]$, $u' = \left[\begin{smallmatrix} u'_t \\ u'_b \end{smallmatrix}\right]$, $v = Pu = \left[\begin{smallmatrix} v_t \\ v_b \end{smallmatrix}\right]$, and $v' = Pu' = \left[\begin{smallmatrix} v'_t \\ v'_b \end{smallmatrix}\right]$, where u_t, u'_t, u_b, and u'_b are $n \times 1$ vectors and v_t, v'_t, v_b, and v'_b are $d \times 1$ vectors. Notice that $u_t < u'_t$ and $u_b > u'_b$, because $u \subset u'$, so $\Pi u_t < \Pi u'_t$ and $\Pi u_b > \Pi u'_b$, because $\Pi \geq 0$ and Π has no row with all zeros. It follows that $v_t < v'_t$ and $v_b > v'_b$ so $Pu = v \subset v' = Pu'$. □

Definition 3.12 (*Safe Container*). For a given container $\mathcal{F}$, we say that $\mathcal{F}$ is safe if $P\overline{x}(\mathcal{F}) \subseteq x_{th}$.

For the example of Figure 1, one can simply find $\overline{x}(\mathcal{F})$ defined in Equation (31) for the current container $\mathcal{F}$ represented in Figure 4. Because only nodes 1 and 3 are of interest, then $P\overline{x}(\mathcal{F}) = [v_1^{(ub)}\ v_3^{(ub)}\ v_1^{(lb)}\ v_3^{(lb)}]^T$, where $v_1^{(ub)}$ and $v_3^{(ub)}$ are the worst-case voltage drop on nodes 1 and 3, respectively, and $v_1^{(lb)}$ and $v_3^{(lb)}$ are the worst-case voltage overshoot on nodes 1 and 3, respectively, under all possible current waveforms $i(t) \in \mathcal{F}$. In Figure 5, we show the values of $v_1^{(ub)}$, $v_3^{(ub)}$, $v_1^{(lb)}$, and $v_3^{(lb)}$ that clearly satisfy the voltage thresholds, so $P\overline{x}(\mathcal{F}) \subseteq x_{th}$, and $\mathcal{F}$ is safe.

Thus, we are interested to discover a safe container $\mathcal{F}$ so, due to Equation (32) and P being SP, we get $Px^{(opt)}(\mathcal{F}) \subseteq x_{th}$, and the grid is safe. We will see below that a safe container $\mathcal{F}$ can be expressed as a set of constraints on the circuit currents that load the grid, thereby providing a set of constraints that are sufficient to guarantee grid safety.

4. MAXIMAL CONTAINERS

This section contains the bulk of the theoretical contributions of this article and is organized as follows. First, we will establish a necessary and sufficient condition for a container to be safe. We will find, however, that there is an infinity of possible safe containers. The question becomes the following: Which safe container should we choose? Recall that a container can be used to drive parts of the design process such as automated floorplanning and placement, and, hence, the "larger" a container is, the more flexibility is provided for the rest of the design stages. We are interested in containers that allow as much flexibility as possible in the circuit loading currents, which we will refer to as *maximal containers*. The following definition captures the notion of maximal containers in mathematical terms.

Definition 4.1. Let $\mathcal{E}$ be a collection of subsets of $\mathbb{R}^m$ and let $\mathcal{X} \in \mathcal{E}$. We say that $\mathcal{X}$ is *maximal* in $\mathcal{E}$ if there does not exist another $\mathcal{Y} \in \mathcal{E}$, $\mathcal{Y} \neq \mathcal{X}$, such that $\mathcal{X} \subseteq \mathcal{Y}$.

Notice that a container $\mathcal{F}$ is a subset of $\mathbb{R}^m$. In the following, we will identify a set $\mathcal{E}$ that is an infinite collection of safe containers. These containers will be of the form of Equation (36). In fact, we will show that these are the only interesting containers. Finally, we provide the necessary and sufficient conditions for a container $\mathcal{X}$ to be maximal in $\mathcal{E}$. These conditions are given in Theorem 4.13 and depend on several results proved in Sections 4.1 and 4.2.

Let $T = Q^{-1}$, so

$$T = I_{2n} - \tilde{F}. \tag{34}$$

Let $u \in \mathbb{R}^{2n}$ and define the sets $\mathcal{U}$ and $\mathcal{F}(u)$ as follows:

$$\mathcal{U} \triangleq \{u \in \mathbb{R}^{2n} : Pu \subseteq x_{th}\}, \tag{35}$$

$$\mathcal{F}(u) \triangleq \{I \in \mathbb{R}^m : I \geq 0,\ RI \in Tu\}, \tag{36}$$

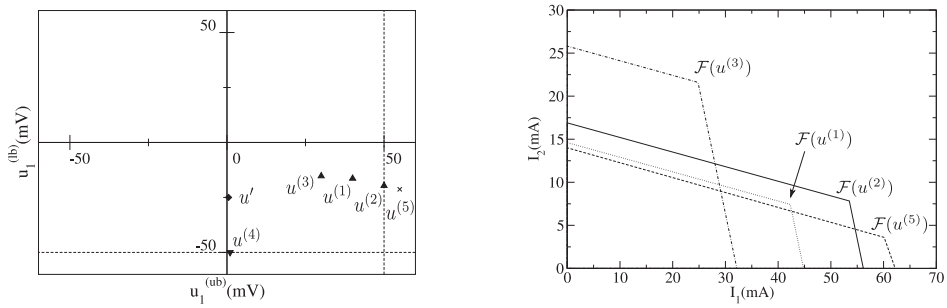
and notice that

$$Tu \subseteq Tu' \Longrightarrow \mathcal{F}(u) \subseteq \mathcal{F}(u'), \quad \forall u, u' \in \mathbb{R}^{2n}. \tag{37}$$

To graphically illustrate the relation between the sets $\mathcal{U}$ and $\mathcal{F}(u)$, consider the circuit in Figure 1 and suppose that only node 1 is required to satisfy the voltage safety specifications $|v_1(t)| \leq 50$mV. Notice that in this case we have $n = 5$ and $d = 1$, so for any $u \in \mathbb{R}^{10}$, $Pu \in \mathbb{R}^2$, and, for any $u \in \mathcal{U}$, we have $Pu \subseteq x_{th} = [\begin{smallmatrix} 50 \\ -50 \end{smallmatrix}]$mV. In Figure 6, we show several instances of u and plot their corresponding Pu and $\mathcal{F}(u)$.

For the instance $u^{(2)}$, the resulting current container is $\mathcal{F}(u^{(2)})$, which, as defined in Equation (36), can be represented using the following set of inequalities, with $I = [\begin{smallmatrix} I_1 \\ I_2 \end{smallmatrix}] \geq 0$:

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -0.051 \end{bmatrix} \leq \begin{bmatrix} 0.76 & 0.26 \\ 0.30 & 0.27 \\ 0.26 & 1.53 \\ 0.04 & 0.03 \\ -0.86 & -0.78 \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \leq \begin{bmatrix} 0.043 \\ 0.017 \\ 0.022 \\ 0.002 \\ 0 \end{bmatrix}. \tag{38}$$



(a) Examples of Pu, where the top part of Pu is denoted as $u_1^{(ub)}$ and the lower part of Pu is denoted as $u_1^{(lb)}$.

(b) Examples of $\mathcal{F}(u)$ with $\mathcal{F}(u^{(4)}) = \mathcal{F}(u') = \phi$.

Fig. 6. Examples of Pu and $\mathcal{F}(u)$.

The above set of inequalities on the current variable I defines a region in $\mathbb{R}^2$ as shown in Figure 6.

LEMMA 4.2. *For any $u \in \mathbb{R}^{2n}$, $0 \in \mathcal{F}(u)$ if and only if $0 \in Tu$.*

PROOF. Let $u \in \mathbb{R}^{2n}$ with $0 \in Tu$, so for $I = 0$ we have $RI = 0 \in Tu$, from which $0 \in \mathcal{F}(u)$. Conversely, let $u \in \mathbb{R}^{2n}$ with $0 \in \mathcal{F}(u)$, and then $0 \in Tu$ due to Equation (36). □

Definition 4.3. For any $u \in \mathbb{R}^{2n}$, u is said to be *feasible* if $\mathcal{F}(u)$ is not empty, otherwise it is *infeasible*.

For illustration, notice that in Figure 6, the instance $u^{(4)}$ generates a current container $\mathcal{F}(u^{(4)})$ represented using the following set of inequalities, with $I = [\begin{smallmatrix} I_1 \\ I_2 \end{smallmatrix}] \geq 0$:

$$\begin{bmatrix} -0.032 \\ 0 \\ 0 \\ 0 \\ -0.041 \end{bmatrix} \leq \begin{bmatrix} 0.76 & 0.26 \\ 0.30 & 0.27 \\ 0.26 & 1.53 \\ 0.04 & 0.03 \\ -0.86 & -0.78 \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \leq \begin{bmatrix} -0.002 \\ 0.014 \\ 0.040 \\ 0.002 \\ -0.01 \end{bmatrix}. \tag{39}$$

It is easy to check that the above set of inequalities is unsatisfiable for any $I \geq 0$, so $\mathcal{F}(u^{(4)}) = \phi$ and $u^{(4)}$ is infeasible, whereas the current container $\mathcal{F}(u^{(2)})$, as shown in Figure 6, is non-empty, so $u^{(2)}$ is feasible.

LEMMA 4.4. *For any feasible $u \in \mathbb{R}^{2n}$, we have $\overline{x}(\mathcal{F}(u)) \subseteq u$.*

PROOF. For any feasible $u \in \mathbb{R}^{2n}$, we have $RI \in Tu$, for all $I \in \mathcal{F}(u)$, due to Ref. (36), so

$$\operatorname*{eopt}_{I \in \mathcal{F}(u)} (RI) \subseteq Tu. \tag{40}$$

Because Q is SP, due to Lemma E.8 (in the appendix), it follows that

$$Q \operatorname*{eopt}_{I \in \mathcal{F}(u)} (RI) \subseteq QTu = u. \tag{41}$$

Therefore, $\overline{x}(\mathcal{F}(u)) = Q\,\mathrm{eopt}_{I \in \mathcal{F}(u)}(RI) \subseteq u$, and the proof is complete. □

LEMMA 4.5. *For any non-empty container $\mathcal{J} \subset \mathbb{R}^m_+$, $\mathcal{J}$ is safe if and only if $\exists u \in \mathcal{U}$ such that $\mathcal{J} \subseteq \mathcal{F}(u)$.*

PROOF. The proof is in two parts.

Proof of the "if direction": Let $\mathcal{J} \subseteq \mathcal{F}(u)$ for some $u \in \mathcal{U}$, it follows that $\text{eopt}_{I \in \mathcal{J}}(RI) \subseteq \text{eopt}_{I \in \mathcal{F}(u)}(RI)$, from which $\overline{x}(\mathcal{J}) \subseteq \overline{x}(\mathcal{F}(u))$, due to Lemma E.8. Notice that $\mathcal{F}(u)$ is not empty, because $\mathcal{J}$ is not empty, so u is feasible. Using Lemma 4.4 and Equation (19), we get $\overline{x}(\mathcal{J}) \subseteq u$, which, because P is SP and $u \in \mathcal{U}$, gives $P\overline{x}(\mathcal{J}) \subseteq x_{th}$.

Proof of the "only if direction": Let $\mathcal{J} \subset \mathbb{R}^m_+$ be a non-empty set with $P\overline{x}(\mathcal{J}) \subseteq x_{th}$, and let $u = \overline{x}(\mathcal{J})$, so $Pu \subseteq x_{th}$, from which $u \in \mathcal{U}$. Multiplying $u = \overline{x}(\mathcal{J}) = Q\,\text{eopt}_{I \in \mathcal{J}}(RI)$ with T, we get

$$\underset{I \in \mathcal{J}}{\text{eopt}}(RI) = Tu, \tag{42}$$

so, $\forall I \in \mathcal{J}$, we have $RI \in Tu$, which, coupled with $I \geq 0$, gives $\mathcal{J} \subseteq \mathcal{F}(u)$. □

Definition 4.6. Define the set of containers:

$$\mathcal{S} \triangleq \{\mathcal{F}(u) : u \in \mathcal{U}\}. \tag{43}$$

It should be clear from Lemma 4.5 that all containers of interest are members of $\mathcal{S}$ or subsets of members of $\mathcal{S}$. Note that if $\mathcal{J} \subseteq \mathcal{F}(u)$ for some $u \in \mathcal{U}$, with $\mathcal{J} \neq \mathcal{F}(u)$, then, clearly, $\mathcal{F}(u)$ is a better choice than $\mathcal{J}$. Choosing $\mathcal{J}$ would be unnecessarily limiting, while $\mathcal{F}(u)$ would allow more flexibility in the circuit loading currents. Therefore, it is enough to consider *only* containers of the form $\mathcal{F}(u)$ with $u \in \mathcal{U}$. This is why the definitions (35), (36), and (43) are important, and we refer to $\mathcal{S}$ as the set of safe containers.

Referring to Figure 6, the instance $u^{(5)} \notin \mathcal{U}$ and generates the current container $\mathcal{F}(u^{(5)})$, represented using the following set of inequalities, with $I = \left[\begin{smallmatrix} I_1 \\ I_2 \end{smallmatrix}\right] \geq 0$:

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -0.055 \end{bmatrix} \leq \begin{bmatrix} 0.76 & 0.26 \\ 0.30 & 0.27 \\ 0.26 & 1.53 \\ 0.04 & 0.03 \\ -0.86 & -0.78 \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \leq \begin{bmatrix} 0.047 \\ 0.019 \\ 0.021 \\ 0.002 \\ 0 \end{bmatrix}. \tag{44}$$

It can be easily verified by computing $\overline{x}(\cdot)$ in Definition 3.10 that the current container defined by the above set of inequalities gives $P\overline{x}(\mathcal{F}(u^{(5)}) = \left[\begin{smallmatrix} 55 \\ -21 \end{smallmatrix}\right] mV \not\subseteq x_{th}$ and thus is an unsafe container. However, $u^{(2)} \in \mathcal{U}$ generates the current container defined by the set of inequalities (38) and satisfies $P\overline{x}(\mathcal{F}(u^{(2)}) \subseteq x_{th}$, so $\mathcal{F}(u^{(2)})$ is a safe container.

Going further, if $\mathcal{F}(u_1) \subseteq \mathcal{F}(u_2)$ with $\mathcal{F}(u_1) \neq \mathcal{F}(u_2)$, then, clearly, $\mathcal{F}(u_2)$ is a better choice than $\mathcal{F}(u_1)$. In a sense, the "larger" the container, the better. Thus, we are only interested in containers $\mathcal{F}(u)$ that are maximal in $\mathcal{S}$.

In the example of Figure 6, the current container $\mathcal{F}(u^{(1)})$ represented using the following set of inequalities, for $I = \left[\begin{smallmatrix} I_1 \\ I_2 \end{smallmatrix}\right] \geq 0$:

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -0.042 \end{bmatrix} \leq \begin{bmatrix} 0.76 & 0.26 \\ 0.30 & 0.27 \\ 0.26 & 1.53 \\ 0.04 & 0.03 \\ -0.86 & -0.78 \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \leq \begin{bmatrix} 0.034 \\ 0.014 \\ 0.022 \\ 0.002 \\ 0 \end{bmatrix} \tag{45}$$

and the current container $\mathcal{F}(u^{(2)})$ represented using Equation (38) satisfy $\mathcal{F}(u^{(1)}) \subseteq \mathcal{F}(u^{(2)})$, with $\mathcal{F}(u^{(1)}) \neq \mathcal{F}(u^{(2)})$, so $\mathcal{F}(u^{(1)})$ is not maximal in $\mathcal{S}$, whereas $\mathcal{F}(u^{(2)})$ can be shown to be maximal in $\mathcal{S}$.

Maximality is a highly desirable property, and so the purpose of the rest of this section is to give necessary and sufficient conditions for a container to be maximal in

$\mathcal{S}$. We will see that the maximality of $\mathcal{F}(u)$ depends on crucial properties of u. Note that $0 \in \mathcal{U}$ for any $x_{th} \ni 0$, and $0 \in \mathcal{F}(0)$, so $u = 0$ is always feasible. It follows that $\mathcal{S}$ always contains a non-empty set as a member, so ϕ is never maximal in $\mathcal{S}$; this will be useful below.

4.1. Extremal

Throughout the rest of the article, whenever the product of a number of matrices A_i by a vector v is followed by the notation $|_i$, as in $A_1 A_2 \cdots A_k v|_i$, the expression shall denote the ith entry of the vector resulting from the product $A_1 A_2 \cdots A_k v$.

Definition 4.7. For any $u \in \mathcal{U}$, we say that u is extremal in $\mathcal{U}$ if $\exists k \in \{1, \ldots, 2d\}$ such that $Pu|_k = x_{th,k}$.

Notice that in Figure 6, $Pu^{(2)}|_1 = x_{th,1}$ so $u^{(2)}$ is extremal in $\mathcal{U}$, whereas $Pu^{(1)}|_k \neq x_{th,k}$, $\forall k$, so $u^{(1)}$ is not extremal in $\mathcal{U}$.

LEMMA 4.8. *If $\mathcal{F}(u)$ is maximal in $\mathcal{S}$, then u is feasible and extremal in $\mathcal{U}$.*

The proof of Lemma 4.8 is available in Appendix B as Lemma B.2.

4.2. Irreducible

Definition 4.9. We say that $u \in \mathbb{R}^{2n}$ is *reducible* if there exists $u' \subseteq u$, $u' \neq u$, with $\mathcal{F}(u') = \mathcal{F}(u)$, otherwise u is said to be *irreducible*.

We will see that irreducibility of u is a crucial property that is required for maximality of $\mathcal{F}(u)$. The proofs of Lemmas 4.10, 4.11, and 4.12 are available in Appendix C as Lemmas C.2, C.3, and C.5.

LEMMA 4.10. *For any feasible $u \in \mathbb{R}^{2n}$, let $u' = \overline{x}(\mathcal{F}(u))$, and it follows that $\mathcal{F}(u') = \mathcal{F}(u)$.*

LEMMA 4.11. *For any $u \in \mathbb{R}^{2n}$, u is irreducible if and only if it is feasible and $\overline{x}(\mathcal{F}(u)) = u$.*

Note that if u is irreducible and extremal in $\mathcal{U}$, then $Pu|_k = x_{th,k}$ for some k, and so $P\overline{x}(\mathcal{F}(u))|_k = x_{th,k}$.

LEMMA 4.12. *For any $u \in \mathbb{R}^{2n}$, u is irreducible if and only if*

$$Tu \subseteq Tu' \iff \mathcal{F}(u) \subseteq \mathcal{F}(u'), \quad \forall u' \in \mathbb{R}^{2n}. \tag{46}$$

4.3. Maximality

As pointed out earlier, we are interested in safe containers that are maximal in $\mathcal{S}$. We now present our main result that gives the necessary and sufficient conditions for maximality.

THEOREM 4.13. *$\mathcal{F}(u)$ is maximal in $\mathcal{S}$ if and only if u is irreducible and extremal in $\mathcal{U}$.*

PROOF. The proof is in two parts.

Proof of the "if direction": We give a proof by contradiction. Let $u \in \mathcal{U}$ be irreducible and extremal in $\mathcal{U}$, but suppose that $\mathcal{F}(u)$ is not maximal in $\mathcal{S}$, so $\exists u' \in \mathcal{U}$ such that $\mathcal{F}(u) \subseteq \mathcal{F}(u')$, with $\mathcal{F}(u) \neq \mathcal{F}(u')$. Because $\mathcal{F}(u) \neq \mathcal{F}(u')$, then clearly $Tu \neq Tu'$, and, using Lemma 4.12, we have $Tu \subseteq Tu'$. Let $\delta = Tu' - Tu$, so $0 \subseteq \delta$ and $\delta \neq 0$. Recall that Q is SP, from Lemma E.8, so $0 \subset Q\delta$. Let $w = Q\delta$. Denote by w_i, δ_j, and q_{ij} the ith entry

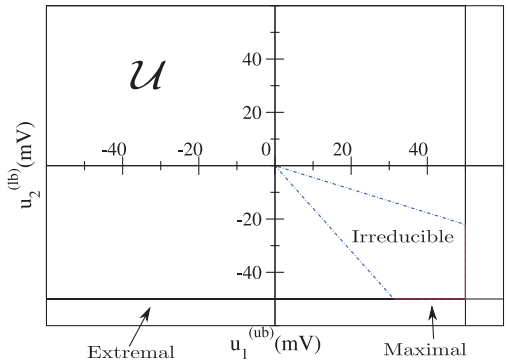


Fig. 7. A graphical representation of the set $\mathcal{U}$, vectors u that are irreducible, vectors u that are extremal in $\mathcal{U}$, and vectors u that have $\mathcal{F}(u)$ maximal in $\mathcal{S}$.

of w, the jth entry of δ, and the (i, j)th entry of Q, respectively. Notice that

$$w_i = \sum_{j=1}^{2n} q_{ij}\delta_j, \tag{47}$$

where $q_{ij} \neq 0, \forall i, j$, due to Lemma E.6. For $i \in \{1, \ldots, n\}$, we have $q_{ij}\delta_j \geq 0, \forall j$, because Q is SP and $0 \subseteq \delta$, which, combined with $q_{ij} \neq 0$ and $\delta \neq 0$, gives $w_i > 0$. Similarly, for $i \in \{n+1, \ldots, 2n\}$, we have $q_{ij}\delta_j \leq 0, \forall j$, because Q is SP and $0 \subseteq \delta$, which, combined with $q_{ij} \neq 0$ and $\delta \neq 0$, gives $w_i < 0$. Therefore, we have $0 \subset w = Q\delta = u' - u$. Then $u \subset u'$, due to Equation (26), and $Pu \subset Pu' \subseteq x_{th}$, making use of Lemma 3.11 and the final step due to $u' \in \mathcal{U}$, so u is not extremal in $\mathcal{U}$, and we have a contradiction that completes the proof.

Proof of the "only if direction": We give a proof by contradiction. Given that $\mathcal{F}(u)$ is maximal in $\mathcal{S}$, we know from Lemma 4.8 that u is feasible and extremal in $\mathcal{U}$. Suppose u is reducible, so $\overline{x}(\mathcal{F}(u)) \neq u$, because we already have that u is feasible. Recall that $\overline{x}(\mathcal{F}(u)) \subseteq u$, from which $P\overline{x}(\mathcal{F}(u)) \subseteq Pu \subseteq x_{th}$, because P is SP and $u \in \mathcal{U}$. Let $u' = \overline{x}(\mathcal{F}(u)) \neq u$, so $u' \in \mathcal{U}$ and $Tu' = T\overline{x}(\mathcal{F}(u)) = \text{eopt}_{I \in \mathcal{F}(u)}(RI)$. Let $\delta = Tu - Tu' = Tu - \text{eopt}_{I \in \mathcal{F}(u)}(RI)$, and then we have $0 \subseteq \delta$ due to $\text{eopt}_{I \in \mathcal{F}(u)}(RI) \subseteq Tu$ and $\delta \neq 0$ (due to $u' \neq u$). Recall that Q is SP, from Lemma E.8, and every element of Q is non-zero, due to Lemma E.6, so $0 \subset Q\delta = u' - u$; a more detailed explanation of this step was presented in the first part of the proof. Consequently, we have $u' \subset u$, due to Equation (26), so $Pu' \subset Pu \subseteq x_{th}$, making use of Lemma 3.11, and the final step is due to $u \in \mathcal{U}$, so u' is not extremal in $\mathcal{U}$. Therefore, by Lemma 4.8, $\mathcal{F}(u')$ is not maximal in $\mathcal{S}$. However, $\mathcal{F}(u) = \mathcal{F}(u')$, due to Lemma 4.10, so $\mathcal{F}(u)$ is not maximal in $\mathcal{S}$, a contradiction that completes the proof. □

In Figure 7, we give a graphical representation of the set $\mathcal{U}$ for the same example as in Figure 6. Sweeping over all values of $u \in \mathcal{U}$ and checking the conditions of Lemma E.5, we can discover the set of irreducible vectors u. We represent the set of vectors u that are irreducible as the double-shaded polygon. Notice that the set of vectors u that are extremal in $\mathcal{U}$ have Pu on the boundary of $\mathcal{U}$. Thus, the set of vectors u that have $\mathcal{F}(u)$ maximal in $\mathcal{S}$ is the intersection of both irreducible and extremal sets, due to Theorem 4.13, represented in red on the boundary of $\mathcal{U}$.

This important theoretical result forms the basis for our choice of practical constraints generation algorithms that are guaranteed to give maximal containers, as we will see in the next section. Recall that whenever u is irreducible and extremal in $\mathcal{U}$, then $P\overline{x}(\mathcal{F}(u))|_k = x_{th,k}$, for some k, so the bound on the voltage variation at the kth

grid node would be equal to its maximum allowable voltage variation. In other words, a maximal container always causes some node(s) on the grid to experience the maximum allowable voltage variation, at least based on the $\overline{x}(\cdot)$ bound.

5. APPLICATIONS

So far, we have shown that a container $\mathcal{F}(u)$ is maximal in $\mathcal{S}$ if and only if u satisfies the conditions of Theorem 4.13. Note that it is possible to find a container $\mathcal{F}(u)$ that is maximal in $\mathcal{S}$ but does not include the *zero state*, that is, the state $I = 0$. We believe that users are interested in containers that include the zero state, and, thus, we will enforce this constraint when searching for maximal containers $\mathcal{F}(u)$. In this section, we will describe some design objectives and corresponding algorithms for finding specific maximal safe containers $\mathcal{F}(u)$, with the additional condition that $0 \in \mathcal{F}(u)$. These algorithms will each be formulated as a maximization of a certain design objective $g(u)$. Lemma D.1 in Appendix D establishes a sufficient condition on $g(\cdot)$ for which the algorithms proposed below produce maximal containers, based on Theorem 4.13.

5.1. Peak Power Dissipation

An interesting quality metric for a power grid is the peak total power dissipation that it can safely support in the underlying circuit. We refer here to the instantaneous power dissipation, which is conservatively approximated by $V_{dd} \sum_{j=1}^{m} i_j(t)$. Thus, we are interested in a safe container that is maximal in $\mathcal{S}$ and that allows the highest possible $\sum_{\forall j} I_j$. Generally, one might be interested in the highest *weighted* sum of the individual currents, that is, $\sum_{\forall j} q_j I_j$, where $q_j > 0$ is a user-specified weight on the jth current source. This will allow certain areas of the die to support larger power dissipation than other areas. However, in this article, we assume that all current sources have equal weights, and, hence, we will be finding the peak total power dissipation that the grid can safely support.

For any $u \in \mathcal{U}$, we define $\sigma(u)$ to be the largest sum of current source values allowed under $\mathcal{F}(u)$:

$$\sigma(u) \stackrel{\triangle}{=} \max_{I \in \mathcal{F}(u)} \left(\sum_{j=1}^{m} I_j \right), \tag{48}$$

and we define σ^* to be the largest $\sigma(u)$ achievable over all possible $u \in \mathcal{U}$ with $0 \in \mathcal{F}(u)$, that is,

$$\sigma^* \stackrel{\triangle}{=} \max_{\substack{u \in \mathcal{U} \\ 0 \in \mathcal{F}(u)}} (\sigma(u)) = \max_{\substack{u \in \mathcal{U} \\ 0 \in Tu}} (\sigma(u)), \tag{49}$$

where the second equality is due to Lemma 4.2.

Let $u_p \in \mathcal{U}$ be such that $\sigma(u_p) = \sigma^*$ and $I^* \in \mathcal{F}(u_p)$ be such that $\sum_{j=1}^{m} I_j^* = \sigma^*$. In general, u_p and I^* may not be unique. Based on Equations (35) and (36), and making use of Lemma 4.2, we can express the combined Equations (48) and (49) as the following LP:

$$\textbf{LP1:} \quad \begin{array}{rll} \sigma^* = & \text{Maximize} & \sum_{j=1}^{m} I_j \\ & \text{subject to} & RI \in Tu \\ & & 0 \in Tu \\ & & Pu \subseteq x_{th} \\ & & I \geq 0. \end{array} \tag{50}$$

Let $\mathcal{D}$ be the feasible region of the LP (50):

$$\mathcal{D} \stackrel{\triangle}{=} \{(I, u) : I \geq 0, RI \in Tu, 0 \in Tu, Pu \subseteq x_{th}\}, \tag{51}$$

so from the above we have

$$\sigma^* = \max_{(I,u)\in\mathcal{D}} \left(\sum_{j=1}^{m} I_j \right). \tag{52}$$

Notice that $(0,0) \in \mathcal{D}$, so $\mathcal{D}$ is not empty and all of σ^*, u_p, and I^* are well defined. Therefore, the container $\mathcal{F}(u_p) = \{I \in \mathbb{R}^m : I \geq 0, RI \in Tu_p\} \neq \phi$ provides the desired current constraints, $\forall t \in \mathbb{R}$,

$$i(t) \geq 0$$
$$Ri(t) \in Tu_p.$$

The following lemma establishes the maximality of $\mathcal{F}(u_p)$, based on Lemma D.1.

LEMMA 5.1. *$\mathcal{F}(u_p)$ is maximal in $\mathcal{S}$.*

PROOF. Recall that I^* and u_p are well defined and $(I^*, u_p) \in \mathcal{D}$, so $RI^* \in Tu_p$ and $I^* \geq 0$, from which u_p is feasible. We will prove that $\sigma(\cdot)$ satisfies the conditions of Lemma D.1, from which $\mathcal{F}(u_p)$ is maximal in $\mathcal{S}$. First, notice that for any $u, u' \in \mathcal{U}$, if $\mathcal{F}(u') = \mathcal{F}(u)$, then it follows that $\sigma(u') = \sigma(u)$, due to Equation (48). It remains to prove that for any $u, u' \in \mathcal{U}$, with $0 \in Tu$ and $0 \in Tu'$, if $Tu' \supset Tu$, then $\sigma(u') > \sigma(u)$.

For any $u \in \mathcal{U}$, let $I \in \mathcal{F}(u)$ be such that $\sigma(u) = \sum_{j=1}^{m} I_j$. Let $\lambda = \min_{\forall i}(|Tu'|_i - Tu|_i|)/\max_{\forall i,j}(|r_{ij}|)$, which is well defined due to $R \neq 0$, and $\lambda > 0$ because $Tu \subset Tu'$. Also, let $e_1 \in \mathbb{R}^m$ be the vector whose first entry is 1 and all other entries are 0 and let $I' = I + \lambda e_1$. Because $\lambda > 0$, we have $\lambda e_1 \geq 0$, $\lambda e_1 \neq 0$, $I' \geq I \geq 0$, and $I' \neq I$, so $\sum_{j=1}^{m} I'_j > \sum_{j=1}^{m} I_j$. Denote by c_j the jth column of R and notice that

$$RI' = RI + \lambda Re_1 = RI + \lambda c_1, \tag{53}$$
$$= RI + \frac{\min_{\forall i}\left(|Tu'|_i - Tu|_i|\right)}{\max_{\forall i,j}(|r_{ij}|)} c_1. \tag{54}$$

Let $\mathbb{1}_{2n}$ be the $2n \times 1$ vector whose first n entries are 1 and the rest are -1. Because $c_1/\max_{\forall i,j}(|r_{ij}|) \in \mathbb{1}_{2n}$, notice that

$$\frac{\min_{\forall i}\left(|Tu'|_i - Tu|_i|\right)}{\max_{\forall i,j}(|r_{ij}|)} c_1 \in \min_{\forall i}(|Tu'|_i - Tu|_i|)\mathbb{1}_{2n}, \tag{55}$$

which, combined with $RI \in Tu$ because $(I,u) \in \mathcal{D}$, and due to Lemma 3.3, gives

$$RI + \frac{\min_{\forall i}\left(|Tu'|_i - Tu|_i|\right)}{\max_{\forall i,j}(|r_{ij}|)} c_1 \in Tu + \min_{\forall i}(|Tu'|_i - Tu|_i|)\mathbb{1}_{2n}. \tag{56}$$

Therefore, using Equation (54), it follows that

$$RI' \in Tu + \min_{\forall i}(|Tu'|_i - Tu|_i|)\mathbb{1}_{2n}. \tag{57}$$

Also, notice that, for any $k \in \{1, \ldots, n\}$, because $Tu \subset Tu'$, we have

$$\min_{\forall i}(|Tu'|_i - Tu|_i|) \leq |Tu'|_k - Tu|_k| = Tu'|_k - Tu|_k. \tag{58}$$

Likewise, for any $k \in \{n+1, \ldots, 2n\}$, we have

$$-\min_{\forall i}(|Tu'|_i - Tu|_i|) \geq -|Tu'|_k - Tu|_k| = Tu'|_k - Tu|_k. \tag{59}$$

Combining Equations (58) and (59), we get

$$\min_{\forall i}(|Tu'|_i - Tu|_i|)\mathbb{1}_{2n} \subseteq Tu' - Tu. \tag{60}$$

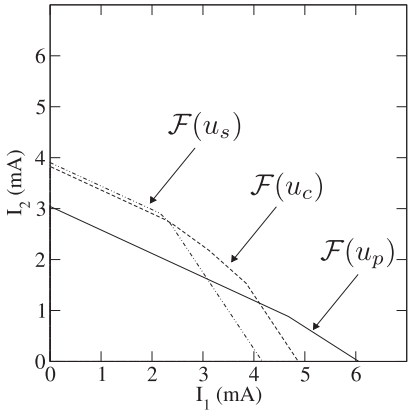


Fig. 8. An example of $\mathcal{F}(u_p)$, $\mathcal{F}(u_s)$, and $\mathcal{F}(u_c)$.

This, combined with $Tu \subseteq Tu$ and making use of Equation (21), gives

$$Tu + \min_{\forall i}(|Tu'|_i - Tu|_i|)\mathbb{1}_{2n} \subseteq Tu + Tu' - Tu = Tu'. \tag{61}$$

Therefore, due to Equations (57) and (61), we get

$$RI' \in Tu'. \tag{62}$$

This, coupled with $u' \in \mathcal{U}$, means that $(I', u') \in \mathcal{D}$, so $\sigma(u') \geq \sum_{j=1}^{m} I'_j > \sum_{j=1}^{m} I_j = \sigma(u)$, from which $\sigma(\cdot)$ satisfies the conditions of Lemma D.1, and $\mathcal{F}(u_p)$ is maximal in $\mathcal{S}$. □

As an example, the LP (50) is run on a 17 nodes grid with two current sources, and the resulting container is shown in Figure 8. Notice that this method, to allow the maximum peak power, may generate a container that is skewed in a way that imposes a tight constraint on current in certain locations of the die (such as at $i_2(t)$) while allowing larger current in other locations (such as at $i_1(t)$). Other approaches are possible to avoid this skew and even out the current budgets, as we will see next.

5.2. Uniform Current Distribution—The Sphere Approach

The design team may be interested in a grid that safely supports a uniform current distribution across the die to allow a placement that provides a uniform temperature distribution. We can generate constraints that allow that objective by searching for a safe maximal container $\mathcal{F}(u)$ that contains the hypersphere in current space that has the largest volume or the largest radius θ. In other words, this method aims to "raise the minimum" and avoid the skew indicated above.

For any $s \times 1$ or $1 \times s$ vector γ, we will use the standard notation $\|\gamma\| = \sqrt{\sum_{j=1}^{s} \gamma_j^2} \geq 0$ to denote the Euclidean norm of γ. Let $S(\theta) \subset \mathbb{R}^m$ denote the hypersphere with radius $\theta \geq 0$, centered at the origin, and let $S^+(\theta) = S(\theta) \cap \mathbb{R}^m_+$ be the part of that hypersphere that is in the first quadrant of $\mathbb{R}^m$, that is, $S^+(\theta) = \{I \geq 0 : \|I\| \leq \theta\}$. In the following lemma, we will derive a necessary and sufficient condition for $S^+(\theta) \subseteq \mathcal{F}(u)$, which will be useful in the rest of this section.

Let R^+ and R^- be the $n \times m$ matrices that consist of the non-negative elements and non-positive elements of R, respectively, so, with r_{ij}, r_{ij}^+, and r_{ij}^- denoting the (i, j)th entries of R, R^+, and R^-, respectively, we have for any (i, j)

$$r_{ij}^+ = \begin{cases} r_{ij}, & \text{if } r_{ij} \geq 0; \\ 0, & \text{otherwise.} \end{cases} \qquad r_{ij}^- = \begin{cases} r_{ij}, & \text{if } r_{ij} \leq 0; \\ 0, & \text{otherwise.} \end{cases} \tag{63}$$

Denote by r_i, r_i^+, and r_i^- the ith rows of R, R^+, and R^-, respectively. Let ν^+ and ν^- be $n \times 1$ vectors, where $\nu_i^+ = \|r_i^+\| \geq 0$ and $\nu_i^- = \|r_i^-\| \geq 0$, and let $\tilde{\nu} \triangleq \left[\begin{smallmatrix} \nu^+ \\ -\nu^- \end{smallmatrix}\right]$.

LEMMA 5.2. *For any $\theta \geq 0$ and $u \in \mathbb{R}^{2n}$ with $0 \in \mathcal{F}(u)$, $S^+(\theta) \subseteq \mathcal{F}(u)$ if and only if $\theta\tilde{\nu} \subseteq Tu$.*

The proof of Lemma 5.2 is available in Appendix D, as Lemma D.2.

For any $u \in \mathcal{U}$ with $0 \in \mathcal{F}(u)$, we define $\Theta(u)$ to be the largest $\theta \geq 0$ for which $S^+(\theta) \subseteq \mathcal{F}(u)$ or, equivalently, for which $\theta\tilde{\nu} \in Tu$ is satisfied, so

$$\Theta(u) \triangleq \max_{\substack{S^+(\theta) \subseteq \mathcal{F}(u) \\ \theta \geq 0}} (\theta) = \max_{\substack{\theta\tilde{\nu} \subseteq Tu \\ \theta \geq 0}} (\theta), \tag{64}$$

and we define Θ^* to be the largest $\Theta(u)$ achievable over all possible $u \in \mathcal{U}$ with $0 \in \mathcal{F}(u)$, that is,

$$\Theta^* \triangleq \max_{\substack{u \in \mathcal{U} \\ 0 \in \mathcal{F}(u)}} (\Theta(u)) = \max_{\substack{u \in \mathcal{U} \\ 0 \in Tu}} \big(\Theta(u)\big), \tag{65}$$

where the second equality is due to Lemma 4.2.

Let $u_s \in \mathcal{U}$ be a vector at which the above maximization attains its maximum. In other words, $u_s \in \mathcal{U}$ is such that $\Theta(u_s) = \Theta^*$ and $S^+(\Theta^*) \subseteq \mathcal{F}(u_s)$. In general, u_s may not be unique. Based on Equations (35) and (36), we can express the combined Equations (64) and (65) as the following linear program:

$$\begin{aligned} \Theta^* = \text{Maximize} \quad & \theta \\ \text{subject to} \quad & \theta\tilde{\nu} \subseteq Tu \\ & 0 \in Tu \\ & Pu \subseteq x_{th} \\ & \theta \geq 0. \end{aligned} \tag{66}$$

Notice that $0 \in \theta\tilde{\nu}$, for any $\theta \geq 0$, so the constraints $\theta \geq 0$ and $\theta\tilde{\nu} \subseteq Tu$ in Equation (66) automatically guarantee that $0 \in Tu$, and so the constraint $0 \in Tu$ in Equation (66) is redundant. Therefore, Equation (66) can be expressed as follows:

LP2:

$$\begin{aligned} \Theta^* = \text{Maximize} \quad & \theta \\ \text{subject to} \quad & \theta\tilde{\nu} \subseteq Tu \\ & Pu \subseteq x_{th} \\ & \theta \geq 0. \end{aligned} \tag{67}$$

Let $\mathcal{R}$ be the feasible region of the LP (67):

$$\mathcal{R} \triangleq \{(\theta, u) : \theta \geq 0, \theta\tilde{\nu} \subseteq Tu, Pu \subseteq x_{th}\}, \tag{68}$$

so from the above, we have

$$\Theta^* = \max_{(\theta,u) \in \mathcal{R}} (\theta). \tag{69}$$

Notice that $(0, 0) \in \mathcal{R}$, so $\mathcal{R}$ is not empty and both Θ^* and u_s are well defined. Therefore, the container $\mathcal{F}(u_s) = \{I \in \mathbb{R}^m : I \geq 0, RI \in Tu_s\} \neq \phi$ provides the desired current constraints, $\forall t \in \mathbb{R}$,

$$\begin{aligned} i(t) &\geq 0 \\ Ri(t) &\in Tu_s. \end{aligned}$$

Lemma D.3 in Appendix D establishes the maximality of $\mathcal{F}(u_s)$, based on Lemma D.1.

5.3. Combined Objective—The Sphere Approach

Thus far, we have presented two algorithms for current constraints generation. The first algorithm (50) aims to maximize the peak power dissipation that the grid can safely support in the underlying circuit; however, it generates a skewed container in a way that imposes a tight constraint on the currents in certain locations on the die. The second algorithm (67) aims to uniformly distribute the power budgets across the circuit by "raising the minimum," but this approach does not necessarily allow for a large peak total power dissipation. One may be interested in a middle scenario; a container that is maximal in $\mathcal{S}$ tries to maximize the peak power dissipation that the grid can safely support and tries to support a uniform current distribution across the die. In this section, we will develop a constraints generation algorithm, essentially a combination of Equations (52) and (69), that allows this type of design objective.

Recall that Equation (48) maximizes the sum of the m current sources attached to the grid, while Equation (64) maximizes the current radius for which the part of the hypersphere in the first quadrant is contained in $\mathcal{F}(u)$. Therefore, there is a clear disproportionality between the dimensions of both objective functions that motivates the following. For any $u \in \mathcal{U}$, we define $\xi(u)$ to be the largest value of the following combined objective allowed under $\mathcal{F}(u)$:

$$\xi(u) \triangleq \max_{\substack{I \in \mathcal{F}(u) \\ S^+(\theta) \subseteq \mathcal{F}(u)}} \left[\left(\sum_{j=1}^{m} I_j \right) + m\theta \right], \tag{70}$$

$$= \max_{\substack{RI \in Tu \\ \theta\tilde{\nu} \subseteq Tu \\ I, \theta \geq 0}} \left[\left(\sum_{j=1}^{m} I_j \right) + m\theta \right], \tag{71}$$

and we define ξ^* to be the largest $\xi(u)$ achievable under all possible $u \in \mathcal{U}$ with $0 \in \mathcal{F}(u)$, so

$$\xi^* \triangleq \max_{\substack{u \in \mathcal{U} \\ 0 \in \mathcal{F}(u)}} (\xi(u)) = \max_{\substack{u \in \mathcal{U} \\ 0 \in Tu}} (\xi(u)), \tag{72}$$

where the second equality is due to Lemma 4.2.

Let $u_c \in \mathcal{U}$ be a vector at which the above maximization attains its maximum. In other words, $u_c \in \mathcal{U}$ is such that $\xi(u_c) = \xi^*$. Also, let ζ and ω be such that $(\sum_{j=1}^{m} \zeta_j) + m\omega = \xi^*$, where $\zeta \in \mathcal{F}(u_c)$ and $\omega\tilde{\nu} \subseteq Tu_c$, where $\tilde{\nu}$ is as defined in Section 5.2. In general, u_c, ζ, and ω may not be unique. Based on Equations (35) and (36), we can express the combined Equations (71) and (72) as the following LP:

$$\begin{aligned} \xi^* = \text{ Maximize } & \left(\sum_{j=1}^{m} I_j\right) + m\theta \\ \text{subject to } & RI \in Tu \\ & \theta\tilde{\nu} \subseteq Tu \\ & 0 \in Tu \\ & Pu \subseteq x_{th} \\ & I, \theta \geq 0. \end{aligned} \tag{73}$$

Notice that $0 \in \theta\tilde{\nu}$, for any $\theta \geq 0$, so the constraints $\theta \geq 0$ and $\theta\tilde{\nu} \subseteq Tu$ in Equation (73) automatically guarantee that $0 \in Tu$, and so the constraint $0 \in Tu$ in Equation (73) is redundant. Therefore, Equation (73) can be expressed as follows:

$$\textbf{LP3:} \quad \begin{aligned} \xi^* = \text{ Maximize } & \left(\sum_{j=1}^{m} I_j\right) + m\theta \\ \text{subject to } & RI \in Tu \\ & \theta\tilde{\nu} \subseteq Tu \\ & Pu \subseteq x_{th} \\ & I, \theta \geq 0. \end{aligned} \tag{74}$$

Let $\mathcal{C}$ be the feasible region of the LP (74),

$$\mathcal{C} \triangleq \{\, (I, \theta, u) : \ \theta\tilde{\nu} \subseteq Tu, \ \ RI \in Tu, \ \ I, \theta \geq 0, \ \ \ Pu \subseteq x_{th} \,\}, \tag{75}$$

so from the above, we have

$$\xi^* = \max_{(I,\theta,u)\in\mathcal{C}} \left[\left(\sum_{j=1}^{m} I_j \right) + m\theta \right]. \tag{76}$$

Notice that $(0, 0, 0) \in \mathcal{C}$ so $\mathcal{C}$ is not empty, and all of ξ^*, u_c, ζ, and ω are well defined. Therefore, the container $\mathcal{F}(u_c) = \{I \in \mathbb{R}^m : I \geq 0, RI \in Tu_c\}$ provides the desired current constraints, $\forall t \in \mathbb{R}$:

$$\begin{aligned} i(t) &\geq 0 \\ Ri(t) &\in Tu_c. \end{aligned}$$

Lemma D.4 in Appendix D establishes the maximality of $\mathcal{F}(u_c)$, based on Lemma D.1.

6. IMPLEMENTATION

In this section, we discuss the implementation of one of the above algorithms, namely Equation (74), as the implementation of the rest of the algorithms is similar.

One way to construct the feasible region of Equation (74) is to compute $T = I_{2n} - \tilde{F}$. Recall from Definition 3.8 that $\tilde{F}$ requires knowledge of the non-negative and non-positive elements in F, where F is defined in Equation (12). Notice from Equation (12) that the explicit computation of F in turn requires the computation of $D^{-1}B$, which has two major drawbacks: (1) this requires the computation of all the columns of D^{-1} that is prohibitively expensive and (2) D^{-1} is a dense matrix, so F, $\tilde{F}$, and T are also dense, and the constraints of the LP (74) are dense. A key observation is that the explicit representation of T is unnecessary, and we can avoid the computation of D^{-1}, as we will show below, using the simple change of variables in Equations (77) and (78).

For any $u \in \mathbb{R}^{2n}$ and $I \in \mathbb{R}^m$, let $z \in \mathbb{R}^{2n_v}$ and $y \in \mathbb{R}^{n_v}$ be defined as follows:

$$Kz = \hat{B}u, \tag{77}$$

$$Dy = HI, \tag{78}$$

where

$$K = \begin{bmatrix} D & 0 \\ 0 & D \end{bmatrix}, \qquad \hat{B} = \begin{bmatrix} B & 0 & 0 & 0 \\ 0 & 0 & B & 0 \end{bmatrix}. \tag{79}$$

Recall that D is non-singular, so K is also non-singular, and its inverse is given by

$$K^{-1} = \begin{bmatrix} D^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix}. \tag{80}$$

Define the following matrices:

$$\hat{H} \triangleq \begin{bmatrix} I_{n_v} & 0 \\ 0 & 0 \\ 0 & I_{n_v} \\ 0 & 0 \end{bmatrix}, \quad H' \triangleq \begin{bmatrix} I_{n_v} \\ 0 \end{bmatrix}, \quad \hat{F} \triangleq \tilde{F} - \hat{H}K^{-1}\hat{B}, \quad \hat{R} \triangleq R - H'D^{-1}H, \tag{81}$$

where I_{n_v} is the $n_v \times n_v$ identity matrix, $\hat{H}$ is a $2n \times 2n_v$ matrix, and H' is an $n \times n_v$ matrix.

Notice that $Kz = \hat{B}u \iff z = K^{-1}\hat{B}u$. But $z = K^{-1}\hat{B}u \implies \hat{H}z = \hat{H}K^{-1}\hat{B}u$, and $\hat{H}z = \hat{H}K^{-1}\hat{B}u \implies z = K^{-1}\hat{B}u$, because $\hat{H}^T\hat{H} = I_{2n_v}$. Therefore,

$$Kz = \hat{B}u \iff \hat{H}z = \hat{H}K^{-1}\hat{B}u, \tag{82}$$
$$\iff \hat{F}u + \hat{H}z = (\hat{F} + \hat{H}K^{-1}\hat{B})u, \tag{83}$$
$$\iff \hat{F}u + \hat{H}z = \tilde{F}u. \tag{84}$$

Also, notice that $Dy = HI \iff y = D^{-1}HI$. But $y = D^{-1}HI \implies H'y = H'D^{-1}HI$, and $H'y = H'D^{-1}HI \implies y = D^{-1}HI$, because $H'^T H' = I_{n_v}$. Therefore,

$$Dy = HI \iff H'y = H'D^{-1}HI, \tag{85}$$
$$\iff \hat{R}I + H'y = (\hat{R} + H'D^{-1}H)I, \tag{86}$$
$$\iff \hat{R}I + H'y = RI. \tag{87}$$

Therefore, for any $u \in \mathbb{R}^{2n}$, $I \in \mathbb{R}^m$, and $\theta \in \mathbb{R}$, we have

$$\left.\begin{array}{l} RI \in Tu \\ \theta\tilde{\nu} \subseteq Tu \end{array}\right\} \iff \left\{\begin{array}{l} RI \in (I_{2n} - \tilde{F})u \\ \theta\tilde{\nu} \subseteq (I_{2n} - \tilde{F})u \end{array}\right., \tag{88}$$

$$\iff \left\{\begin{array}{l} \hat{R}I + H'y \in (I_{2n} - \hat{F})u - \hat{H}z \\ \theta\tilde{\nu} \subseteq (I_{2n} - \hat{F})u - \hat{H}z \\ Kz = \hat{B}u \\ Dy = HI \end{array}\right., \tag{89}$$

where in Equation (88) we have used the fact that $T = I_{2n} - \tilde{F}$ and in Equation (89) we have used Equations (84) and (87). With this, LP3 in Equation (74) can be expressed as follows:

$$\textbf{LP3':} \quad \begin{array}{rll} \xi^* = & \text{Maximize} & \left(\sum_{j=1}^{m} I_j\right) + m\theta \\ & \text{subject to} & \hat{R}I + H'y \in (I_{2n} - \hat{F})u - \hat{H}z \\ & & \theta\tilde{\nu} \subseteq (I_{2n} - \hat{F})u - \hat{H}z \\ & & Kz = \hat{B}u \\ & & Dy = HI \\ & & Pu \subseteq x_{th} \\ & & I, \theta \geq 0. \end{array} \tag{90}$$

Notice that K and $\hat{B}$ are sparse matrices that can be constructed easily from the matrices D and B. Furthermore, notice that constructing the matrices $\hat{R}$ and $\hat{F}$ requires the computation of $D^{-1}M$, the computation of the inverse of the diagonal matrix E, and some matrix multiplications. The computation of $D^{-1}M$ does not require the full inverse of D; it only requires an LU factorization of the matrix D and n_l forward/backward solves. The result of $D^{-1}M$ is used to compute the matrices $E^{-1}M^T D^{-1}B$ and $E^{-1}M^T D^{-1}H$ in $\hat{F}$ and $\hat{R}$, respectively. With this, it is easy to see that the constraints of Equation (90) do not require the full inverse of D thus, it is easier to construct as compared to Equation (74), and the constraints of Equation (90) are much sparser than Equation (74).

Finally, notice that the computation of $\tilde{\nu}$ requires the computation of the rows of R. More precisely, finding the upper part of $\tilde{\nu}$, denoted as $\tilde{\nu}_u$, requires the elements of $D^{-1}H$, which can be done using $m \ll n_v$ linear system solves. Notice that the full $D^{-1}H$ does not have to be stored in memory. Finding the lower part of $\tilde{\nu}$, denoted as

Table I. Comparison of the Three Approaches

Power Grid	Peak Power		Uniform Current Distribution		Combined Objective	
Name	$P(u_p)$ in mW	$\Theta(u_p)$ in μA	$P(u_s)$ in mW	$\Theta(u_s)$ in μA	$P(u_c)$ in mW	$\Theta(u_c)$ in μA
G1	0.48	0.98	0.11	2.88	0.43	2.58
G2	0.96	1.77	0.31	3.58	0.85	3.42
G3	1.45	1.04	0.45	2.86	1.31	2.74
G4	3.09	1.22	1.15	3.22	2.72	3.12
G5	5.77	1.43	2.35	3.43	5.05	3.36
G6	8.55	1.58	3.61	3.56	7.42	3.50
G7	18.69	1.28	8.38	3.29	16.75	3.25
G8	33.52	1.11	15.59	3.11	30.86	3.06
G9	70.04	0.75	33.95	2.73	66.21	2.71
G10	97.24	1.57	44.54	3.57	83.96	3.53
G11	118.53	1.58	54.86	3.55	103.13	3.54

$\tilde{\nu}_l$, requires the computation of $-E^{-1}M^T D^{-1}H$. Recall that $D^{-1}M$ is already computed and stored in memory to construct $\hat{F}$, so that $-E^{-1}M^T D^{-1}H = -E^{-1}(D^{-1}M)^T H$, because D^{-1} is symmetric, can be easily computed using matrix transpose and matrix multiplications.

7. RESULTS

To compare the different tradeoffs of the current containers generated by each of the above three algorithms and their runtime efficiency, we implemented LP1, LP2, and LP3 given in Equations (50), (67), and (74), respectively, using C++. Recall that these algorithms were transformed into simpler forms by avoiding the computation of the full D^{-1} matrix. The implementation details of LP1 and LP2 are similar to those of LP3, which is explained in Section 6. We tested them on a number of power grids with a 1.1V supply voltage that was generated based on user specifications, including grid dimensions, metal layers, pitch and width per layer, and C4 pads and current source distributions, consistent with 65nm technology. With these specifications, the grids are automatically generated, after which we introduce non-uniformity in the grid to model the real-world scenario. The maximizations were performed using the Mosek optimization package [MOSEK ApS 2015]. All results were obtained using a 3.4GHz Linux machine with 32GB of RAM.

The number of variables and constraints required for each LP are shown in Table III. Furthermore, the specifications of the generated power grids are given in columns 1–4 of Table II. Also, the total CPU runtime for setting up and solving LP1, LP2, and LP3 are given in columns 5–7 of Table II. Note that the CPU time in columns 6 and 7 of Table II include the time required for computing $\tilde{\nu}$ and $\tilde{\eta}$, respectively. For example, on a 600k-node grid, LP1 took 5.8h, LP2 took 9.6h, and LP3 took 10.8h.

In Table I, we present the resultsof the three LPs. Denote by $P(u) \triangleq V_{dd} \times \sigma(u)$ the peak power dissipation allowed under $\mathcal{F}(u)$. To study the difference between the containers generated using LP1, LP2, and LP3, we used the following method. First, we computed the peak power dissipation achievable under all containers, which are $P(u_p)$, $P(u_s)$, and $P(u_c)$, and the largest current radius for which the part of the hypersphere in the first quadrant is contained in all containers, which are $\Theta(u_p)$, $\Theta(u_s)$, and $\Theta(u_c)$. For instance, on a 560k-node grid, the peak power dissipation achievable under $\mathcal{F}(u_p)$, $\mathcal{F}(u_s)$, and $\mathcal{F}(u_c)$ is 97.24mW, 44.54mW, and 83.96mW, respectively, and the largest current radius for which the part of the hypersphere in the first quadrant is contained in $\mathcal{F}(u_p)$, $\mathcal{F}(u_s)$, and $\mathcal{F}(u_c)$ is 1.57μA, 3.57μA, and 3.53μA, respectively. The results show that $P(u_s) \ll P(u_p)$ and $\Theta(u_p) \ll \Theta(u_s)$ on all grids. Thus, both $\mathcal{F}(u_p)$ and $\mathcal{F}(u_s)$ provide

Table II. Runtime of the Three Approaches

Power Grid				Peak Power	Uniform Current Distribution	Combined Objective
Name	Nodes	Current Sources	C4 Connections	Total Time	Total Time	Total Time
G1	2,027	156	27	1.7s	1.3s	2.3s
G2	4,499	306	71	5.2s	5.0s	5.8s
G3	7,774	552	97	10.2s	11.5s	11.9s
G4	17,160	1,190	199	31.0s	33.6s	36.5s
G5	30,117	2,070	360	1.3min	1.2min	1.6min
G6	46,701	3,192	556	2.6min	3.1min	3.7min
G7	104,530	7,140	1,164	14.8min	16.0min	14.3min
G8	184,523	12,656	1,994	21.7min	1.0h	48.8min
G9	412,927	28,056	4,375	1.6h	6.3h	4.2h
G10	561,344	38,220	6,027	3.0h	8.6h	6.0h
G11	662,708	45,156	7,119	5.8h	9.6h	10.8h

Table III. Number of Variables and Constraints for All Three LPs

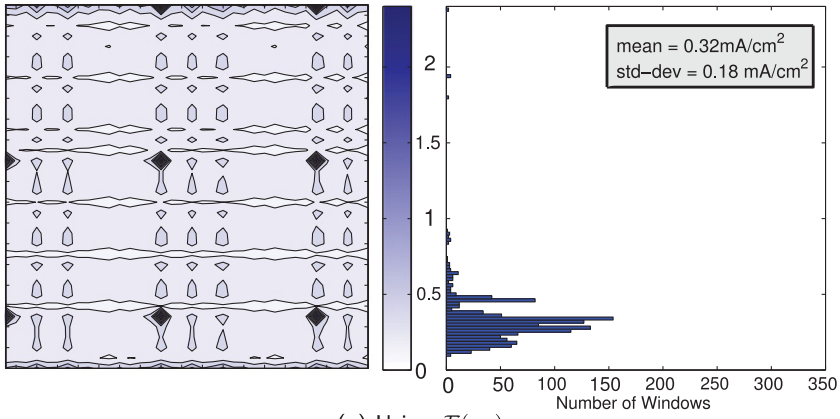
	LP1 (50)	LP2 (67)	LP3 (74)
Number of Variables	5N+2n+m	4N+2n+1	4N+2n+m+1
Number of Constraints	6N+2n	5N+2n	6N+2n

a distinct tradeoff for the chip design team. Moreover, the results show that $P(u_c) \approx P(u_p)$ and $\Theta(u_c) \approx \Theta(u_s)$. Therefore, the combined objective approach in LP3 gives the best features of the peak power dissipation and the uniform current distribution approaches.

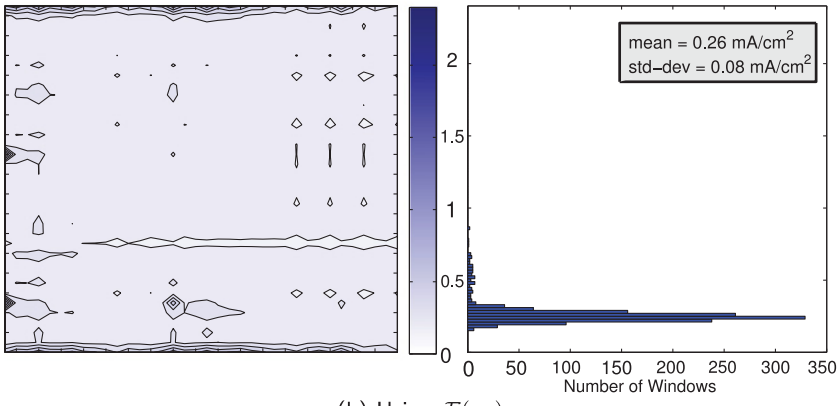
Another way to compare the three approaches, LP1, LP2, and LP3, is to look at the *power density*, that is, the power dissipation per unit area of the die, allowed by the three resulting containers. To assess this, we maximize the allowed power (current) within a small window of the die surface, and we do this for every position of that window across the die. We divide the die area into $\kappa \times \kappa$ of these windows and compute the peak power dissipation inside each, as allowed by $\mathcal{F}(u_p)$, $\mathcal{F}(u_s)$, and $\mathcal{F}(u_c)$. In Figure 9, we present contour plots for $\kappa = 35$ for the peak power densities under $\mathcal{F}(u_p)$, $\mathcal{F}(u_s)$, and $\mathcal{F}(u_c)$, respectively, on a 100k-node grid. Note that the current constraints based on $\mathcal{F}(u_p)$ allow higher current densities at certain spots but also include some spots with very small and restricted current density budgets. This large spread in power densities can lead to thermal hotspots. This may be avoided by using $\mathcal{F}(u_s)$, which, as expected and as seen in the figure, provides a uniform distribution of power densities across the die area compared to $\mathcal{F}(u_p)$, which is reflected in a smaller standard deviation. Of course, $\mathcal{F}(u_p)$ supports larger overall peak power dissipation than $\mathcal{F}(u_s)$, which is reflected in a larger mean. The current constraints based on $\mathcal{F}(u_c)$ provide a power density distribution over a smaller range compared to $\mathcal{F}(u_p)$ and allow for larger power dissipation compared to $\mathcal{F}(u_s)$. Clearly, $\mathcal{F}(u_c)$ is superior to $\mathcal{F}(u_p)$ and $\mathcal{F}(u_s)$, providing the best features in those containers.

8. CONCLUSION

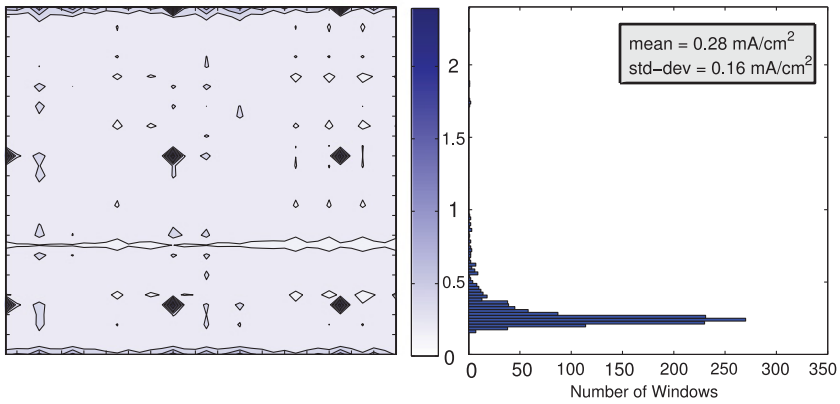
Efficient and early power grid verification is a key step in modern chip design. This has been extensively addressed in the recent literature, introducing novel simulation-based and vectorless approaches, both of which have their shortcomings. In this article, we adopt a recently introduced framework, namely the *inverse* problem of vectorless verification, that generates circuit current constraints to guarantee power grid safety. We extend the applicability of this framework to allow for inductance. We develop some key theoretical results to allow the generation of constraints that correspond to maximal current spaces. We then apply these results to provide two constraints



(a) Using $\mathcal{F}(u_p)$



(b) Using $\mathcal{F}(u_s)$



(c) Using $\mathcal{F}(u_c)$

Fig. 9. Contour plots for peak power density across the layout and the corresponding histograms. The color bar units are mA/cm^2.

generation algorithms that target key quality metrics of the grid: maximum power dissipation and uniformity of the power spread across the die. Finally, we present a combination of both quality metrics that proved to be superior to the other algorithms.

APPENDIXES

APPENDIX A: SP MATRICES

LEMMA A.1. *Let X be a $2n \times 2n$ matrix represented as:*

$$X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}, \tag{91}$$

where X_{11}, X_{12}, X_{21}, and X_{22} are $n \times n$ matrices. X is SP if and only if

$$X_{11} \geq 0\,, \quad X_{12} \leq 0\,, \quad X_{21} \leq 0\,, \text{ and } \quad X_{22} \geq 0. \tag{92}$$

PROOF. The proof is in two parts.

Proof of the "if direction": Let X be a $2n \times 2n$ matrix that satisfies Equation (92). Let $u = \left[\begin{smallmatrix} u_t \\ u_b \end{smallmatrix}\right]$ and $v = \left[\begin{smallmatrix} v_t \\ v_b \end{smallmatrix}\right]$ be any two $2n \times 1$ vectors, where u_t, u_b, v_t, and v_b are $n \times 1$, and let $u \subseteq v$. Because $u_t \leq v_t$, then $X_{11}u_t \leq X_{11}v_t$, and because $u_b \geq v_b$, then $X_{12}u_b \leq X_{12}v_b$, which gives $X_{11}u_t + X_{12}u_b \leq X_{11}v_t + X_{12}v_b$. Likewise, because $u_t \leq v_t$, then $X_{21}u_t \geq X_{21}v_t$, and because $u_b \geq v_b$, then $X_{22}u_b \geq X_{22}v_b$, which gives $X_{21}u_t + X_{22}u_b \geq X_{21}v_t + X_{22}v_b$, so $Xu \subseteq Xv$ and X is SP.

Proof of the "only if direction": Let X be SP, and let $u = \left[\begin{smallmatrix} u_t \\ u_b \end{smallmatrix}\right]$ and $v = \left[\begin{smallmatrix} v_t \\ v_b \end{smallmatrix}\right]$ be any two $2n \times 1$ vectors, where u_t, u_b, v_t, and v_b are $n \times 1$ vectors, and let $u \subseteq v$ or, equivalently,

$$u_t \leq v_t \quad \text{and} \quad u_b \geq v_b. \tag{93}$$

Because $Xu \subseteq Xv$, we have that

$$X_{11}u_t + X_{12}u_b \ \leq X_{11}v_t + X_{12}v_b, \tag{94}$$

$$X_{21}u_t + X_{22}u_b \ \geq X_{21}v_t + X_{22}v_b. \tag{95}$$

Let $u_t = u_b = 0$, $v_b = 0$, and $v_t = e_k$, where $e_k \in \mathbb{R}^n$ is the vector whose kth entry is 1 and all other entries are 0. Because this assignment satisfies Equation (93), then Equations (94) and (95) lead to $X_{11}e_k \geq 0$ and $X_{21}e_k \leq 0$, for every $k \in \{1, \ldots, n\}$. This means that $X_{11} \geq 0$ and $X_{21} \leq 0$. Likewise, let $u_t = u_b = 0$, $v_t = 0$, and $v_b = -e_k$. Because this assignment satisfies Equation (93), then Equations (94) and (95) lead to $-X_{12}e_k \geq 0$ and $-X_{22}e_k \leq 0$ for every $k \in \{1, \ldots, n\}$. This means that $X_{12} \leq 0$ and $X_{22} \geq 0$, which completes the proof. □

LEMMA A.2. *If X and Y are SP, then XY and $(X+Y)$ are SP.*

PROOF. Suppose that X and Y are $2n \times 2n$ SP matrices. For any two $2n \times 1$ vectors $u \subseteq v$, we have $Yu \subseteq Yv$, and $X(Yu) \subseteq X(Yv)$, so $XYu \subseteq XYv$ and XY is SP. Furthermore, $Xu \subseteq Xv$ and $Yu \subseteq Yv$, so $(X+Y)u = Xu + Yu \subseteq Xv + Yv = (X+Y)v$ and so $(X+Y)$ is SP. □

APPENDIX B: PROOF OF LEMMA 4.8

LEMMA B.1 ([JOSHI 2001]). *Let $\mathcal{G} = \{y \in \mathbb{R}^m : Zy \leq w\}$ be a non-empty convex polytope, where Z is an $r \times m$ matrix and w is an $r \times 1$ vector. Also, let z_i and w_i be the ith row of Z and the ith element of w, respectively. Then there exists a $y \in \mathcal{G}$ such that $z_i y = w_i$ for some $i \in \{1, \ldots, r\}$.*

LEMMA B.2. *If $\mathcal{F}(u)$ is maximal in $\mathcal{S}$, then u is feasible and extremal in $\mathcal{U}$.*

PROOF. We will prove the contrapositive. Let $u \in \mathcal{U}$ be either infeasible or not extremal in $\mathcal{U}$; we will prove that $\mathcal{F}(u)$ is not maximal in $\mathcal{S}$. If u is infeasible, then $\mathcal{F}(u) = \phi$, which we already know is not maximal in $\mathcal{S}$. Now consider the case when u is feasible but not extremal in $\mathcal{U}$. In other words, we have $Pu \subset x_{th}$, so $\epsilon \triangleq \min_{\forall i}(|Pu|_i - x_{th,i}|) > 0$. Let $\mathbb{1}_{2d}$ be the $2d \times 1$ vector whose first d entries are 1 and the rest are -1, so $Pu \subseteq x_{th} - \epsilon\mathbb{1}_{2d}$. Also, let $\mathbb{1}_{2n}$ be the $2n \times 1$ vector whose first n entries are 1 and the rest are -1. Because P has exactly one 1 in each row, it follows that $P\mathbb{1}_{2n} = \mathbb{1}_{2d}$. Also, let $q = Q\mathbb{1}_{2n}$, so $\delta \triangleq \max_{\forall i} |q_i| > 0$ because Q is non-singular, and let $u' = u + (\epsilon/\delta)q$. Notice that $(1/\delta)q \subseteq \mathbb{1}_{2n}$, due to the definition of δ, so $(\epsilon/\delta)Pq \subseteq \epsilon P\mathbb{1}_{2n}$ because ϵP is SP, from which $Pu + (\epsilon/\delta)Pq \subseteq x_{th} - \epsilon\mathbb{1}_{2d} + \epsilon P\mathbb{1}_{2n} = x_{th}$, due to $P\mathbb{1}_{2n} = \mathbb{1}_{2d}$. Therefore, we have $Pu' \subseteq x_{th}$, so $u' \in \mathcal{U}$. Also, notice that $Tu' = Tu + (\epsilon/\delta)Tq = Tu + (\epsilon/\delta)TQ\mathbb{1}_{2n} = Tu + (\epsilon/\delta)\mathbb{1}_{2n}$, so $Tu \subset Tu'$, because $(\epsilon/\delta) > 0$. We have so far established that there exists $u' \in \mathcal{U}$ with $Tu \subset Tu'$, so $\mathcal{F}(u) \subseteq \mathcal{F}(u')$, due to Equation (37). It only remains to prove that $\mathcal{F}(u) \neq \mathcal{F}(u')$. Notice that $\mathcal{F}(u') \neq \phi$, because u is feasible and $\mathcal{F}(u) \subseteq \mathcal{F}(u')$. Also, for any $y, z \in \mathcal{F}(u')$ and $0 \leq \alpha \leq 1$, we have $\alpha y + (1-\alpha)z \geq 0$ and $R[\alpha y + (1-\alpha)z] = \alpha Ry + (1-\alpha)Rz \in \alpha Tu' + (1-\alpha)Tu' = Tu'$, so $\mathcal{F}(u')$ is convex. Therefore, due to Lemma B.1, there exists an $I \in \mathcal{F}(u')$ such that:

$$r_i I = t_i u' \qquad \text{or} \qquad r_i I = t_{n+i} u' \tag{96}$$

for some $i \in \{1, \ldots, n\}$, where r_i is the ith row of R, t_i is the ith row of T, and t_{n+i} is the $(n+i)$th row of T. Suppose, towards a contradiction, that $I \in \mathcal{F}(u)$, from which

$$r_i I \leq t_i u \qquad \text{and} \qquad r_i I \geq t_{n+i} u, \qquad \forall i \in \{1, \ldots, n\}. \tag{97}$$

Therefore, due to Equations (96) and (97), we have

$$t_i u' \leq t_i u \qquad \text{or} \qquad t_{n+i} u' \geq t_{n+i} u, \tag{98}$$

that contradicts $Tu \subset Tu'$, so $I \notin \mathcal{F}(u)$, $\mathcal{F}(u) \neq \mathcal{F}(u')$, $\mathcal{F}(u)$ is not maximal in $\mathcal{S}$, and the proof is complete. □

APPENDIX C: PROOF OF LEMMAS 4.10–4.12

LEMMA C.1. *For any feasible $u \in \mathbb{R}^{2n}$ and any $z \in \mathbb{R}^{2n}$ such that $0 \subseteq Tz \subseteq T(u - \overline{x}(\mathcal{F}(u)))$, let $u' = u - z$, it follows that $\mathcal{F}(u') = \mathcal{F}(u)$.*

PROOF. For any $I \in \mathcal{F}(u')$, we have $I \geq 0$ and $RI \in Tu' = Tu - Tz \subseteq Tu$, due to Equation (25) and $0 \subseteq Tz$, so $I \in \mathcal{F}(u)$. It follows that $\mathcal{F}(u') \subseteq \mathcal{F}(u)$. Conversely, for any $I \in \mathcal{F}(u)$, we have $I \geq 0$ and

$$RI \in \operatorname*{eopt}_{I \in \mathcal{F}(u)}(RI) = T\,\overline{x}(\mathcal{F}(u)). \tag{99}$$

Notice that for any z with $0 \subseteq Tz \subseteq T(u - \overline{x}(\mathcal{F}(u)))$, we have $Tu' = Tu - Tz \supseteq Tu - T(u - \overline{x}(\mathcal{F}(u)))$, due to Equations (23) and (21), so $Tu' \supseteq T\,\overline{x}(\mathcal{F}(u))$. Combining this with Equation (99), we get $RI \in Tu'$, so $I \in \mathcal{F}(u')$. Therefore, $\mathcal{F}(u) \subseteq \mathcal{F}(u')$ from which $\mathcal{F}(u') = \mathcal{F}(u)$, and the proof is complete. □

LEMMA C.2. *For any feasible $u \in \mathbb{R}^{2n}$, let $u' = \overline{x}(\mathcal{F}(u))$, it follows that $\mathcal{F}(u') = \mathcal{F}(u)$.*

PROOF. Let $z = u - \overline{x}(\mathcal{F}(u))$, so $Tz = Tu - T\overline{x}(\mathcal{F}(u)) = Tu - \operatorname{eopt}_{I \in \mathcal{F}(u)}(RI) \supseteq 0$, the last step due to the definition of $\mathcal{F}(u)$ and Equation (25). As a result, z satisfies the conditions of Lemma C.1. Let $u' = u - z = \overline{x}(\mathcal{F}(u))$. Then, by Lemma C.1, $\mathcal{F}(u') = \mathcal{F}(u)$. □

LEMMA C.3. *For any $u \in \mathbb{R}^{2n}$, u is irreducible if and only if it is feasible and $\overline{x}(\mathcal{F}(u)) = u$.*

PROOF. The proof is in two parts.

Proof of the "if direction": The proof is by contradiction. Let u be feasible with $\overline{x}(\mathcal{F}(u)) = u$, and suppose that u is reducible so there exists $u' \subseteq u$, $u' \neq u$, with $\mathcal{F}(u') = \mathcal{F}(u)$. Notice that $\mathcal{F}(u)$ is not empty, because u is feasible, so $\mathcal{F}(u')$ is not empty and u' is feasible. Therefore, we get

$$u' - \overline{x}(\mathcal{F}(u')) = u' - \overline{x}(\mathcal{F}(u)) = u' - u + u - \overline{x}(\mathcal{F}(u)).$$

Because $\overline{x}(\mathcal{F}(u')) \subseteq u'$, due to Lemma 4.4, it follows that $u' - u + u - \overline{x}(\mathcal{F}(u)) \supseteq 0$, due to Equation (25), so $u - \overline{x}(\mathcal{F}(u)) \supseteq u - u' \supseteq 0$, the final step due to $u' \subseteq u$ and Equation (25). But $u - \overline{x}(\mathcal{F}(u)) = 0$ due to $\overline{x}(\mathcal{F}(u)) = u$, so $u' = u$, and we have a contradiction that completes the proof.

Proof of the "only if direction": We will prove the contrapositive. Let u be either infeasible or $\overline{x}(\mathcal{F}(u)) \neq u$, and we will prove that u is reducible. If u is infeasible, then $\mathcal{F}(u) = \phi$ and $u \neq 0$ (recall that $u = 0$ is always feasible), and it is easy to find another infeasible u' with $u' \subseteq u$ and $u' \neq u$, as follows. Let $u' = \frac{1}{2}u$, from which $Tu' = \frac{1}{2}Tu$. Suppose that there exists $I \in \mathcal{F}(u')$, that is, $\exists I \geq 0$ such that $RI \in \frac{1}{2}Tu$, then $2I \geq 0$ and $R(2I) \in Tu$, so$2I \in \mathcal{F}(u)$ that contradicts that u is infeasible; it follows that u' is infeasible. Therefore, we have found $u' \subseteq u$, $u' \neq u$, with $\mathcal{F}(u') = \mathcal{F}(u) = \phi$, which means that u is reducible. If u is feasible and $\overline{x}(\mathcal{F}(u)) \neq u$, then let $u' = \overline{x}(\mathcal{F}(u))$, so $\overline{x}(\mathcal{F}(u)) \subseteq u$, due to Lemma 4.4, leads to $u' \subseteq u$, $u' \neq u$, with $\mathcal{F}(u') = \mathcal{F}(u)$ due to Lemma 4.10, and u is reducible. □

LEMMA C.4. *For any feasible $u \in \mathbb{R}^{2n}$, let $u' = \overline{x}(\mathcal{F}(u))$, it follows that u' is irreducible.*

PROOF. Because $u' = \overline{x}(\mathcal{F}(u))$, it follows from Lemma 4.10 that $\mathcal{F}(u') = \mathcal{F}(u)$, so u' is feasible and $\overline{x}(\mathcal{F}(u')) = \overline{x}(\mathcal{F}(u))$. With this, notice that $u' - \overline{x}(\mathcal{F}(u')) = u' - \overline{x}(\mathcal{F}(u)) = 0$, from which $\overline{x}(\mathcal{F}(u')) = u'$. Using Lemma 4.11, it follows that u' is irreducible, and the proof is complete. □

LEMMA C.5. *For any $u \in \mathbb{R}^{2n}$, u is irreducible if and only if:*

$$Tu \subseteq Tu' \iff \mathcal{F}(u) \subseteq \mathcal{F}(u'), \quad \forall u' \in \mathbb{R}^{2n}. \tag{100}$$

PROOF. The proof is in two parts.

Proof of the "if direction": We give a proof by contradiction, given Equation (100) and supposing u is reducible, so it is either infeasible or $\overline{x}(\mathcal{F}(u)) \neq u$. If u is infeasible, then $\mathcal{F}(u) = \phi \subseteq \mathcal{F}(u')$, for any $u' \in \mathbb{R}^{2n}$, so $Tu \subseteq Tu'$, for any $u' \in \mathbb{R}^{2n}$, due to Equation (100). But this is impossible, because we can always find a $u' \in \mathbb{R}^{2n}$ that violates $Tu \subseteq Tu'$, as follows. Let $\mathbb{1}_{2n}$ be the $2n \times 1$ vector whose first n entries are 1 and the rest are -1, and let $w = Q\mathbb{1}_{2n}$ so $Tw = \mathbb{1}_{2n} \geq 0$, and let $u' = u - w$ so $Tu - Tu' = Tw \geq 0$, and, hence, $Tu' \subseteq Tu$, due to Equation (25), with $Tu' \neq Tu$, because $Tw = \mathbb{1}_{2n} \neq 0$. This violates $Tu \subseteq Tu'$. Therefore, it must be that u is feasible and $\overline{x}(\mathcal{F}(u)) \neq u$. Let $u' = \overline{x}(\mathcal{F}(u))$, so $\mathcal{F}(u') = \mathcal{F}(u)$ due to Lemma 4.10, with $Tu' = T\,\overline{x}(\mathcal{F}(u))$. Recall that $T\,\overline{x}(\mathcal{F}(u)) = \text{eopt}_{I \in \mathcal{F}(u)}(RI) \subseteq Tu$, and $T\,\overline{x}(\mathcal{F}(u)) \neq Tu$ due to $\overline{x}(\mathcal{F}(u)) \neq u$, so $Tu' \subseteq Tu$, $Tu' \neq Tu$. This means that we have $\mathcal{F}(u) \subseteq \mathcal{F}(u')$ while $Tu \not\subseteq Tu'$, which contradicts (100), and the proof is complete.

Proof of the "only if direction": Let u be irreducible, so u is feasible with $\overline{x}(\mathcal{F}(u)) = u$. Due to Equation (37), it only remains to prove that $\forall u' \in \mathbb{R}^{2n}$, $\mathcal{F}(u) \subseteq \mathcal{F}(u') \implies Tu \subseteq Tu'$. Notice that $\mathcal{F}(u')$ is non-empty, because $\mathcal{F}(u) \neq \phi$ and $\mathcal{F}(u) \subseteq \mathcal{F}(u')$, from which u'

is feasible. Because u and u' are feasible, and using $u = \overline{x}(\mathcal{F}(u))$, notice that

$$\begin{aligned} Tu' - Tu &= Tu' - T\,\overline{x}(\mathcal{F}(u)) \\ &= Tu' - \operatorname*{eopt}_{I\in\mathcal{F}(u)}(RI) \\ &\supseteq Tu' - \operatorname*{eopt}_{I\in\mathcal{F}(u')}(RI) \supseteq 0, \end{aligned}$$

where we used $\text{eopt}_{I\in\mathcal{F}(u')}(RI) \supseteq \text{eopt}_{I\in\mathcal{F}(u)}(RI)$, because $\mathcal{F}(u) \subseteq \mathcal{F}(u')$, making use of Equations (23), (21), and (25). Therefore, $Tu' - Tu \supseteq 0$, so $Tu \subseteq Tu'$ due to Equation (25) and the proof is complete. □

APPENDIX D: APPLICATIONS

LEMMA D.1. *Given a real-valued function $g(\cdot) : \mathbb{R}^{2n} \to \mathbb{R}$ such that, for any $u, u' \in \mathcal{U}$, with $0 \in Tu$ and $0 \in Tu'$, we have (i) $g(u') = g(u)$ if $\mathcal{F}(u') = \mathcal{F}(u)$ and (ii) $g(u') > g(u)$ if $Tu' \supset Tu$. Furthermore, let*

$$g^* \triangleq \max_{\substack{u\in\mathcal{U}\\ 0\in Tu}}[g(u)], \tag{101}$$

and let $u^* \in \mathcal{U}$ be feasible with $0 \in Tu^*$ and $g(u^*) = g^*$. It follows that $\mathcal{F}(u^*)$ is maximal in $\mathcal{S}$.

PROOF. We will prove that u^* is irreducible and extremal in $\mathcal{U}$, so$\mathcal{F}(u^*)$ is maximal in $\mathcal{S}$, due to Theorem 4.13. The proof is in two parts.

First, we will prove that u^* is extremal in $\mathcal{U}$; the proof is *by contradiction*. Let $u \in \mathcal{U}$ be feasible with $0 \in Tu$ and $g(u) = g^*$, and suppose that u is not extremal in $\mathcal{U}$, so $Pu \subset x_{th}$. Let $\epsilon \triangleq \min_{\forall i}(|Pu|_i - x_{th,i}|) > 0$, and let $\mathbb{1}_{2n}$ be the $2n\times 1$ vector whose first n entries are 1 and the rest are -1. Because P has exactly one 1 in each row, it follows that $P\mathbb{1}_{2n} = \mathbb{1}_{2d}$, so $Pu \subseteq x_{th} - \epsilon\mathbb{1}_{2d}$, due to the definition of ϵ. Also, let $q = Q\mathbb{1}_{2n}$, so $\delta \triangleq \max_{\forall i}|q_i| > 0$, because Q is non-singular, and let $u' = u + (\epsilon/\delta)q$, for which, clearly, $u' \neq u$. Notice that $(1/\delta)q \subseteq \mathbb{1}_{2n}$, due to the definition of δ, so $(\epsilon/\delta)Pq \subseteq \epsilon P\mathbb{1}_{2n}$, because ϵP is SP, from which $Pu + (\epsilon/\delta)Pq \subseteq x_{th} - \epsilon\mathbb{1}_{2d} + \epsilon P\mathbb{1}_{2n} = x_{th}$, due to $P\mathbb{1}_{2n} = \mathbb{1}_{2d}$. Therefore, we have $Pu' \subseteq x_{th}$, so $u' \in \mathcal{U}$. Note that $Tu' = Tu + (\epsilon/\delta)Tq = Tu + (\epsilon/\delta)TQ\mathbb{1}_{2n} = Tu + (\epsilon/\delta)\mathbb{1}_{2n}$, and, because $(\epsilon/\delta) > 0$, we get $Tu' \supset Tu$, so $0 \in Tu'$, due to $0 \in Tu$. It follows that $g(u') > g(u) = g^*$ with $u' \neq u$, which contradicts Equation (101). Therefore, u is extremal in $\mathcal{U}$, so u^* is extremal in $\mathcal{U}$, which completes the first part of the proof.

Next, we will prove that u^* is irreducible; the proof is *by contradiction*. Let $u \in \mathcal{U}$ be feasible with $0 \in Tu$ and $g(u) = g^*$, and suppose that u is reducible; then, by Lemma 4.11, we must have $\overline{x}(\mathcal{F}(u)) \neq u$. Let $u' = \overline{x}(\mathcal{F}(u))$, so $\mathcal{F}(u') = \mathcal{F}(u)$ due to Lemma 4.10. Because $u' \subseteq u$ due to Lemma 4.4, from which $Pu' \subseteq Pu$ because P is SP, then $u' \in \mathcal{U}$. Note that $Tu' = T\,\overline{x}(\mathcal{F}(u)) = \text{eopt}_{I\in\mathcal{F}(u)}(RI)$. Furthermore, because $0 \in Tu$, we have $0 \in \mathcal{F}(u)$, due to Lemma 4.2, so $0 \in \text{eopt}_{I\in\mathcal{F}(u)}(RI)$ due to Equation (36), from which $0 \in Tu'$, and the conditions of the lemma provide that $g(u') = g(u) = g^*$. Let $\delta = Tu - Tu'$. Note that $Tu' = T\,\overline{x}(\mathcal{F}(u)) = \text{eopt}_{I\in\mathcal{F}(u)}(RI) \subseteq Tu$, due to Equation (36), and $Tu \neq T\,\overline{x}(\mathcal{F}(u))$, due to $\overline{x}(\mathcal{F}(u)) \neq u$, from which $\delta \supseteq 0$ and $\delta \neq 0$. Combining this with Q being SP, from Lemma E.8, and every element of Q is non-zero, from Lemma E.6, so we have $0 \subset Q\delta = u - u'$. Consequently, we have $u' \subset u$, due to Equation (26), so $Pu' \subset Pu \subseteq x_{th}$, making use of Lemma 3.11 and the final step due to $u \in \mathcal{U}$, so u' is not extremal in $\mathcal{U}$. But this contradicts the first part of the proof. It follows that u is irreducible, so u^* is irreducible. Therefore, $\mathcal{F}(u^*)$ is maximal in $\mathcal{S}$. □

LEMMA D.2. *For any $\theta \geq 0$ and $u \in \mathbb{R}^{2n}$ with $0 \in \mathcal{F}(u)$, $S^+(\theta) \subseteq \mathcal{F}(u)$ if and only if $\theta\bar{v} \subseteq Tu$.*

PROOF. Let t_k be the kth row of T. Because $0 \in \mathcal{F}(u)$, it follows that $t_i u \geq 0$ and $t_{(n+i)} u \leq 0$, $\forall i \in \{1, \ldots, n\}$, due to Lemma 4.2. Also, notice that $\theta\tilde{\nu} \subseteq Tu$ if and only if $\theta\nu_i^+ \leq t_i u$ and $-\theta\nu_i^- \geq t_{(n+i)} u$, $\forall i \in \{1, \ldots, n\}$. The proof is in two parts.

Proof of the "if direction": Let $\theta\tilde{\nu} \subseteq Tu$. For any $I \in S^+(\theta)$, we have $I \geq 0$ and $\|I\| \leq \theta$, so $\nu_i^+ \|I\| \leq \nu_i^+ \theta \leq t_i u$ and $-\nu_i^- \|I\| \geq -\nu_i^- \theta \geq t_{(n+i)} u$, $\forall i \in \{1, \ldots, n\}$, where we have used the fact that $\nu_i^+ \geq 0$ and $-\nu_i^- \leq 0$. Notice that

$$r_i I \leq r_i^+ I = |r_i^+ I| \leq \|r_i^+\| \|I\| = \nu_i^+ \|I\|, \tag{102}$$

where the first two steps are due to $I \geq 0$ and the third step is due to the Cauchy-Schwarz inequality (see Saad [2003]). Therefore, it follows that $r_i I \leq t_i u$. Similarly, notice that

$$r_i I \geq r_i^- I = -|r_i^- I| \geq -\|r_i^-\| \|I\| = -\nu_i^- \|I\|, \tag{103}$$

so $r_i I \geq t_{(n+i)} u$. Thus, $RI \in Tu$, so $I \in \mathcal{F}(u)$ and $S^+(\theta) \subseteq \mathcal{F}(u)$.

Proof of the "only if direction": Let $S^+(\theta) \subseteq \mathcal{F}(u)$. For any $i \in \{1 \ldots, n\}$, notice that if $r_i^+ = 0$, then $\nu_i^+ = 0$ and $\theta\nu_i^+ \leq t_i u$, because $t_i u \geq 0$. Otherwise, if $r_i^+ \neq 0$, then let $I = \theta \frac{(r_i^+)^T}{\nu_i^+} \geq 0$. Notice that $\|I\| = \theta$, because $\|(r_i^+)^T\| = \|r_i^+\| = \nu_i^+$, so $I \in S^+(\theta)$, from which $I \in \mathcal{F}(u)$, that is, $r_i I \leq t_i u$. Therefore, we have

$$r_i I = \theta r_i \frac{(r_i^+)^T}{\nu_i^+} \leq t_i u. \tag{104}$$

But $r_i (r_i^+)^T = r_i^+ (r_i^+)^T = \|r_i^+\|^2 = (\nu_i^+)^2$, so

$$\theta\nu_i^+ \leq t_i u. \tag{105}$$

Similarly, if $r_i^- = 0$, then $\nu_i^- = 0$ and $-\theta\nu_i^- \geq t_{(n+i)} u$, because $t_{(n+i)} u \leq 0$. Otherwise, if $r_i^- \neq 0$, then let $I' = -\theta \frac{(r_i^-)^T}{\nu_i^-} \geq 0$. Notice that $\|I'\| = \theta$, because $\|(r_i^-)^T\| = \|r_i^-\| = \nu_i^-$, so $I' \in S^+(\theta)$, from which $I' \in \mathcal{F}(u)$, that is, $r_i I' \geq t_{(n+i)} u$. Therefore, we have

$$r_i I' = -\theta r_i \frac{(r_i^-)^T}{\nu_i^-} \geq t_{(n+i)} u. \tag{106}$$

But $r_i (r_i^-)^T = r_i^- (r_i^-)^T = \|r_i^-\|^2 = (\nu_i^-)^2$, so

$$-\theta\nu_i^- \geq t_{(n+i)} u. \tag{107}$$

From Equations (105) and (107), it follows that $\theta\tilde{\nu} \subseteq Tu$. □

LEMMA D.3. *$\mathcal{F}(u_s)$ is maximal in $\mathcal{S}$.*

PROOF. Recall that Θ^* and u_s are well defined and $(\Theta^*, u_s) \in \mathcal{R}$, so $\Theta^*\tilde{\nu} \subseteq Tu_s$ and $\Theta^* \geq 0$. We will prove that $\Theta(\cdot)$ satisfies the conditions of Lemma D.1, from which it would follow that $\mathcal{F}(u_s)$ is maximal in $\mathcal{S}$. First, notice that for any $u, u' \in \mathcal{U}$, if $\mathcal{F}(u') = \mathcal{F}(u)$, then it follows that $\Theta(u') = \Theta(u)$, due to Equation (64). It remains to prove that for any $u, u' \in \mathcal{U}$, with $0 \in Tu$ and $0 \in Tu'$, if $Tu' \supset Tu$, then $\Theta(u') > \Theta(u)$.

Let $\lambda = \min_{\forall i} \left(|Tu'|_i - Tu|_i|\right) / \max_{\forall i}(|\tilde{\nu}_i|)$, which is well defined because $\tilde{\nu} \neq 0$ due to $R \neq 0$, and let $\theta' = \Theta(u) + \lambda$. Because $Tu' \supset Tu$, it follows that $\lambda > 0$ and $\theta' > \Theta(u) \geq 0$. Therefore,

$$\theta'\tilde{\nu} = \Theta(u)\tilde{\nu} + \frac{\min_{\forall i} \left(|Tu'|_i - Tu|_i|\right)}{\max_{\forall i}(|\tilde{\nu}_i|)} \tilde{\nu}, \tag{108}$$

$$\subseteq Tu + \min_{\forall i} \left(|Tu'|_i - Tu|_i|\right) \mathbb{1}_{2n}, \tag{109}$$

where $\mathbb{1}_{2n}$ is the $2n \times 1$ vector whose first n entries are 1 and the rest are -1, and in Equation (109) we used $(\Theta(u), u) \in \mathcal{R}$ and $\tilde{\nu}/\max_{\forall i}(|\tilde{\nu}_i|) \subseteq \mathbb{1}_{2n}$. Notice that, for any $k \in \{1, \ldots, n\}$, because $Tu \subset Tu'$, we have

$$\min_{\forall i} \left(|Tu'|_i - Tu|_i|\right) \leq |Tu'|_k - Tu|_k| = Tu'|_k - Tu|_k. \tag{110}$$

Likewise, for any $k \in \{n+1, \ldots, 2n\}$, we have

$$-\min_{\forall i} \left(|Tu'|_i - Tu|_i|\right) \geq -|Tu'|_k - Tu|_k| = Tu'|_k - Tu|_k. \tag{111}$$

Combining Equations (110) and (111), we get

$$\min_{\forall i} \left(|Tu'|_i - Tu|_i|\right) \mathbb{1}_{2n} \subseteq Tu' - Tu. \tag{112}$$

Therefore, due to Equations (109) and (112) and making use of Equation (21), we get

$$\theta' \tilde{\nu} \subseteq Tu + Tu' - Tu = Tu'. \tag{113}$$

This, coupled with $u' \in \mathcal{U}$, means that $(\theta', u') \in \mathcal{R}$, so $\Theta(u') \geq \theta' > \Theta(u)$, from which $\Theta(\cdot)$ satisfies the conditions of Lemma D.1 and $\mathcal{F}(u_s)$ is maximal in $\mathcal{S}$. □

Lemma D.4. *$\mathcal{F}(u_c)$ is maximal in $\mathcal{S}$.*

Proof. Recall that ζ, ω, and u_c are well defined and $(\zeta, \omega, u_c) \in \mathcal{C}$, so $R\zeta \in Tu_c$, $\omega\tilde{\nu} \subseteq Tu_c$, and $\zeta, \omega \geq 0$. We will prove that $\xi(\cdot)$ satisfies the conditions of Lemma D.1, from which it would follow that $\mathcal{F}(u_c)$ is maximal in $\mathcal{S}$. First, notice that for any $u, u' \in \mathcal{U}$, if $\mathcal{F}(u') = \mathcal{F}(u)$, then it follows that $\xi(u') = \xi(u)$, due to Equation (71). It remains to prove that for any $u, u' \in \mathcal{U}$, with $0 \in Tu$ and $0 \in Tu'$, if $Tu' \supset Tu$, then $\xi(u') > \xi(u)$.

For any $u \in \mathcal{U}$, there must exist a vector $I \in \mathcal{F}(u)$ and θ, where $\theta\tilde{\nu} \subseteq Tu$, such that $\sum_{j=1}^{m} I_j + m\theta = \xi(u)$. Let $\lambda = \min_{\forall i} \left(|Tu'|_i - Tu|_i|\right) / \max_{\forall i,j}(|r_{ij}|)$, which is well defined due to $R \neq 0$. Because $Tu \subset Tu'$, it follows that $\lambda > 0$. Also, let $e_1 \in \mathbb{R}^m$ be the vector whose first entry is 1 and all other entries are 0, and let $I' = I + \lambda e_1$. Because $\lambda > 0$, we have $\lambda e_1 \geq 0$, $\lambda e_1 \neq 0$, $I' \geq I \geq 0$, and $I' \neq I$, so $\sum_{j=1}^{m} I'_j + m\theta > \sum_{j=1}^{m} I_j + m\theta = \xi(u)$. Denote by c_j the jth column of R and notice that

$$RI' = RI + \lambda Re_1 = RI + \lambda c_1, \tag{114}$$

$$= RI + \frac{\min_{\forall i} \left(|Tu'|_i - Tu|_i|\right)}{\max_{\forall i,j}(|r_{ij}|)} c_1. \tag{115}$$

Let $\mathbb{1}_{2n}$ be the $2n \times 1$ vector whose first n entries are 1 and the rest are -1. Because $c_1/\max_{\forall i,j}(|r_{ij}|) \in \mathbb{1}_{2n}$, notice that

$$\frac{\min_{\forall i} \left(|Tu'|_i - Tu|_i|\right)}{\max_{\forall i,j}(|r_{ij}|)} c_1 \in \min_{\forall i}(|Tu'|_i - Tu|_i|)\mathbb{1}_{2n}, \tag{116}$$

which, combined with $RI \in Tu$ because $(I, \theta, u) \in \mathcal{C}$, and due to Lemma 3.3, gives

$$RI + \frac{\min_{\forall i} \left(|Tu'|_i - Tu|_i|\right)}{\max_{\forall i,j}(|r_{ij}|)} c_1 \in Tu + \min_{\forall i}(|Tu'|_i - Tu|_i|)\mathbb{1}_{2n}. \tag{117}$$

Therefore, using Equation (115), it follows that

$$RI' \in Tu + \min_{\forall i}(|Tu'|_i - Tu|_i|)\mathbb{1}_{2n}. \tag{118}$$

Also, notice that, for any $k \in \{1, \ldots, n\}$, because $Tu \subset Tu'$, we have

$$\min_{\forall i}(|Tu'|_i - Tu|_i|) \leq |Tu'|_k - Tu|_k| = Tu'|_k - Tu|_k. \tag{119}$$

Likewise, for any $k \in \{n+1, \ldots, 2n\}$, we have

$$-\min_{\forall i}(|Tu'|_i - Tu|_i|) \geq -|Tu'|_k - Tu|_k| = Tu'|_k - Tu|_k. \tag{120}$$

Combining Equations (119) and (120), we get

$$\min_{\forall i}(|Tu'|_i - Tu|_i|)\mathbb{1}_{2n} \subseteq Tu' - Tu. \tag{121}$$

This, combined with $Tu \subseteq Tu$ and making use of Equation (21), gives

$$Tu + \min_{\forall i}(|Tu'|_i - Tu|_i|)\mathbb{1}_{2n} \subseteq Tu + Tu' - Tu = Tu'. \tag{122}$$

Therefore, due to Equations (118) and (122), we get

$$RI' \in Tu'. \tag{123}$$

Also, we have $\theta\tilde{v} \subseteq Tu \subset Tu'$. Therefore, we have $I' \in \mathcal{F}(u')$ and θ satisfying $\theta\tilde{v} \subseteq Tu'$, with $\xi(u') \geq \sum_{j=1}^{m} I'_j + m\theta > \xi(u)$, so $\xi(\cdot)$ satisfies the conditions of Lemma D.1, and $\mathcal{F}(u_c)$ is maximal in $\mathcal{S}$. □

APPENDIX E: PROPERTIES OF THE MATRICES

In the following, we present key theoretical results that are useful to carry out some of the above proofs. First, we prove that the $n_v \times n_v$ matrix D given in Equation (8) is irreducible, which is required to prove $D^{-1} > 0$, a key result that is useful to prove Lemma E.5. Second, we prove that every element of Q is non-zero. This result is established in Lemma E.6, depends on the result of Lemma E.5, and is a key result in proving Theorem 4.13 and Lemma D.1. Finally, we prove that Q is SP in Lemma E.8, depending on Lemma E.7, which is used in the proofs of Theorem 4.13, Lemmas D.1, 4.4, and 4.5.

Definition E.1 (*Directed Graph*). A *directed graph* $\mathcal{G}$ is the combination of a set of vertices $\mathcal{V}(\mathcal{G})$ and a set of ordered pairs of vertices from $\mathcal{V}(\mathcal{G})$, called *directed edges*, $\mathcal{E}(\mathcal{G})$. If (v_i, v_j) is a directed edge of $\mathcal{G}$, then it is said to have a *direction* from v_i to v_j.

Definition E.2 (*Directed Path*). A *directed path* in a graph $\mathcal{G}$ is a sequence of vertices $v_0, v_1, \ldots, v_k$ where $(v_{i-1}, v_i) \in \mathcal{E}(\mathcal{G})$ for all $i \in \{1, \ldots, k\}$. The vertex v_k is said to be *reachable* from v_0, denoted as $v_0 \to v_k$.

Definition E.3 (*Strongly Connected*). A directed graph is said to be *strongly connected* if, for every pair of vertices, u, v, we have $u \to v$.

Any square $n \times n$ matrix A can be used to generate a graph $\mathcal{G}(A)$, defined as the directed graph on n vertices $\{v_1, \ldots, v_n\}$, in which $(v_i, v_j) \in \mathcal{E}(\mathcal{G}(A))$ if and only if $a_{ij} \neq 0$, where a_{ij} is the (i, j)th element of A. If the graph is strongly connected, then the matrix A is said to be *irreducible* (see Berman and Plemmons [1994]).

LEMMA E.4. *The $n_v \times n_v$ matrix D given in Equation* (8) *is irreducible.*

PROOF. Recall that $D = G + B + ME^{-1}M^T$, where G is an irreducible matrix with positive diagonal entries and non-positive off-diagonal entries and B is a non-negative diagonal matrix with a positive diagonal. We will show that $\mathcal{G}(D)$ is strongly connected, so D is irreducible. We start by showing that $ME^{-1}M^T$ has non-negative diagonal entries and non-positive off-diagonal entries. Notice that

$$(E^{-1})_{ij} = \begin{cases} \frac{\Delta t}{l_{ii}} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}. \tag{124}$$

Therefore, if $X = E^{-1}M^T$, then we have

$$x_{ij} = \sum_{k=1}^{n_l} (E^{-1})_{ik}(M^T)_{kj}, \quad \begin{matrix} i = 1, \ldots, n_l \\ j = 1, \ldots, n_v \end{matrix}, \tag{125}$$

so $x_{ij} = \frac{\Delta t}{l_{ii}} m_{ji}$. Also, if $W = ME^{-1}M^T$, then

$$w_{ij} = \sum_{k=1}^{n_l} m_{ik}x_{kj} = \sum_{k=1}^{n_l} m_{ik}\frac{\Delta t}{l_{kk}} m_{jk}, \quad \begin{matrix} i = 1, \ldots, n_v \\ j = 1, \ldots, n_v \end{matrix}. \tag{126}$$

By definition, every column of the matrix M contains either one non-zero entry or two non-zero entries where one of them is $+1$ and the other is -1. It follows that, for any $i \neq j$, we have $m_{ik}m_{jk} \leq 0$, for any k, so $w_{ij} \leq 0$, $\forall i \neq j$. Also, we have $m_{ik}m_{ik} \geq 0$, for any k, so $w_{ii} \geq 0$, $\forall i$. Therefore, $W = ME^{-1}M^T$ has non-negative diagonal entries and non-positive off-diagonal entries. Therefore, $B + ME^{-1}M^T$ has positive diagonal entries and non-positive off-diagonal entries, similarly to G. It follows that $d_{ij} \neq 0$ whenever $g_{ij} \neq 0$, so $\mathcal{E}(\mathcal{G}(G)) \subseteq \mathcal{E}(\mathcal{G}(D))$. Therefore, $\mathcal{G}(D)$ is strongly connected, so D is irreducible. □

Lemma E.5. *For the $n \times n$ matrix F given in Equation* (12)*, and its extension $\tilde{F}$ according to Definition 3.8, the directed graph $\mathcal{G}(\tilde{F})$ is strongly connected.*

Proof. We can represent the $2n \times 2n$ matrix $Z \triangleq \tilde{F}$ as follows:

$$Z = \begin{bmatrix} Z_{11} & Z_{12} & Z_{13} & Z_{14} \\ Z_{21} & Z_{22} & Z_{23} & Z_{24} \\ Z_{31} & Z_{32} & Z_{33} & Z_{34} \\ Z_{41} & Z_{42} & Z_{43} & Z_{44} \end{bmatrix}, \tag{127}$$

$$= \begin{bmatrix} (D^{-1}B)^+ & (D^{-1}M)^+ & (D^{-1}B)^- & (D^{-1}M)^- \\ (-E^{-1}M^TD^{-1}B)^+ & (I_{n_l} - E^{-1}M^TD^{-1}M)^+ & (-E^{-1}M^TD^{-1}B)^- & (I_{n_l} - E^{-1}M^TD^{-1}M)^- \\ (D^{-1}B)^- & (D^{-1}M)^- & (D^{-1}B)^+ & (D^{-1}M)^+ \\ (-E^{-1}M^TD^{-1}B)^- & (I_{n_l} - E^{-1}M^TD^{-1}M)^- & (-E^{-1}M^TD^{-1}B)^+ & (I_{n_l} - E^{-1}M^TD^{-1}M)^+ \end{bmatrix}, \tag{128}$$

where, using the notation introduced in Definition 3.8,

$$Z_{11} = Z_{33} = (D^{-1}B)^+, \tag{129}$$
$$Z_{13} = Z_{31} = (D^{-1}B)^-, \tag{130}$$
$$Z_{12} = Z_{34} = (D^{-1}M)^+, \tag{131}$$
$$Z_{14} = Z_{32} = (D^{-1}M)^-, \tag{132}$$
$$Z_{21} = Z_{43} = (-E^{-1}M^TD^{-1}B)^+, \tag{133}$$
$$Z_{23} = Z_{41} = (-E^{-1}M^TD^{-1}B)^-, \tag{134}$$
$$Z_{22} = Z_{44} = (I_{n_l} - E^{-1}M^TD^{-1}M)^+, \tag{135}$$
$$Z_{24} = Z_{42} = (I_{n_l} - E^{-1}M^TD^{-1}M)^-. \tag{136}$$

The matrix Z can be used to construct a graph $\mathcal{G}(Z)$ whose vertices are $\mathcal{V} = \{v_1, v_2, \ldots, v_{2n}\}$ and whose directed edges are (v_i, v_j) for every $z_{ij} \neq 0$, where z_{ij} is the (i, j)th element of Z. Let $\mathcal{E}$ denote the set of edges of $\mathcal{G}(Z)$. Also, consider a partition $\mathcal{V}_1 = \{v_1, v_2, \ldots, v_{n_v}\}$, $\mathcal{V}_2 = \{v_{n_v+1}, v_{n_v+2}, \ldots, v_n\}$, $\mathcal{V}_3 = \{v_{n+1}, \ldots, v_{n+n_v}\}$, and $\mathcal{V}_4 = \{v_{n+n_v+1}, \ldots, v_{2n}\}$ of $\mathcal{V}$. For any two vertices $u, v \in \mathcal{V}$, define a binary-valued function $\beta(u, v)$ as follows:

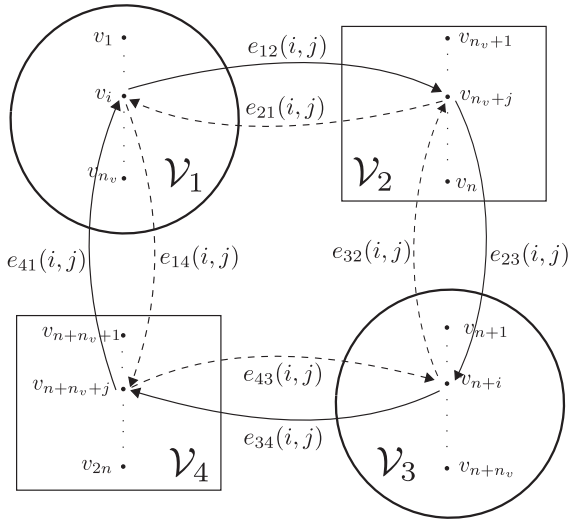


Fig. 10. A high-level representation of $\mathcal{G}(Z)$.

$$\beta(u,v) = \begin{cases} 1, & \text{if } u \leftrightarrow v; \\ 0, & \text{otherwise.} \end{cases} \tag{137}$$

where $u \leftrightarrow v$ is a shorthand for $u \rightarrow v$ and $v \rightarrow u$. It should be clear that $\beta(\cdot)$ is transitive, that is, for any three vertices $u, v, w \in \mathcal{V}$, if $\beta(u,v) = 1$ and $\beta(v,w) = 1$, then $\beta(u,w) = 1$, and $\beta(\cdot)$ is commutative, that is, $\beta(u,v) = \beta(v,u)$. In the following, we will show that for any $u, v \in \mathcal{V}$ we have $\beta(u,v) = 1$, so $\mathcal{G}(Z)$ is strongly connected.

We start by proving the following properties on the vertices of $\mathcal{G}(Z)$,

$$\forall u, v \in \mathcal{V}_1 \qquad \beta(u,v) = 1, \tag{138}$$

$$\forall u, v \in \mathcal{V}_3 \qquad \beta(u,v) = 1, \tag{139}$$

$$\forall v \in \mathcal{V}_2, \exists u \in \mathcal{V}_1 \quad \text{such that} \quad \beta(u,v) = 1, \tag{140}$$

$$\forall v \in \mathcal{V}_4, \exists u \in \mathcal{V}_3 \quad \text{such that} \quad \beta(u,v) = 1, \tag{141}$$

which will lead us to the desired result. To better visualize the proof, refer to Figure 10.

Recall that $D^{-1} > 0$, from which $D^{-1}B > 0$, because $B = C/\Delta t \geq 0$ is a diagonal matrix with non-zero diagonal elements. Therefore, we have $(D^{-1}B)^+ = D^{-1}B > 0$ which, due to Equation (129), gives

$$u \rightarrow v, \qquad \forall u, v \in \mathcal{V}_1, \tag{142}$$

$$u \rightarrow v, \qquad \forall u, v \in \mathcal{V}_3, \tag{143}$$

This proves Equations (138) and (139).

For any $i \in \{1, 2, \ldots, n_v\}$ and $j \in \{1, 2, \ldots, n_l\}$, notice that $v_i \in \mathcal{V}_1$, $v_{n_v+j} \in \mathcal{V}_2$, $v_{n+i} \in \mathcal{V}_3$, and $v_{n+n_v+j} \in \mathcal{V}_4$, as shown in Figure 10. Also, for any $i \in \{1, 2, \ldots, n_v\}$ and $j \in \{1, 2, \ldots, n_l\}$, we define the following indexing scheme and notation for certain edges in $\mathcal{E}$:

$$e_{12}(i,j) = (v_i, v_{n_v+j}) \in \mathcal{E} \iff (Z_{12})_{ij} \neq 0, \tag{144}$$
$$e_{21}(i,j) = (v_{n_v+j}, v_i) \in \mathcal{E} \iff (Z_{21})_{ij} \neq 0, \tag{145}$$
$$e_{23}(i,j) = (v_{n_v+j}, v_{n+i}) \in \mathcal{E} \iff (Z_{23})_{ij} \neq 0, \tag{146}$$
$$e_{32}(i,j) = (v_{n+i}, v_{n_v+j}) \in \mathcal{E} \iff (Z_{32})_{ij} \neq 0, \tag{147}$$
$$e_{34}(i,j) = (v_{n+i}, v_{n+n_v+j}) \in \mathcal{E} \iff (Z_{34})_{ij} \neq 0, \tag{148}$$
$$e_{43}(i,j) = (v_{n+n_v+j}, v_{n+i}) \in \mathcal{E} \iff (Z_{43})_{ij} \neq 0, \tag{149}$$
$$e_{41}(i,j) = (v_{n+n_v+j}, v_i) \in \mathcal{E} \iff (Z_{41})_{ij} \neq 0, \tag{150}$$
$$e_{14}(i,j) = (v_i, v_{n+n_v+j}) \in \mathcal{E} \iff (Z_{14})_{ij} \neq 0. \tag{151}$$

Then

$$Z_{12} = Z_{34} \text{ from (131) leads to } \quad e_{12}(i,j) \in \mathcal{E} \iff e_{34}(i,j) \in \mathcal{E}, \tag{152}$$
$$Z_{14} = Z_{32} \text{ from (132) leads to } \quad e_{14}(i,j) \in \mathcal{E} \iff e_{32}(i,j) \in \mathcal{E}, \tag{153}$$
$$Z_{21} = Z_{43} \text{ from (133) leads to } \quad e_{21}(i,j) \in \mathcal{E} \iff e_{43}(i,j) \in \mathcal{E}, \tag{154}$$
$$Z_{23} = Z_{41} \text{ from (134) leads to } \quad e_{23}(i,j) \in \mathcal{E} \iff e_{41}(i,j) \in \mathcal{E}. \tag{155}$$

Now let $X = D^{-1}M$ be a $n_v \times n_l$ matrix and $Y = -E^{-1}M^T D^{-1}B$ be a $n_l \times n_v$ matrix, and notice that $Y = -E^{-1}X^T B$, where $E^{-1} \geq 0$ and $B \geq 0$ are diagonal matrices with non-zero diagonal elements, so $y_{ij} = -\frac{1}{e_{jj}}(X^T)_{ji}b_{ii} = -\frac{1}{e_{jj}}x_{ij}b_{ii}$. Thus, the corresponding non-zero elements of Y and X^T have opposite signs. Two things follow from this:

1) Considering the positive elements of Y,

$$(Y^+)_{ji} \neq 0 \iff y_{ji} > 0 \iff x_{ij} < 0 \iff (X^-)_{ij} \neq 0.$$

But $Y^+ = Z_{21}$ and $X^- = Z_{14}$, so

$$(Z_{21})_{ji} \neq 0 \iff (Z_{14})_{ij} \neq 0,$$

from which, due to Equations (145) and (151),

$$e_{21}(i,j) \in \mathcal{E} \iff e_{14}(i,j) \in \mathcal{E}. \tag{156}$$

2) Considering the negative elements of Y:

$$(Y^-)_{ji} \neq 0 \iff y_{ji} < 0 \iff x_{ij} > 0 \iff (X^+)_{ij} \neq 0.$$

But $Y^- = Z_{23}$ and $X^+ = Z_{12}$, so

$$(Z_{23})_{ji} \neq 0 \iff (Z_{12})_{ij} \neq 0,$$

from which, due to Equations (146) and (144),

$$e_{23}(i,j) \in \mathcal{E} \iff e_{12}(i,j) \in \mathcal{E}. \tag{157}$$

Furthermore, let x_j and m_j be the jth columns of X and M, respectively. Note that $m_j \neq 0$, by definition of M, and D^{-1} is non-singular, so $x_j = D^{-1}m_j \neq 0$. Therefore, $x_{ij} \neq 0$ for some $i \in \{1, 2, \ldots, n_v\}$, so either $(Z_{12})_{ij} \neq 0$ or $(Z_{32})_{ij} \neq 0$, depending on whether $x_{ij} > 0$ or $x_{ij} < 0$. By Equations (145) and (147), it follows that *either* $e_{12}(i,j) \in \mathcal{E}$ *or* $e_{32}(i,j) \in \mathcal{E}$.

This being said, and due to Equations (152)–(157), for every i and j, we have either $e_{12}(i,j)$, $e_{23}(i,j)$, $e_{34}(i,j)$, and $e_{41}(i,j) \in \mathcal{E}$ or $e_{32}(i,j)$, $e_{21}(i,j)$, $e_{14}(i,j)$, and $e_{43}(i,j) \in \mathcal{E}$. Thus, for any vertex in $\mathcal{V}_2$ and the corresponding vertex in $\mathcal{V}_4$, as indexed by j, there exists a cycle connecting all the partitions of $\mathcal{V}$ and passing through these two vertices as shown in Figure 10. This completes the proof of Equations (140) and (141).

Now we are ready to show that for any two vertices $u, v \in \mathcal{V}$, we have $\beta(u,v) = 1$. Notice that $\forall u, v \in \mathcal{V}_1 \cup \mathcal{V}_2$, $\beta(u,v) = 1$, due to the following:

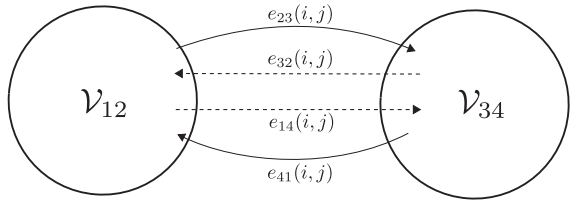


Fig. 11. A high-level representation of $\mathcal{G}(Z)$.

—if u or $v \in \mathcal{V}_1$, then, clearly, $u \leftrightarrow v$, either due to Equation (138) or due to Equations (138) and (140) and due to transitivity of $\beta(\cdot)$.
—if $u, v \in \mathcal{V}_2$, then there exist $w, w' \in \mathcal{V}_1$ such that $\beta(w, u) = 1$ and $\beta(w', v) = 1$, due to Equation (140), which, combined with $\beta(\cdot)$ being commutative and transitive, and due to Equation (138), gives $\beta(u, v) = 1$.

Therefore, we will combine $\mathcal{V}_1$ and $\mathcal{V}_2$ as $\mathcal{V}_{12} \triangleq \mathcal{V}_1 \cup \mathcal{V}_2$, so $\mathcal{V}_{12}$ is strongly connected. Likewise, $\mathcal{V}_{34} \triangleq \mathcal{V}_3 \cup \mathcal{V}_4$ is strongly connected, due to Equation (139) and Equation (141). W can now look at $\mathcal{G}(Z)$ using the simple representation of Figure 11, where $\mathcal{V}_{12}$ and $\mathcal{V}_{34}$ are strongly connected. Moreover, because either $e_{23}(i, j)$ and $e_{41}(i, j) \in \mathcal{E}$ or $e_{14}(i, j)$ and $e_{32}(i, j) \in \mathcal{E}$, then $\mathcal{G}(Z)$ is strongly connected. □

LEMMA E.6. *Every element of Q is non-zero.*

PROOF. Let $Z = \tilde{F}$. Note that $Q = (I_{2n} - Z)^{-1} = \sum_{k=0}^{\infty} Z^k$, because $\rho(Z) < 1$ [Saad 2003]. In the following, we will first show that $|Z^k| = |Z|^k$, for every integer $k \geq 1$, starting with the block-form of Z^i in the following notation:

$$Z^i = \begin{bmatrix} Z_{11}^{(i)} & Z_{12}^{(i)} \\ Z_{21}^{(i)} & Z_{22}^{(i)} \end{bmatrix}. \tag{158}$$

Recall that $Z = \tilde{F}$ is SP, due to Definition 3.8 and Lemma A.1, so Z^k is SP, due to Lemma A.2, which, due to Lemma A.1, gives $Z_{11}^{(i)} \geq 0$, $Z_{12}^{(i)} \leq 0$, $Z_{21}^{(i)} \leq 0$, and $Z_{22}^{(i)} \geq 0$.

Let $Y \triangleq |Z|$, where $|Z|$ is the matrix consisting of the absolute values of the elements of Z, and represent Y^i as follows:

$$Y^i = \begin{bmatrix} Y_{11}^{(i)} & Y_{12}^{(i)} \\ Y_{21}^{(i)} & Y_{22}^{(i)} \end{bmatrix}. \tag{159}$$

We will prove by induction that $Y^i = |Z^i|$ for every $i \geq 1$. Notice that for $i = 1$, the result is trivially true, and suppose that $Y^{k-1} = |Z^{k-1}|$, so

$$Y^k = Y^{k-1}Y = |Z^{k-1}||Z| = \begin{bmatrix} Z_{11}^{(k-1)} & -Z_{12}^{(k-1)} \\ -Z_{21}^{(k-1)} & Z_{22}^{(k-1)} \end{bmatrix} \begin{bmatrix} Z_{11} & -Z_{12} \\ -Z_{21} & Z_{22} \end{bmatrix} \tag{160}$$

$$= \begin{bmatrix} Z_{11}^{(k-1)}Z_{11} + Z_{12}^{(k-1)}Z_{21} & -Z_{11}^{(k-1)}Z_{12} - Z_{12}^{(k-1)}Z_{22} \\ -Z_{21}^{(k-1)}Z_{11} - Z_{22}^{(k-1)}Z_{21} & Z_{21}^{(k-1)}Z_{12} + Z_{22}^{(k-1)}Z_{22} \end{bmatrix}, \tag{161}$$

while

$$Z^k = Z^{k-1}Z = \begin{bmatrix} Z_{11}^{(k-1)} & Z_{12}^{(k-1)} \\ Z_{21}^{(k-1)} & Z_{22}^{(k-1)} \end{bmatrix} \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}, \tag{162}$$

$$= \begin{bmatrix} Z_{11}^{(k-1)}Z_{11} + Z_{12}^{(k-1)}Z_{21} & Z_{11}^{(k-1)}Z_{12} + Z_{12}^{(k-1)}Z_{22} \\ Z_{21}^{(k-1)}Z_{11} + Z_{22}^{(k-1)}Z_{21} & Z_{21}^{(k-1)}Z_{12} + Z_{22}^{(k-1)}Z_{22} \end{bmatrix}. \tag{163}$$

Recall that Z^k is SP which, due to Lemma A.1, means that the diagonal blocks in Equation (163) are non-negative and the off-diagonal blocks are non-positive. Thus, comparing Equations (161) and (163), we see that $Y^k = |Z^k|$, for every k. Furthermore, notice that $\mathcal{G}(Z) = \mathcal{G}(|Z|)$, due to the definition of $\mathcal{G}(\cdot)$, and $\mathcal{G}(Z)$ is strongly connected, due to Lemma E.5, so $\mathcal{G}(|Z|)$ is strongly connected and $|Z|$ is irreducible (see Berman and Plemmons [1994]). This, combined with $|Z| \geq 0$, leads to $|Z|^p > 0$, for some integer $p \geq 1$ (see Berman and Plemmons [1994]). Therefore, $|Z^p| > 0$, due to $|Z^p| = |Z|^p$, so every element of Z^p is non-zero. Let $z_{ij}^{(k)}$ be the (i, j)th element of Z^k, so the (i, j)th element of $Q = \sum_{k=0}^{\infty} Z^k$ is

$$q_{ij} = \sum_{k=0}^{p-1} z_{ij}^{(k)} + z_{ij}^{(p)} + \sum_{k=p+1}^{\infty} z_{ij}^{(k)} \neq 0,$$

where we used the fact that $z_{ij}^{(k)}$ have the same sign, $\forall k$, due to Z^k being SP, $\forall k$, and Lemma A.1, and $z_{ij}^{(p)} \neq 0$, because every element of Z^p is non-zero. It follows that every element of Q is non-zero. □

Lemma E.7. *Let $u^{(k)}, v^{(k)} \in \mathbb{R}^{2n}$ be sequences of vectors. If $u^{(k)} \subseteq v^{(k)}$, $\forall k \geq 1$, and if $u \triangleq \lim_{k\to\infty} u^{(k)}$ and $v \triangleq \lim_{k\to\infty} v^{(k)}$ exist, then $u \subseteq v$.*

Proof. Let $u^{(k)} = \begin{bmatrix} u_t^{(k)} \\ u_b^{(k)} \end{bmatrix}$, $v^{(k)} = \begin{bmatrix} v_t^{(k)} \\ v_b^{(k)} \end{bmatrix}$, $u = \begin{bmatrix} u_t \\ u_b \end{bmatrix}$, and $v = \begin{bmatrix} v_t \\ v_b \end{bmatrix}$, where $u_t^{(k)}$, $v_t^{(k)}$, u_t, v_t, $u_b^{(k)}$, $v_b^{(k)}$, u_b, and v_b are $n \times 1$ vectors. In the following, we will show that $u_t \leq v_t$ and $u_b \geq v_b$, so $u \subseteq v$.

Let $w_t^{(k)} = v_t^{(k)} - u_t^{(k)} \geq 0$, $\forall k$, and let $w_t \triangleq \lim_{k\to\infty} w_t^{(k)} = \lim_{k\to\infty}(v_t^{(k)} - u_t^{(k)}) = \lim_{k\to\infty} v_t^{(k)} - \lim_{k\to\infty} u_t^{(k)} = v_t - u_t$. If $w_t < 0$, then there exists an integer $N \geq 1$ such that $|w_t^{(k)} - w_t| < -w_t$, $\forall k \geq N$, due to the definition of limits [Bartle and Sherbert 1992]. It follows that $w_t < w_t^{(k)} - w_t < -w_t$, $\forall k \geq N$, so $w_t^{(k)} < 0$, $\forall k \geq N$, and we have a contradiction. Therefore, $w_t = v_t - u_t \geq 0$ and $u_t \leq v_t$. Similarly, let $w_b^{(k)} = u_b^{(k)} - v_b^{(k)} \geq 0$, $\forall k$, and let $w_b \triangleq \lim_{k\to\infty} w_b^{(k)} = \lim_{k\to\infty}(u_b^{(k)} - v_b^{(k)}) = \lim_{k\to\infty} u_b^{(k)} - \lim_{k\to\infty} v_b^{(k)} = u_b - v_b$. If $w_b < 0$, then there exists an integer $N \geq 1$ such that $|w_b^{(k)} - w_b| < -w_b$, $\forall k \geq N$, due to the definition of limits [Bartle and Sherbert 1992]. It follows that $w_b < w_b^{(k)} - w_b < -w_b$, $\forall k \geq N$, so $w_b^{(k)} < 0$, $\forall k \geq N$, and we have a contradiction. Therefore, $w_b = u_b - v_b \geq 0$, $u_b \geq v_b$, and $u \subseteq v$. □

Lemma E.8. *Q is SP.*

Proof. Let $Z = \tilde{F}$. Recall that, because $\rho(Z) < 1$, then the summation $\sum_{k=0}^{\infty} Z^k$ exists, and we have $Q = (I_{2n} - Z)^{-1} = \sum_{k=0}^{\infty} Z^k$ [Saad 2003]. The proof is by induction. Notice that Z^0 is SP, due to Lemma A.1. Suppose that Z^{k-1} is SP, then $Z^k = Z^{k-1}Z$ is also SP, due to Lemma A.2. Therefore, Z^k is SP, for any $k \geq 0$. Also, notice that, for any two

$2n \times 1$ vectors $u \subseteq v$ and for any integer $p \geq 1$, we have

$$\left[\sum_{k=1}^{p} Z^k\right] u = \sum_{k=1}^{p} [Z^k u] \subseteq \sum_{k=1}^{p} [Z^k v] = \left[\sum_{k=1}^{p} Z^k\right] v, \tag{164}$$

where in the second step we used the fact that Z^k is SP, for any $k \geq 0$, and that the finite sum of SP matrices is SP, due to Lemma A.2. Because $\lim_{p\to\infty}[\sum_{k=1}^{p} Z^k]$ exists and converges to $(I_{2n} - Z)^{-1}$, taking the limits on both sides of Equation (164), due to Lemma E.7, leads to

$$Qu = (I_{2n} - Z)^{-1} u \subseteq (I_{2n} - Z)^{-1} v = Qv. \tag{165}$$

Hence, Q is SP, which completes the proof. □

REFERENCES

N. H. Abdul Ghani and F. N. Najm. 2011. Fast vectorless power grid verification under an RLC model. *IEEE Trans. Comput.-Aid. Des. Integr. Circ. Syst.* 30, 5 (May 2011), 691–703.

R. G. Bartle and D. R. Sherbert. 1992. *Introduction to Real Analysis*. Wiley, New York, NY.

A. Berman and R. J. Plemmons. 1994. *Nonnegative Matrices in the Mathematical Science*. Society for Industrial and Applied Mathematics.

M. Fawaz and F. N. Najm. 2016. Fast vectorless RLC verification. (unpublished). DOI:http://dx.doi.org/10.1109/TCAD.2016.2589899

K. D. Joshi. 2001. *Applied Discrete Structures*. New Age International Pvt Ltd Publishers.

D. Kouroussis and F. N. Najm. 2003. A static pattern-independent technique for power grid voltage integrity verification. In *Proceedings of the ACM/IEEE 40th Design Automation Conference (DAC-03)*. 99–104.

A. Krstic and K.-T. Cheng. 1997. Vector generation for maximum instantaneous current through supply lines for CMOS circuits. In *Proceedings of the 34th Design Automation Conference*. 383–388.

J. D. Lambert. 1991. *Numerical Methods for Ordinary Differential Systems: The Initial Value Problem*. John Wiley & Sons, Inc., New York, NY.

W.-H. Lee, S. Pant, and D. Blaauw. 2004. Analysis and reduction of on-chip inductance effects in power supply grids. In *Proceedings of the IEEE International Symposium on Quality Electronic Design (ISQED)*. San Jose, CA, 131–136.

MOSEK ApS. 2015. *The MOSEK C Optimizer API Manual. Version 7.1 (Revision 28)*. Retrieved from DOI:http://docs.mosek.com/7.1/toolbox/index.html.

Z. Moudallal and F. N. Najm. 2015. Generating circuit current constraints to guarantee power grid safety. In *Proceedings of the IEEE/ACM Asia and South Pacific Design Automation Conference (ASP-DAC)*. Tokyo, Japan, 358–365.

Z. Moudallal and F. N. Najm. 2016. Generating current budgets to guarantee power grid safety. *IEEE Trans. Comput.-Aid. Des. Integr. Circ. Syst.* 35, 11 (Nov. 2016), 1914–1927.

A. Muramatsu, M. Hashimoto, and H. Ondera. 2005. Effects of on-chip inductance on power distribution grid. In *Proceedings of the ACM International Symposium on Physical Design*. San Francisco, CA, 63–69.

S. Pant, D. Blaauw, V. Zolotov, S. Sundareswaran, and R. Panda. 2004. A stochastic approach to power grid analysis. In *Proceedings of the ACM/IEEE 41st Design Automation Conference (DAC-04)*. 171–176.

Y. Saad. 2003. *Iterative Methods for Sparse Linear Systems*. SIAM, Philadelphia, PA.

N. Srivastava, X. Qi, and K. Banerjee. 2005. Impact of on-chip inductance on power distribution network design for nanometer scale integrated circuits. In *Proceedings of the IEEE International Symposium on Quality Electronic Design (ISQED)*. 341–351.

R. S. Varga. 1962. *Matrix Iterative Analysis*. Prentice-Hall, Inc., Englewood Cliffs, NJ.

Received September 2016; revised February 2017; accepted February 2017