Asynchronous and Distributed Tracking of Time-Varying Fixed Points

Andrey Bernstein
National Renewable Energy Laboratory

April 11th, 2019 / Autonomous Energy Systems Workshop

Acknowledgments

Emiliano Dall'Anese

University of Colorado Boulder

Andrea Simonetto

IBM

Iterative Algorithms

Iterative algorithms expressed in the generic form:

$$\mathbf{x}^{(k+1)} = f(\mathbf{x}^{(k)}),\ k = 1, 2, \ldots$$

Convergence to a fixed point $\mathbf{x}^* = f(\mathbf{x}^*)$ hinges on conditions on f such as

- contraction map
- an α-averaged operator
- paracontraction map
- $\ldots$

Iterative Algorithms

Iterative algorithms expressed in the generic form:

$$\mathbf{x}^{(k+1)} = f(\mathbf{x}^{(k)}),\ k = 1, 2, \ldots$$

What happens if

- f is time-varying?
  Application examples: time-varying optimization, time-varying load-flow.
- the iteration is distributed and computed asynchronously?
  Application examples: distributed optimization, multi-area load-flow.
- only approximate map $\tilde{f}$ is available?
  Application examples: optimization with approximate gradient, feedback-based optimization.

Related Work

- Bertsekas, Mathematical Programming, 1983 [Asynchronous, stationary]
- Frommer and Szyld, Journal of Computational and Applied Mathematics, 2000 [Asynchronous, common fixed point]
- Fullmer et al, CDC 2016, 2017 [Asynchronous, paracontractions, common fixed point]
- Simonetto, arxiv, 2017. [Time-varying, synchronous]
- This talk is based on: Bernstein and Dall'Anese, arxiv, CDC 2018.

Time-Varying Setting

Given a sequence of mappings $f^{(t)} : \mathbb{R}^m \to \mathbb{R}^m$, a batch solution given by the iteration

$$\mathbf{x}^{(k+1,t)} = f^{(t)}(\mathbf{x}^{(k,t)}),\ k = 1, 2, \ldots$$
$$\mathbf{x}^{(k,t)} \to \mathbf{x}^{(*,t)} \ \text{ as } \ k \to \infty.$$

- t is the time index
- k is the iteration index

Time-Varying Setting

Given a sequence of mappings $f^{(t)} : \mathbb{R}^m \to \mathbb{R}^m$, a batch solution given by the iteration

$$\mathbf{x}^{(k+1,t)} = f^{(t)}(\mathbf{x}^{(k,t)}),\ k = 1, 2, \ldots$$
$$\mathbf{x}^{(k,t)} \to \mathbf{x}^{(*,t)} \ \text{ as } \ k \to \infty.$$

- t is the time index
- k is the iteration index

A running (or online) algorithm:

$$\mathbf{x}^{(t+1)} = f^{(t)}(\mathbf{x}^{(t)}),$$

- performs one iteration at every time step

Time-Varying Setting

Given a sequence of mappings $f^{(t)} : \mathbb{R}^m \to \mathbb{R}^m$, a batch solution given by the iteration

$$\mathbf{x}^{(k+1,t)} = f^{(t)}(\mathbf{x}^{(k,t)}),\ k = 1, 2, \ldots$$
$$\mathbf{x}^{(k,t)} \to \mathbf{x}^{(*,t)} \ \text{ as } \ k \to \infty.$$

- t is the time index
- k is the iteration index

A running (or online) algorithm:

$$\mathbf{x}^{(t+1)} = f^{(t)}(\mathbf{x}^{(t)}),$$

- performs one iteration at every time step

Tracking question: $\boxed{\|\mathbf{x}^{(t)} - \mathbf{x}^{(*,t)}\| \leq ?}$

Online Algorithm

Assumption

$\{f^{(t)}\}$ are self-maps on a common set $\mathcal{D} \subseteq \mathbb{R}^m$ and contractions with common coefficient $L < 1$.

Therefore, there exists a sequence of unique fixed points $\{x^{(*,t)}\}$. Define:

$$\sigma^{(t)} := \|\mathbf{x}^{(*,t+1)} - \mathbf{x}^{(*,t)}\| \leq \sigma.$$

Online Algorithm

Assumption

$\{f^{(t)}\}$ are self-maps on a common set $\mathcal{D} \subseteq \mathbb{R}^m$ and contractions with common coefficient $L < 1$.

Therefore, there exists a sequence of unique fixed points $\{x^{(*,t)}\}$. Define:

$$\sigma^{(t)} := \|\mathbf{x}^{(*,t+1)} - \mathbf{x}^{(*,t)}\| \leq \sigma.$$

Online algorithm to track $\{x^{(*,t)}\}$:

$$\boxed{\mathbf{x}^{(t+1)} = \tilde{f}^{(t)}(\mathbf{x}^{(t)}), \quad t = 1, 2, \ldots}$$

where approximate maps $\{\tilde{f}^{(t)} : \mathcal{D} \to \mathcal{D}\}$ satisfy:

$$\|f^{(t)}(\mathbf{x}) - \tilde{f}^{(t)}(\mathbf{x})\| \leq e_f^{(t)} \leq e_f, \ \forall \mathbf{x} \in \mathcal{D}. \tag{1}$$

Tracking Result

Theorem

For each t, it holds that

$$\|\mathbf{x}^{(t)} - \mathbf{x}^{(*,t)}\| \leq L^t \|\mathbf{x}^{(1)} - \mathbf{x}^{(*,1)}\| + \sum_{\tau=1}^{t} L^{\tau-1} \left(e_f^{(t-\tau)} + \sigma^{(t-\tau)} \right).$$

In particular,

$$\limsup_{t\to\infty} \|\mathbf{x}^{(t)} - \mathbf{x}^{(*,t)}\| \leq \frac{e_f + \sigma}{1 - L}.$$

Distributed Setting

A network of agents with dependency graph $\mathcal{G} = (\mathcal{N}, \mathcal{A})$

- $\mathcal{N} := \{1, \ldots, N\}$ is the set of agents
- $\mathcal{A}$ is the set of directed edges, which represent information exchanges that are required in order to perform iteration

$$\mathcal{N}_i := \{j \in \mathcal{N} : (j, i) \in \mathcal{A}\}$$

Distributed Setting

A network of agents with dependency graph $\mathcal{G} = (\mathcal{N}, \mathcal{A})$

- $\mathcal{N} := \{1, \ldots, N\}$ is the set of agents
- $\mathcal{A}$ is the set of directed edges, which represent information exchanges that are required in order to perform iteration

$$\mathcal{N}_i := \{j \in \mathcal{N} : (j, i) \in \mathcal{A}\}$$

Distributed online algorithm:

$$\boxed{\mathbf{x}_i^{(t+1)} = \tilde{f}_i^{(t)} \left(\{\mathbf{x}_j^{(t)}\}_{j \in \mathcal{N}_i \cup \{i\}} \right), \quad i \in \mathcal{N}.}$$

Asynchronous Computation

$$\boxed{\mathbf{x}_i^{(t+1)} = \tilde{f}_i^{(t)}\left(\mathbf{x}_i^{(t)}, \{\tilde{\mathbf{x}}_j^{(t)}\}_{j\in\mathcal{N}_i}\right), \quad i \in \mathcal{N},}$$

where

$$\tilde{\mathbf{x}}_j^{(t)} = \mathbf{x}_j^{(D_{i,j}^{(t)})}$$

for some $D_{i,j}^{(t)} \in \{1, \ldots, t\}$.

Asynchronous Computation

$$\boxed{\mathbf{x}_i^{(t+1)} = \tilde{f}_i^{(t)}\left(\mathbf{x}_i^{(t)}, \{\tilde{\mathbf{x}}_j^{(t)}\}_{j\in\mathcal{N}_i}\right), \quad i \in \mathcal{N},}$$

where

$$\tilde{\mathbf{x}}_j^{(t)} = \mathbf{x}_j^{(D_{i,j}^{(t)})}$$

for some $D_{i,j}^{(t)} \in \{1, \ldots, t\}$.

If $D_{i,j}^{(t)} < t$, then $\tilde{\mathbf{x}}_j^{(t)}$ is an outdated copy of the variable associated with agent j.

Asynchronous Computation

$$\boxed{\mathbf{x}_i^{(t+1)} = \tilde{f}_i^{(t)}\left(\mathbf{x}_i^{(t)}, \{\tilde{\mathbf{x}}_j^{(t)}\}_{j\in\mathcal{N}_i}\right), \quad i \in \mathcal{N},}$$

where

$$\tilde{\mathbf{x}}_j^{(t)} = \mathbf{x}_j^{(D_{i,j}^{(t)})}$$

for some $D_{i,j}^{(t)} \in \{1, \ldots, t\}$.

If $D_{i,j}^{(t)} < t$, then $\tilde{\mathbf{x}}_j^{(t)}$ is an outdated copy of the variable associated with agent j.

Assumption

Let

$$T_d := \max_{i\in\mathcal{N}, j\in\mathcal{N}_i} \sup_{t\geq 1} \left\{ t - D_{i,j}^{(t)} \right\}$$

denote the worst-case communication delay, and assume that T_d is bounded; that is, $T_d < \infty$.

Result for ℓ_∞ Contractions

Theorem

The tracking error $\|\mathbf{x}^{(t)} - \mathbf{x}^{(*,t)}\|_\infty$ *can be asymptotically bounded as:*

$$\limsup_{t\to\infty} \|\mathbf{x}^{(t)} - \mathbf{x}^{(*,t)}\|_\infty \leq \frac{e_f + \sigma(1 + LT_d)}{1 - L}.$$

Result for ℓ_2 Contractions

Theorem

Let

$$N_d := \sup_{t \geq 1} \max_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \mathbb{I}\{D_{i,j}^{(t)} < t\} \in [0, m-1] \tag{2}$$

denote the maximum number of variables that are outdated at any given time step at any node. If, in addition, $L < \frac{1}{\sqrt{N_d+1}}$*, then*

$$\limsup_{t \to \infty} \|\mathbf{x}^{(t)} - \mathbf{x}^{(*,t)}\|_2 \leq \frac{e_f + \sigma(1 + L\sqrt{N_d+1}\,T_d)}{1 - L\sqrt{N_d+1}}. \tag{3}$$

Applications

- Time-varying optimization with feedback
- Multi-area load-flow in power network

Time-Varying Optimization with Feedback

Consider regularized time-varying optimization

$$\min_{\mathbf{x} \in \mathcal{X}^{(t)}} g^{(t)}(\mathbf{x}) + \frac{\eta}{2} \|\mathbf{x}\|_2^2$$

- $\mathcal{X}^{(t)}$ convex set
- $g^{(t)}(\mathbf{x})$ convex, with M-Lipschitz gradient, but not necessarily strongly convex

Time-Varying Optimization with Feedback

Consider regularized time-varying optimization

$$\min_{\mathbf{x}\in\mathcal{X}^{(t)}} g^{(t)}(\mathbf{x}) + \frac{\eta}{2}\|\mathbf{x}\|_2^2$$

- $\mathcal{X}^{(t)}$ convex set
- $g^{(t)}(\mathbf{x})$ convex, with M-Lipschitz gradient, but not necessarily strongly convex

Online gradient descent:

$$\mathbf{x}^{(t+1)} := \underbrace{\mathrm{proj}_{\mathcal{X}^{(t)}}\left\{\mathbf{x}^{(t)} - \alpha\left(\nabla g^{(t)}(\mathbf{x}^{(t)}) + \eta\mathbf{x}^{(t)}\right)\right\}}_{f^{(t)}(\mathbf{x}^{(t)})},$$

$f^{(t)}(\cdot)$ is Lipschitz in ℓ_2 norm with constant
$L := \max\{|1-\alpha\eta|, |1-\alpha(M+\eta)|\}$

Time-Varying Optimization with Feedback

Online gradient descent with feedback:

$$\mathbf{x}^{(t+1)} := \underbrace{\mathrm{proj}_{\mathcal{X}^{(t)}} \left\{ \mathbf{x}^{(t)} - \alpha \hat{\mathbf{g}}^{(t)} \right\}}_{\tilde{f}^{(t)}(\mathbf{x}^{(t)})},$$

where $\hat{\mathbf{g}}^{(t)}$ is an estimate or measurement of $\nabla g^{(t)}(\mathbf{x}^{(t)}) + \eta \mathbf{x}^{(t)}$.

Time-Varying Optimization with Feedback

Online gradient descent with feedback:

$$\mathbf{x}^{(t+1)} := \underbrace{\mathrm{proj}_{\mathcal{X}^{(t)}}\left\{\mathbf{x}^{(t)} - \alpha\hat{\mathbf{g}}^{(t)}\right\}}_{\tilde{f}^{(t)}(\mathbf{x}^{(t)})},$$

where $\hat{\mathbf{g}}^{(t)}$ is an estimate or measurement of $\nabla g^{(t)}(\mathbf{x}^{(t)}) + \eta\mathbf{x}^{(t)}$.

To satisfy the condition

$$L := \max\{|1-\alpha\eta|, |1-\alpha(M+\eta)|\} < \frac{1}{\sqrt{N_d+1}},$$

the regularization parameter have to be lower bounded as

$$\eta > \frac{\sqrt{N_d+1}-1}{2}M. \tag{4}$$

Multi-Area Load Flow

The load-flow problem of power network with n PQ-buses and one slack bus, cast as a fixed-point problem:

$$\mathbf{v} = h(\mathbf{v}, \mathbf{s})$$

- $\mathbf{v} \in \mathbb{R}^{2n}$ collects the voltage phasors
- $\mathbf{s} \in \mathbb{R}^{2n}$ collects the active and reactive power injections

Multi-Area Load Flow

The load-flow problem of power network with n PQ-buses and one slack bus, cast as a fixed-point problem:

$$\mathbf{v} = h(\mathbf{v}, \mathbf{s})$$

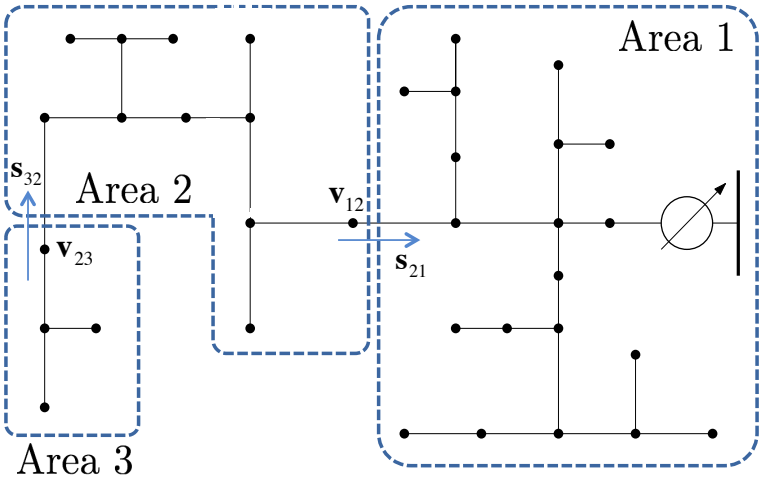
- $\mathbf{v} \in \mathbb{R}^{2n}$ collects the voltage phasors
- $\mathbf{s} \in \mathbb{R}^{2n}$ collects the active and reactive power injections

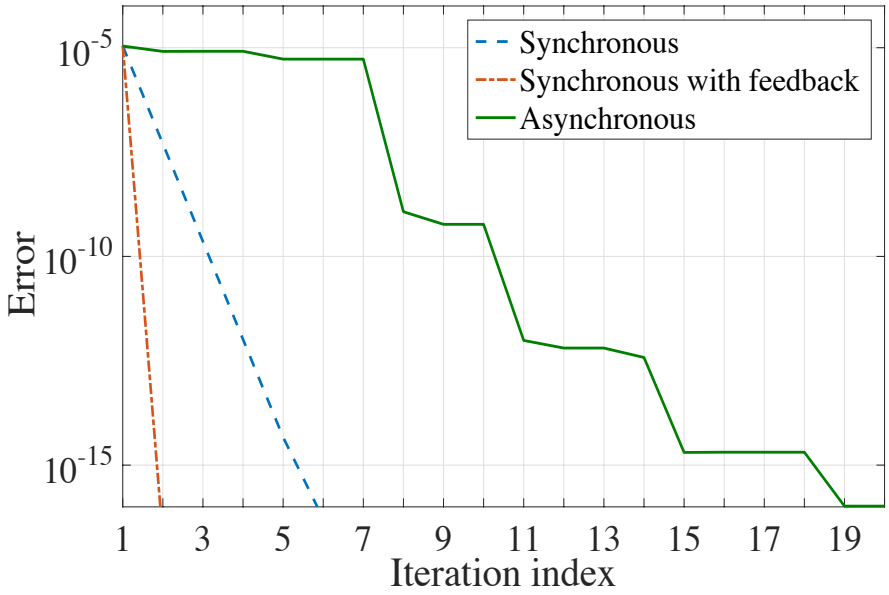
Time-varying load-flow: tracking the solutions of

$$\mathbf{v} = \underbrace{h(\mathbf{v}, \mathbf{s}^{(t)})}_{f^{(t)}(\mathbf{v})}, \; t \in \mathbb{N}.$$

- was recently shown that the map $h(\cdot, \mathbf{s})$ is a contraction and self-map in the ℓ_∞ norm, on some (proper) subset $\mathcal{D}$ of $\mathbb{R}^{2n}$

Multi-Area Load Flow



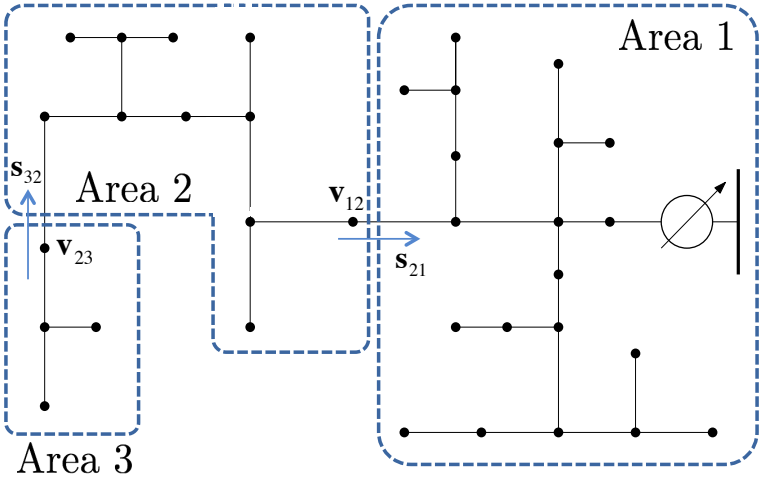


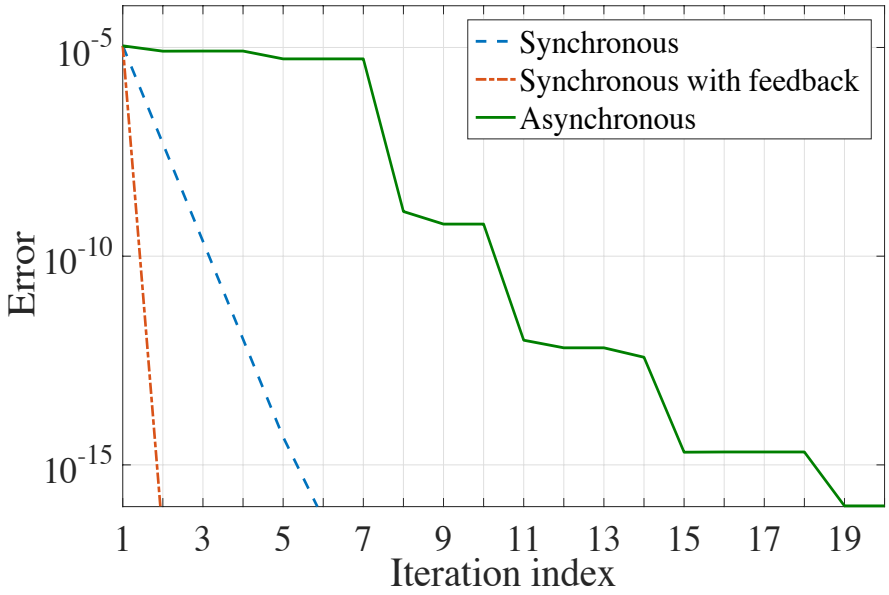
$$\mathbf{v}_1 = h_1(\mathbf{v}_1, \mathbf{s}_1^{(t)}, \overbrace{g_2(\mathbf{v}_2, \mathbf{v}_{12})}^{\mathbf{s}_{21}})$$

$$\mathbf{v}_2 = h_2(\mathbf{v}_2, \mathbf{v}_{12}, \mathbf{s}_2^{(t)}, \overbrace{g_3(\mathbf{v}_3, \mathbf{v}_{23})}^{\mathbf{s}_{32}})$$

$$\mathbf{v}_3 = h_2(\mathbf{v}_3, \mathbf{v}_{23}, \mathbf{s}_3^{(t)}),$$

Multi-Area Load Flow





Feedback-based load flow mappings:

$$\tilde{f}_1^{(t)}(\mathbf{v}_1) := h_1(\mathbf{v}_1, \mathbf{s}_1^{(t)}, \mathbf{s}_{21}^{(t)})$$
$$\tilde{f}_2^{(t)}(\mathbf{v}_1, \mathbf{v}_2) := h_2(\mathbf{v}_2, \mathbf{v}_{12}, \mathbf{s}_2^{(t)}, \mathbf{s}_{32}^{(t)})$$
$$\tilde{f}_3^{(t)}(\mathbf{v}_2, \mathbf{v}_3) := h_3(\mathbf{v}_3, \mathbf{v}_{23}, \mathbf{s}_3^{(t)})$$

Preview of Results for Averaged Operators[1]

Let $f : \mathbb{R}^m \to \mathbb{R}^m$ be a non-expansive mapping:

$$\|f(\mathbf{x}) - f(\mathbf{x}')\| \leq \|\mathbf{x} - \mathbf{x}'\|, \quad \forall \mathbf{x}, \mathbf{x}' \in \mathbb{R}^m.$$

[1] Dall'Anese, Simonetto, and Bernstein, submitted to LCSS and CDC 2019

Preview of Results for Averaged Operators[1]

Let $f : \mathbb{R}^m \to \mathbb{R}^m$ be a non-expansive mapping:

$$\|f(\mathbf{x}) - f(\mathbf{x}')\| \leq \|\mathbf{x} - \mathbf{x}'\|, \quad \forall \mathbf{x}, \mathbf{x}' \in \mathbb{R}^m.$$

For $\alpha \in (0, 1]$, define the α-averaged mapping as:

$$g_\alpha(\mathbf{x}) := (1 - \alpha)\mathbf{x} + \alpha f(\mathbf{x})$$

[1] Dall'Anese, Simonetto, and Bernstein, submitted to LCSS and CDC 2019

Preview of Results for Averaged Operators[1]

Let $f : \mathbb{R}^m \to \mathbb{R}^m$ be a non-expansive mapping:

$$\|f(\mathbf{x}) - f(\mathbf{x}')\| \leq \|\mathbf{x} - \mathbf{x}'\|, \quad \forall \mathbf{x}, \mathbf{x}' \in \mathbb{R}^m.$$

For $\alpha \in (0, 1]$, define the α-averaged mapping as:

$$g_\alpha(\mathbf{x}) := (1 - \alpha)\mathbf{x} + \alpha f(\mathbf{x})$$

Running Mann-Krasnosel'skiĭ iteration:

$$\boxed{\mathbf{x}^{(t+1)} = g_{\alpha_t}^{(t)}(\mathbf{x}^{(t)}) = (1 - \alpha_t)\mathbf{x}^{(t)} + \alpha_t f^{(t)}(\mathbf{x}^{(t)})}$$

Theorem

$$\frac{1}{T}\sum_{t=1}^{T} \alpha_t(1 - \alpha_t)\|\mathbf{x}^{(t)} - f^{(t)}(\mathbf{x}^{(t)})\|^2 \leq \frac{1}{T}\|\mathbf{x}_1 - \mathbf{x}_1^\star\|^2 + r$$

[1] Dall'Anese, Simonetto, and Bernstein, submitted to LCSS and CDC 2019

Conclusion

- Algorithmic framework for tracking fixed points of time-varying contraction mappings
- Tracking results for asynchronous implementation
- Future directions:
  - larger class of mappings
  - a more general asynchronous setting with non-homogeneous update rates.
  - local contraction, corresponding to tracking solutions of non-convex optimization

Thank you

www.nrel.gov

NREL/PR-5D00-74796

This work was authored in part by the National Renewable Energy Laboratory (NREL), operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. This work was supported by the Laboratory Directed Research and Development (LDRD) Program at NREL. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

NREL
Transforming ENERGY