Title
Advanced Metering Infrastructure Data Driven Phase Identification in Smart Grid

Authors
Wang, W.
Yu, N.

Publication Date
2017-09-10

Peer reviewed

GREEN 2017 - The Second International Conference on Green Communications, Computing and Technologies

4 authors, including:

Carlos Becker Westphall
Federal University of Santa Catarina

William Campbell
Birmingham City University

GREEN 2017

The Second International Conference on Green Communications, Computing and Technologies

ISBN: 978-1-61208-588-3

September 10 - 14, 2017

Rome, Italy

GREEN 2017 Editors

Carlos Becker Westphall, Federal University of Santa Catarina, Brazil
Steffen Fries, Siemens AG, Germany
William Campbell, Birmingham City University, UK
Nanpeng Yu, University of California, Riverside, USA

GREEN 2017

Forward

The Second International Conference on Green Communications, Computing and Technologies (GREEN 2017), held between September 10-14, 2017 in Rome, continued the inaugural event focusing on current solutions, stringent requirements for further development, and evaluations of potential directions. The event targeted to bring together academia, research institutes, and industries working towards green solutions.

Expected economic, environmental and society wellbeing impact of green computing and communications technologies led to important research and solutions achievements in recent years. Environmental sustainability, high-energy efficiency, diversity of energy sources, renewable energy resources contributed to new paradigms and technologies for green computing and communication.

Economic metrics and social acceptability are still under scrutiny, despite the fact that many solutions, technologies and products are available. Deployment at large scale and a long term evaluation of benefits are under way in different areas where dedicated solutions are applied.

The conference had the following tracks:

- Improving Green-ness
- Smart Energy and Smart Grid

We take here the opportunity to warmly thank all the members of the GREEN 2017 technical program committee, as well as all the reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and effort to contribute to GREEN 2017. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

We also gratefully thank the members of the GREEN 2017 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope that GREEN 2017 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in the field of green communications, computing and technology. We also hope that Rome, Italy provided a pleasant environment during the conference and everyone found some time to enjoy the historic charm of the city.

GREEN 2017 Chairs

GREEN Steering Committee

Carlos Becker Westphall, Federal University of Santa Catarina, Brazil
Wilfried Elmenreich, University of Klagenfurt, Austria

Bo Nørregaard Jørgensen, University of Southern Denmark, Denmark
Coral Calero, University of Castilla-La Mancha, Spain

GREEN Research/Industry Committee

Steffen Fries, Siemens AG, Germany
Daisuke Mashima, Advanced Digital Sciences Cener, Singapore
Nanpeng Yu, University of California, Riverside, USA
Enrique Romero-Cadaval, University of Extremadura, Spain

GREEN 2017
Committee

GREEN Steering Committee

Carlos Becker Westphall, Federal University of Santa Catarina, Brazil
Wilfried Elmenreich, University of Klagenfurt, Austria
Bo Nørregaard Jørgensen, University of Southern Denmark, Denmark
Coral Calero, University of Castilla-La Mancha, Spain

GREEN Research/Industry Committee

Steffen Fries, Siemens AG, Germany
Daisuke Mashima, Advanced Digital Sciences Cener, Singapore
Nanpeng Yu, University of California, Riverside, USA
Enrique Romero-Cadaval, University of Extremadura, Spain

GREEN 2017 Technical Program Committee

Kouzou Abdellah, Djelfa University, Algeria
Naji Abdenouri, Université Cadi Ayyad, Morocco
Carlos Becker Westphall, Federal University of Santa Catarina, Brazil
Rachid Benchrifa, Mohammed V University, Morocco
Abdelhak Bouchakour, Applied research unit in renewable energy, Ghardaia, Algeria
Mohsine Bouya, Rabat International University, Morocco
Coral Calero, University of Castilla-La Mancha, Spain
William Campbell, Birmingham City University, UK
Massimo Canonico, University of Piemonte Orientale, Italy
Fernando Jose Castor de Lima Filho, Federal University of Pernambuco, Brazil
David (Bong Jun) Choi**,** SUNY Korea / Stony Brook University, Korea
Wilfried Elmenreich, University of Klagenfurt, Austria
Youssef Errami, Chouaïb Doukkali University, El Jadida, Morocco
Steffen Fries, Siemens AG, Germany
Saurabh Garg, University of Tasmania, Australia
Chris Gniady, University of Arizona, USA
Marco Guazzone, University of Piemonte Orientale, Italy
Poria Hasanpor, KTH Royal Institute of Technology, Sweden
Khalil Kassmi, Mohamed Premier University, Morocco
Sedef Akinli Kocak, Ryerson University, Canada
Mohamed Latrach, University of Rennes 1, France
Daisuke Mashima, Advanced Digital Sciences Cener, Singapore
Mª Ángeles Moraga, University of Castilla-La Mancha, Spain
Masayuki Murata, Osaka University Suita, Japan

Copyright Information

Table of Contents

Development of a Solar Furnace with High Insulating Properties Using Date Palm Waste

Rachida Oaddi, Rachid Tiskatine, Mohamed Boulaid, Lahcen Bammou, Ahmed Aharoune, Ahmed Ihlal
Material and Renewable Energy Laboratory, Thermodynamics and Energetic Laboratory
Department of Physics, Faculty of Sciences Ibn Zohr University
Agadir, Morocco
e-mail: rachidaoaddi@gmail.com, tiskatinerachid@gmail.com, mohamed.boulaid@ed.uiz.ac.ma, lahcen.bammou@gmail.com, aharoune@gmail.com, a.ihlal@uiz.ac.ma

*Abstract***—This paper reports the results of a study on thermo-physical properties of two varieties of local date palm wood, called petiole, namely Boufeggous and Hafsa from Tinghir oasis, southern Morocco. The goal is to use this natural material as insulation for a solar furnace to reduce heat loss. Experimental measurements of thermo-physical properties, according to the orientation of the fibers at ambient temperature and atmospheric pressure, have been conducted and analyzed. Furthermore, a scanning microscopy (SEM) analysis of the samples was investigated to characterize their microstructure. Preliminary results deducted from this study were compared with other insulation materials in literature in order to evaluate the interest of these kind of materials for solar cooking application.**

Keywords-Solar furnace; date palm; insulation and efficiency

I. INTRODUCTION

In many thermodynamic processes, the heat flow (gain or loss) can be optimized using special materials to insulate the system and, ultimately, save energy. In solar cooker applications, numerous materials with low thermal conductivity may be used for insulation. However, to ensure optimal insulation, we have to take into account the thermo-physical properties, environmental and economic impact, as well as availability and durability of the used material [1]. At present, there are a number of different insulating materials used for solar furnace manufactured from fiberglass, mineral wool (rock wool), cellulose or polystyrene (expanded) [2]. Although these materials have good physical properties, they are very expensive to acquire and they can be unsafe to human health and to the environment. For instance, the fiberglass could irritate the eyes, skin and the respiratory system. These disadvantages have necessitated research on natural, ecological and economic insulation materials such as flax, cotton, hemp, jute, sisal, kenaf, pineapple, ramie, bamboo, banana, palm etc. The performance of these materials is under study and their development is at an early stage [2][3]. Several authors have analyzed and characterized experimentally some of these natural materials, as described in the rest of this section.

Nguyen et al. [4] studied the thermal performance of hemp shives. They developed a multi-scale homogenization approach that takes into account the shape and orientation of pores and particles in order to predict the thermal performance of hemp as an insulation material. As results, they found that the thermal conductivity of hump increased linearly with increasing saturation degree, density and temperature. Tangjuank [1] presents the thermal properties of insulation material produced from pineapple leaves. As blinder, he used natural rubber latex. According to the results, the thermal conductivity obtained was close to the commercial insulator, with a value of 0.035W/m.K and with density of 210 Kg/m^3. This value ensures that pineapple fibrous material can replace the synthesis insulator.

Agoujdil et al. [5] have carried out an experimental investigation on the thermo-physical, chemical and dielectric properties of three varieties of date palm wood. According to the results, they noticed that date palm wood could be a good candidate as insulating material, in order to use it to reduce building heat loss. Oushabi et al. [2] analyzed a local date palm waste from Errachidia oasis in Morocco, in order to use it as a thermal insulation material in the vessel of refrigerator, cooler and food flask. Their results showed that this material had good thermal properties compared with other synthetic materials.

Djoudi et al. [6] have performed an experimental study and modeling of the effect of date palm fibers addition on thermal properties of plaster concrete. They reported that the thermal conductivity and density of this composite decrease as the fraction of fibers increases. The same results are found by Braiek et al. [7], who have analyzed the thermo physical properties of date palm/gypsum composite, in order to use it as insulating material in building.

The main goal of this paper is to evaluate the possibility of using the date palm wood waste (Phoenix Dactylifera), including petiole as an insulating material to reduce the heat loss in the solar box cooker. No study up to date has been conducted in terms of using the petiole of date palm as insulating material in solar furnace application. This paper is organized as following. Section 2 provides a description of different experimental measurements and samples preparation. The results are interpreted, discussed and compared with some conventional material described in literature in Section 3. We conclude the paper in Section 4.

 ISBN: 978-1-61208-588-3

II. MATERIAL AND METHODOLOGY

A. *Experimental set up*

1) *Solar cooker description*

In the experimental tests, a solar box cooker was used (Figure 1). It consists of the following main components: cover double glazed glass, three reflectors placed on the outer cover of the cooker, thermal insulator placed in the lateral part of the solar cooker and rocks bed for heat storage. The measured parameters are the ambient temperature, temperature of rocks and temperature inside the solar cooker. Type-k thermocouples were used for this purpose.

Figure 1. Solar furnace with three external reflectors.

The main problem this prototype faces is heat loss through the furnace walls as evidenced by thermographic analyses.

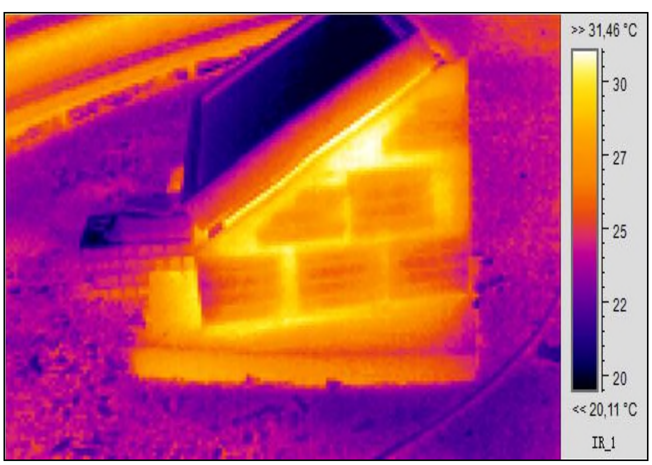


Figure 2. Infrared picture showing heat loss through the furnace walls.

Initially, we have tested glass wool as insulation material and we found out that the heat losses remain high, as we can conclude from Figure 2. Thus, we decided to go for date palm wood as insulation due to its availability, low economic and environmental impact.

B. *Samples preparation*

The natural materials used in this research are from two varieties of local date palm wood called petiole, namely Hafsa and Boufeggous, from Tinghir oasis, southern Morocco.

Two configurations of petiole (P), according to the orientation of fibers, were studied. Figure 3 shows these configurations.

Figure 3. Samples cut out from the petiole (a) petiole sample I longitudinal direction of fibers, (b) in transversal direction of fibers.

C. *Thermal conductivity and diffusivity*

All thermo-physical measurements of date palm samples at room temperature were determined using a Thermal Analyzer TPS 1500 (Figure 4). Transient plane source (TPS), or Hot-Disc method, is highly appreciated technique for measuring thermal properties of materials from a single measurement, with minimum sample preparation [8]. The results are displayed directly on the device screen.

Figure 4. Thermal properties test measurements.

 ISBN: 978-1-61208-588-3

D. *Morphological analysis*

Microscopic examinations of the samples were carried out using a TESCAN VEGA3 LM scanning electron microscope (SEM) in order to analyze the morphology of these samples.

III. RESULTS AND DISCUSSION

A. *Thermo-physical properties*

Thermal conductivity (*k*) is defined as the ability of a material to conduct heat. This parameter is tremendously important to evaluate an insulating material. The results of the thermal conductivity measurements at room temperature are provided in Table 1. The mean value of this parameter of the samples studied is about k = 0.076W/m.K at room temperature. This value is close to or lower compared to the thermal conductivity of other natural insulating materials, for example sisal (k = 0.070W/m.K), banana (k = 0.117 W/m.K), and hemp (0.115 W/m.K).

TABLE I. THERMAL CONDUCTIVITY OF PALM PETIOLE, SISAL, HEMP AND BANANA.

Sample	Thermal conductivity (W/m.K)	Reference
PLH (Hafsa petiole sample in longitudinal direction of fibers)	0.0736 ± 0.001	Present study
PTH (Hafsa petiole sample in transversal direction of fibers)	0.0670 ± 0.001	Present study
PTB (Boufeggous petiole sample in transversal direction of fibers)	0.0893 ± 0.001	Present study
Sisal	0.070	[9]
Hemp	0.115	[10]
Banana	0.117	[11]

For petiole of Hafsa variety, two types of measurements were performed according to the orientation of fibers. Figure 5 shows that orientation of fibers has a weak effect on thermal conductivity and diffusivity. Indeed, the number of fibers is much less in the longitudinal direction than in the transversal direction and, consequently, there should be more thermal resistance across the axis. Therefore, the orientation of the fiber should have a significant effect on the thermo-physical properties of this kind of natural materials. And yet, this behavior was not observed in this work neither in some other similar studies [12].

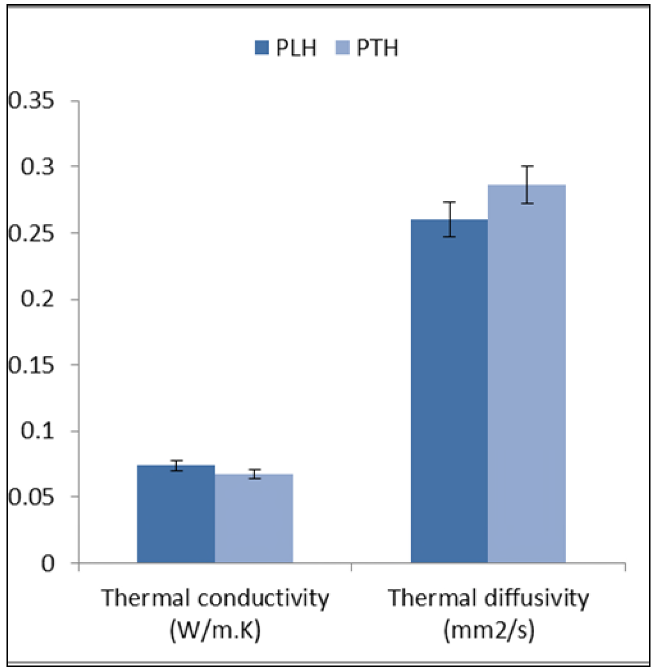
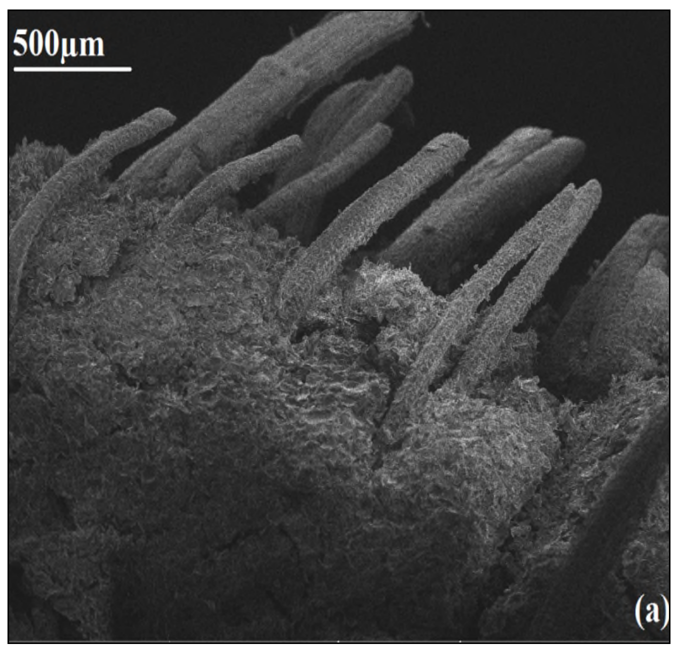


Figure 5. Thermal conductivity and diffusivity of Hafsa petiole samples measured at atmospheric pressure.

B. *Structure and morphology*

Figure 6 presents SEM images of a typical sample of Hafsa petiole in transversal direction of fibers. Observing these microstructures, it can be seen that the sample contains cylindrical fibers with irregular and rough surface containing many impurities. Likewise, the morphology of petiole fibers of date palms is similar to those of coir fiber [13] [14].



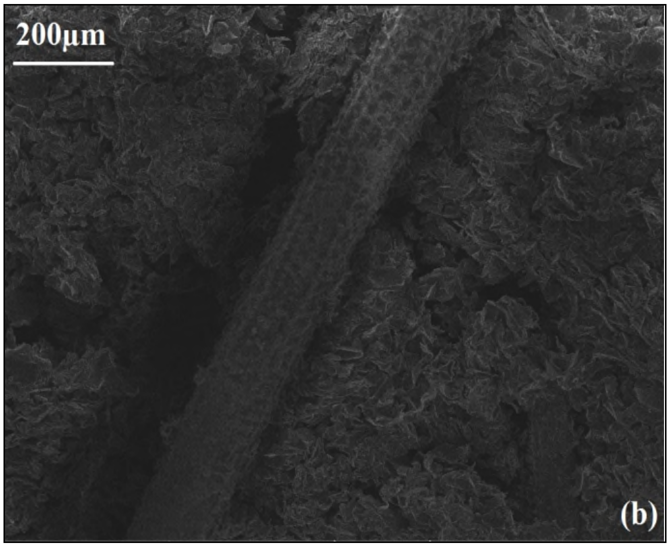


Figure 6. SEM images of Hafsa petiole sample in transversal direction of fibers (PTH). 500µm (a) and 200µm (b).

This kind of natural fiber has a cylindrical and irregular form with many filaments and cells.

IV. CONCLUSION

This study presents the results of experimental measurements and tests conducted on two varieties of date palm material, from Tinghir oasis, south in Morocco. The aim of this research was to evaluate some thermo-physical properties and investigate the possibility to use this local material in building solar furnace insulation.

The findings from this study reveal low thermal conductivity compared to other conventional materials used in this field.

In perspective, for more accurate results and recommendations, other tests will be carried out. The insulation efficiency will also be tested by manufacturing thermal insulation based on these candidate materials and using it as an insulating material in the solar furnace.

ACKNOWLEDGMENTS

This study was supported by The Moroccan Ministry of Higher Education and Research in the framework of (PPR/2015/31) project.

REFERENCES

[1] S. Tangjuank, "Thermal insulation and physical properties of particleboards from pineapple leaves," International Journal of Physical Sciences, vol. 6, no. 19, 2011, pp. 4528–4532.

[2] A. Oushabi, S. Sair, Y. Abboud, O. Tanane, and A. EL Bouari, "Natural thermal-insulation materials composed of renewable resources: characterization of local date palm fibers (LDPE)," J.Mater. Environ. Sci., vol. 6, no. 12, 2015, pp. 3395–3402.

[3] F. Asdrubali, F. D'Alessandro, and S. Schiavoni, "A review of unconventional sustainable building insulation materials," Sustainable Materials and Technologies, vol. 4, 2015, pp. 1–17.

[4] S.T. Nguyen et al., "Modeling thermal conductivity of hemp insulation material: A multiscale homogenization approach, " Building and Environment, vol. 107, 2016, pp. 127-134

[5] B. Agoudjil, A. Benchabane, A. Boudenne, L. Ibos, and M. Fois, "Renewable materials to reduce building heat loss : Characterisation of date palm wood, " Energy and Buildings, vol. 43, 2011, pp. 491–497.

[6] A. Djoudi, M. Khenfer, A. Bali, and T. Bouziani, "Effect of the addition of date palm fibers on thermal properties of plaster concrete: experimental study and modeling, " J. Adhes. Sci. Technol, vol. 28, no. 20, 2014, pp. 2100–2111.

[7] A. Braiek, M. Karkri, A. Adili, L. Ibos, and S. Ben Nasrallah, "Estimation of the thermophysical properties of date palm fibers/gypsum composite for use as insulating materials in building," Energy Build, vol. 140, 2017, pp. 268–279.

[8] P. Krupa and S. Malinaric, "Using the Transient Plane Source Method for Measuring Thermal Parameters of Electroceramics," International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering, vol. 8, No. 5, 2014, pp. 735–740.

[9] G. Kalaprasad, P. Pradeep, G. Mathew, C. Pavithran, and S. Thomas, "Thermal conductivity and thermal diffusivity analyses of low-density polyethylene composites reinforced with sisal, glass and intimately mixed sisal/glass fibers, " Composites Science and Technology, vol. 60, no. 16, 2000, pp. 2967–2977.

[10] T. Behzad and M. Sain, "Measurement and prediction of thermal conductivity for hemp fiber reinforced composites, " Polymer Engineering and Science, vol. 47, no. 7, 2007, pp. 977–983.

[11] S.A. Paul et al., "Effect of fiber loading and chemical treatments on thermophysical properties of banana fiber/polypropylene commingled composite materials," Composites Part A-Applied Science and Manufacturing, vol. 39, no. 9, 2008, pp. 1582–1588.

[12] F.A. Al-Sulaiman, "Date palm fibre reinforced composite as a new insulating, " International Jounrnal of Energy Research, vol. 27, no.14, 2003, pp. 1293–1297.

[13] M.J. John and R.D. Anandjiwala, "Recent developments in chemical modification and characterization of natural fiber-reinforced composites," Polymer Composites, vol. 29, no. 2, 2008, pp. 187–207.

[14] A. Kriker, G. Debicki, A. Bali, M. Khenfer, and M. Chabannet, "Mechanical properties of date palm fibres and concrete reinforced with date palm fibres in hot-dry climate, " Cement and Concrete Composites, vol. 27, no. 5, 2005, pp. 554–564.

 ISBN: 978-1-61208-588-3

An Exploration of the Impact of the Use of Standard Management Models on the Adoption of Green IT

William M. Campbell
School of Computing and Digital Technology
Birmingham City University
Birmingham, UK
B4 7XG
Email: william.campbell@bcu.ac.uk

Abstract—**This paper explores the extent to which senior managers using standard management models as tools for developing corporate strategy, structures and culture are likely to be encouraged to adopt green IT. A range of standard management models are considered: strategic, tactical and operational. Analysis reveals that many standard models, in particular older ones that rely heavily on numbers and take a narrow view of corporate responsibility, are not favourable to the adoption of green IT. Accordingly, managers need to avoid excessive reliance on such models and should consider using models which take account of softer issues, in particular those models which address sustainability directly. There is a need for the development of new management models, which more explicitly integrate traditional bottom line considerations with the wider ethical responsibilities of companies, including sustainability.**

Keywords–Green, Sustainability, Information Technology, Organizational Culture, Management Models.

I. Introduction

The sustainable use of resources is a key issue facing the human race. It is widely accepted that the emission of Greenhouse gases has affected the climate. Other issues include pollution and the careless disposal of waste.

Information Technology makes a major contribution to Greenhouse gas emissions, producing around 2% of global carbon dioxide emissions However, IT can also contribute to the reduction of pollution through technologies such as "smart cities" and environmental monitoring systems.

There has been pressure on individual companies to take note of environmental issues [1]. This has come not only from the need to comply with environmental legislation, but also from consumer pressure and concern about reputation. Many companies now accept that economic performance is not the only measure of success and have adopted a "Triple Bottom Line" of environment, society and economic performance [2] [3].

In determining corporate strategy and organizational structures senior managers often seek guidance from the standard management models taught in business schools. The extent to which these models encourage the adoption of green IT will, therefore, have an effect on the extent to which managers regard green IT as a serious, mainstream issue.

The remainder of the paper is structured as follows: Section 2 looks at the green agenda, focusing in particular on green IT. Section 3 explores management and organizational models which specifically address green IT. Section 4 investigates the extent to which standard management models are favourable to green IT. Finally, there are some concluding observations.

II. The Green Agenda

The definition of sustainability provided by the Brundtland Commission has gained widespread acceptance: "Development that meets the needs of the present without compromising the ability of future generations to meet their needs"[4]. There has been a number of agreements, most recently the Paris Agreement in 2016. Its central aim was to strengthen the global response to the threat of climate change, by keeping the global temperature rise this century well below 2 degrees Celsius above pre-industrial levels and to pursue efforts to limit the temperature increase even further to 1.5 degrees Celsius [5].

Jenkin et al.[6] distinguish between "Green IT" and "Green IS". "Green IT" is the attempt to reduce energy consumption and waste associated with the use of both hardware and software. "Green IS" they define as the use of information systems to support environmental sustainability initiatives, as in "smart cities". Here we use "Green IT" as a generic term, covering both efforts to reduce the environmental damage caused by the use of IT and the use of IT in a positive way to support sustainability objectives.

IT has played an increasingly important role in industry and commerce and makes a substantial contribution to the environmental footprints of companies, through both the use of IT and the construction and disposal of IT equipment [7]. It is estimated that IT is responsible for around 2% of worldwide carbon dioxide emissions [8]. Energy and resources are consumed throughout the IT lifecycle. Furthermore, the Basel Action Network estimates that 80% of electronic waste is sent for recycling to the developing world [9]. Computing equipment contains highly toxic materials such as cadmium.

A number of national and international laws have been introduced to tackle this issue. The European Union Waste and Electronic Equipment (WEEE) directive (2003) requires producers, importers and resellers of electronic equipment to dispose of, refurbish or recycle equipment in an environmentally sound manner. The Japanese Home Electronics Recycling Law (1998) has similar requirements. Sustainability issues should be considered at all stages of the software lifecycle.

Software as a Service (SAAS) and Cloud Computing offer ways for using IT resources more efficiently. Companies purchase data storage and rent software, as required, from external providers. These can be accessed using "thin client" computers.

However, the IT data centres which these technologies require have a major carbon footprint. It is estimated that data centres produce 150 million tonnes of carbon each year. Server virtualization has provided the opportunity for servers to be used more efficiently; this allows several servers to be consolidated as virtual servers on one physical server, enabling sharing of resources and economies of scale.

The application of IT can make a positive contribution to sustainability in various ways. Environmental information systems and "intelligent buildings" help to reduce energy wastage; supply chain information systems optimize routing and transportation [10]. Dao et al. [11] argue for combining IT resources with supply chain management and human resource management within an integrated sustainability framework,

III. Management and Organizational Models for Green IT

Bokolo et al. [12] provide a systematic and up-to-date review of literature on green IT. This illustrates that much effort, across a number of disciplines, has been put into developing models and frameworks for analysing green IT.

Murugesan and Gangadharan [13] divide enterprise green IT strategy into three approaches.

Tactical Incremental Approach. In this approach, the company retains the existing infrastructure and policies and introduces simple measures such as switching off computers when not in use.

Strategic Approach. In this approach, the company develops a comprehensive plan for making its deployment of IT more energy-efficient.

Companies following a *Deep Green Approach* go beyond the *Tactical Incremental Approach*, adopting additional measures such as a carbon offset policy to neutralize greenhouse gas emissions.

One of the mostly widely-cited models is Molla and Cooper's "Green IT Readiness" or "G-Readiness" framework. It divides IT into IT Managerial Capability, IT Human Capability and IT Technical Capability. An organization's green IT maturity is assessed in terms of attitude, policy, practice, technology and governance. There is an accompanying G-Readiness Survey instrument.

Deng and Ji undertook a review of the literature, seeking to identify the motivating factors for companies to adopt green IT [14]. They noted that the literature has "scattered theoretical foundations", but identified the following key underlying theories.

The *Diffusion of Innovation Theory* investigates the process by which innovations spread.

Institutional Theory analyses the pressures which influence the development of organizations. A key institutional pressure is "mimetic isomorphism", the tendency of companies to follow leading companies in their field.

Organizational Culture views organizations as social structures and examines the way shared assumptions and norms emerge. This is discussed later in the section on Cameron and Quinn's Competing Values Framework .

The *Resource Based View* (RBV) [15] takes the view that a company's competitive advantage resides in its ownership of a set of resources that are not easily duplicated by a competitor. These resources can be physical, organizational or social.

Hart [16] extends this to the *Natural Resource Based View* (NRBV), by including resources and capabilities particularly relating to sustainability.

Deng and Ji introduce a theoretical framework for "Organizational Green IT Adoption" (OGITA). This has the external drivers of technological context and institutional pressures; and internal drivers of senior management attitudes, corporate strategy and organizational culture.

However, senior managers looking for guidance on changing company strategy, structures and culture are likely to refer to standard management models. The extent to which these models "favour" green IT will, therefore, have a major impact on its adoption. We discuss this in the next section.

IV. Standard Management Models

Almost a third (31%) of the world's largest 500 companies have a chief executive with an MBA [17]. It is likely that the management models they studied will have influenced them in their later careers. We use a standard, widely used and influential book on management models [18]. We follow its separation of models into strategic, tactical and operational. In each case, we explore the extent to which managers employing these models are likely to be encouraged to adopt green IT.

A. Strategic Management Models

These models help a company to analyse its strategic position and develop strategic plans for the future.

1) Ansoff's Matrix: Ansoff's Matrix is a widely used model for helping companies determine their strategy for developing new products and entering new markets [19]. In terms of products, they would have a choice of retaining existing products or developing new products. In term of markets, they would have a choice of focusing on existing markets or developing new markets. This produces four top-level strategies, as illustrated in Figure 1. The top left quadrant is the "conservative" strategy of focusing on existing products and markets; the bottom right quadrant is the "aggressive" strategy of developing new products and seeking new markets.

The model has been extended to a cube, by introducing a geographical dimension, where companies consider expanding into new countries. This is illustrated in Figure 2.

The Ansoff model advises companies to consider four issues: competitive advantage, potential synergies across the company's core competencies, strategic flexibility (the ability easily to modify strategy to cope with unpredicted events), and the potential for geographical growth.

We now use the OGITA model discussed above to evaluate the extent to which use of the Ansoff Matrix would be likely to encourage companies to adopt green IT. We first consider external pressures.

From a technology perspective, the questions would be:

- Would going green give the company a relative technological advantage?
- Would it be technically challenging?
- Would it make use of core technical competencies within the company?

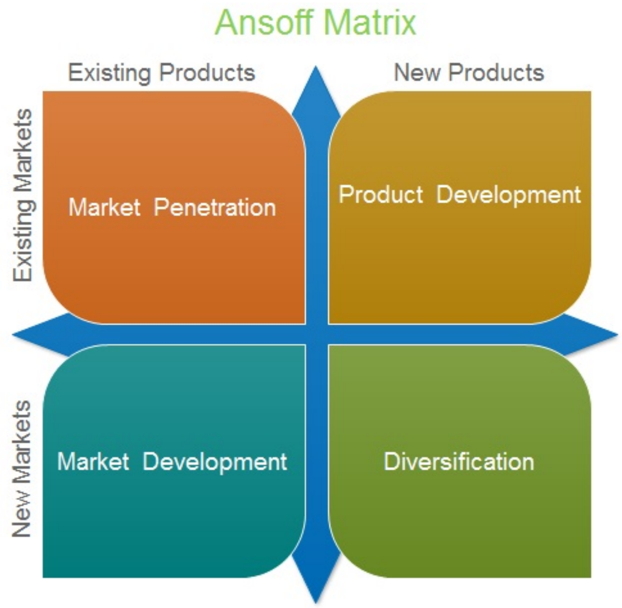


Figure 1. Ansoff Matrix

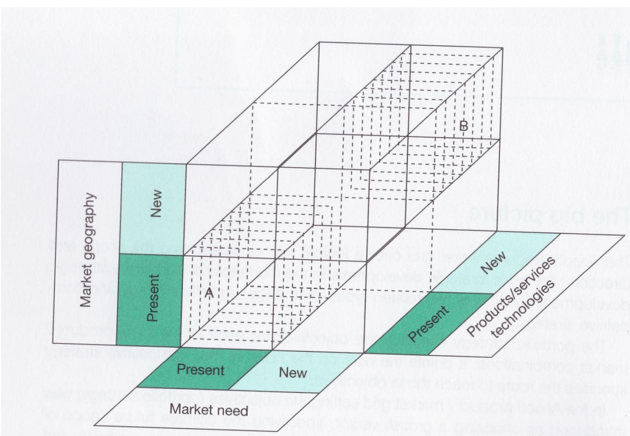


Figure 2. Ansoff Cube [18]

In considering these issues, technological experts within the company would be considering the challenges of developing new products, against the backdrop of the possibility of just going for greater market penetration in existing markets or developing new markets. Unless there was a compelling reason to suppose that the greener product would provide a competitive technical advantage or the existing product would become obsolete because of its poor green credentials, technology experts would be likely to favour avoiding radical changes to the existing product portfolio.

We noted above that there are essentially two types of green IT: those which try to avoid negative environmental impacts of IT-related products and those which use information systems to promote sustainability in applications such as environmental monitoring and smart cities. The latter are likely to involve developing radically new products and be much more challenging in technical terms. They are likely, therefore, to be deemed unattractive.

From the perspective of external institutional pressures, the questions would be:

- Will the company be breaking the law, if it does not make its products greener?
- Will the company become out of step with the market if it does not become greener?
- Does the company face a risk of reputational damage?

Unless the company is driven by a powerful "mimetic isomorphism" pressure, external pressures for greenness are unlikely to be stronger than economic pressures.

Finally, we consider the internal motivations of the OSITA framework. Senior managers tend to be driven by numbers and verifiable evidence. It is likely to be easier to provide clear evidence for the benefits of taking existing products into new markets than to demonstrate that a market will exist for radical new green products. Many green products are "disruptive technologies", for which there is currently no market. As Christensen argues in his influential work "The Innovator's Dilemma", the company culture is frequently hostile to such technologies [20]. Unless there are a number of green champions within the company at a senior level, top managers are likely to favour developing markets and making only incremental changes to existing products.

In summary, the Ansoff model is likely to discourage companies from developing new greener products, because it juxtaposes the challenge of developing radical new products with the easier option of expanding the market for existing products. Insofar as the use of Ansoff's Matrix encourages the adoption of green IT, it is likely to be of a "Tactical Incremental" nature, within Murugesan's taxonomy of green initiatives discussed above.

2) Porter's Five Forces: Porter's Five Forces is one of the most established management models, and has been used for around forty years. It is used by companies contemplating entering a new industry. It identifies five things that need to be considered:

- New entrants
- Substitutes (will it be easy to replace the proposed product with something else?)
- Buyers
- Suppliers (companies which will be below you in the supply chain)
- Existing Competitors

The employment of Porter's Five Forces is likely to discourage companies from developing radical new green products and services, for the same reason as Ansoff's Matrix. As Christenen (discussed above) notes, you cannot analyse a market that does not exist.

3) The BCG Matrix: The Boston Consulting Group Matrix goes back to the 1970s [21] [18]. It is used by companies for planning their product portfolio. It is similar to the Ansoff Matrix, having two dimensions; in this case, the dimensions are the projected Market Share and Market Growth. This again creates four main types of market:

1) high market share, high growth (best)
2) high market share, low growth
3) low market share, high growth

4) low market share, low growth (worst)

What "advice" will this model give? The market for a new green Cloud service is likely to be of the third type. The Cloud market is highly competitive but is likely to grow. The market for a new environmental monitoring system for reservoirs is likely to be of the second type. The market is small and unlikely to grow substantially, but it is a small market and a successful product could have reasonable expectations of dominating it. Few green markets are likely to be of the first type. It seems probable that senior decision makers using the BCG will favour potential new markets of the first type rather than green markets.

4) The Blue Ocean Strategy: This model makes a distinction between a Red Ocean Strategy, where a company seeks to beat the competition in an existing market; and a Blue Ocean Strategy, where a company seeks to develop a brand new market. It encourages companies to focus on the big picture rather than the numbers [22] [18].

Employment of the Blue Ocean model is likely to be positive for the development of green IT products for new applications, such as the Internet of Things.

5) Kay's Distinctive Capabilities: The Kay's Distinctive Capabilities (KDC) model originates from the Resource Based View, discussed above, which regards a company as a collection of skills and capacities, many intangible, which cannot easily be imitated. [23] [18]. KDC separates these into three categories:

- Architecture (features intrinsic to the company and its relationships with customers and suppliers)
- Reputation
- Capacity to innovate

To some extent this model encourages green innovation. It acknowledges the value of a company having a reputation for being ethical. Furthermore, the extension of the RBV discussed above, the Natural Resource Based View, explicitly recognizes that green capabilities are likely to be important in the future. But the model emphasises that it is very difficult to convert innovation into competitive advantage. The success of a radical new and efficient Cloud Computing model will be greatly affected by whether competitors are developing a similar product.

B. Tactical Management Models

These models help a company to organize its process, resources and people. They address "how to" questions.

1) Cameron and Quinn's Competing Values Framework: Anthropology takes the view that organizations are cultures; sociology takes the view that organizations have cultures [24]. Most organizational theory adopts the sociological perspective, regarding culture as an attribute of an organization that can be measured and analysed. Schein [25] defined organizational culture as: "A pattern of shared basic assumptions that the group learned as it solved its problems of external adaptation and internal integration that has worked well enough to be considered valid and hence to be taught to new members as the correct way to perceive, think and feel in relation to those problems."

Schein identified three levels of culture:

- Artifacts, those aspects which are on the surface such as dress and can be easily identified;
- Espoused Values, that is conscious goals, strategies and philosophies;
- Basic Assumptions and Values. These exist at a largely unconscious level, form the inner core of culture and are hard to identify.

Basic Assumptions and Values have the deepest influence and are the most difficult to change. Many attempts at organizational change fail because of a failure to change the underlying culture [26].

Many dimensions of organizational culture have been proposed, for example Hofstede [27]: power distance, uncertainty avoidance, individualism, and masculinity. Cameron and Quinn's "Competing Values Framework" (CVF) originated from a cluster analysis of these dimension schemes. It identifies two key dimensions: Internal Focus and Integration versus External Focus and Differentiation; and Stability and Control versus Flexibility and Discretion [28] [29]. The CVF has been used in many research studies and has been shown to have a high degree of validity [30].

The four key culture types identified by the CVF are illustrated in Figure 3 and may be summarized as follows (Adapted from [29]):

- Hierarchy. Such organizations tend to be bureaucratic. Formal rules and policies hold the organization together. The long-term goals of the organization are stability, predictability and efficiency, Government agencies and the military are typical hierarchical cultures.
- Market. The workplace is results-oriented. Leaders tend to be aggressive and demanding. The glue that holds the organization together is an emphasis on winning. Success is defined in terms of beating the competition and market share.
- Clan. The organization is held together by loyalty, tradition, and collaboration. It is a friendly place to work, where people share a lot of themselves. Leaders are thought of as mentors and coaches. Success is defined in terms of internal climate and concern for people. The organization places a premium on teamwork, participation, and consensus.
- Adhocracy The workplace is dynamic, creative, entrepreneurial and risk-oriented. The emphasis is on being at the leading edge of new knowledge, products, and/or services. The glue that holds the organization together is commitment to experimentation and innovation. Success is defined as the production of innovative and original products and services.

The Organizational Culture Assessment Tool (OCAI) consists of a questionnaire requiring employees to assess their organization, using an ipsative scale, on six characteristics: Dominant Characteristics, Organizational Leadership, Management of Employees, Organization Glue, Strategic Emphases and Criteria for Success. A culture profile diagram can then be produced.

The results can be used for various purposes, e.g.: to calculate the average profile of an organization and identify

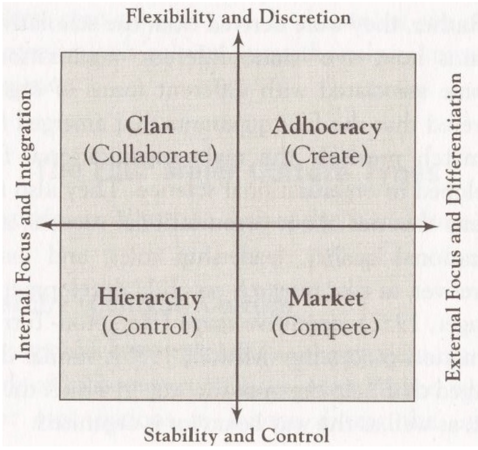


Figure 3. Cameron and Quinn [29]

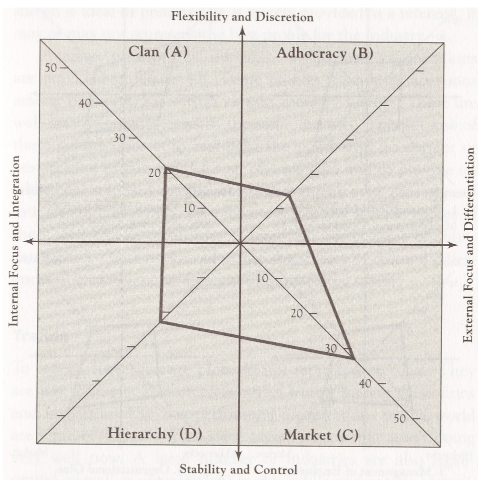


Figure 4. Cameron and Quinn [29]

the main culture types(s); to identify discrepancies between current and preferred culture; and to ascertain the degree of congruence between results produced by different groups of employees. Cameron and Quinn averaged the results for over one thousand companies; this resulted in the average profile in Figure 4.

There has been a considerable amount of research on the relationship between types of organizational culture and effectiveness. Richard et al. [31] conducted a survey of US firms. They found that clan cultures resulted in higher earnings and employee satisfaction.

In the US health industry, Gregory et al. [32] found a positive link between group (clan) culture and patient and physician satisfaction and also a slight link between balanced cultures and satisfaction.

The successful adoption and diffusion of green IT systems is also affected by the organizational culture of companies. Green IT systems are likely to be ‘disruptive technologies’, which are regarded as risky. For example, attempts to reduce energy use associated with data storage through the employment of “cloud computing” may raise fears about security. Green IT systems are, therefore, more likely to be favoured

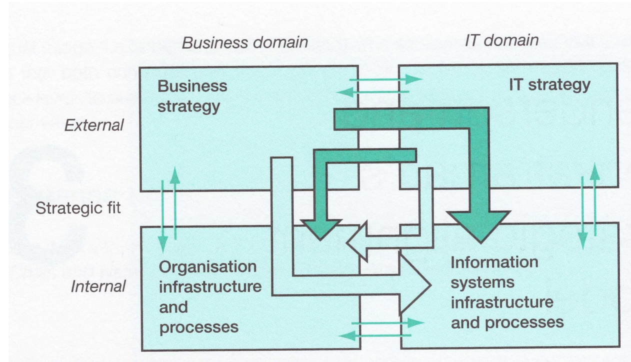


Figure 5. Strategic Alignment Model

by companies with clan or adhocracy cultures, which are non-hierarchical, entrepreneurial and can embrace change.

The use of the Cameron and Quinn model as a framework for discussing the impact of organizational culture on the adoption of green IT is discussed in detail in [33] [34].

2) Beer and Nohria E and O Theories: Beer and Nohria is a modern management model, which explicitly emphasises the value of soft skills and the importance of companies behaving ethically and taking account of their corporate social responsibility [35] [18].

They have two main theories of change:

- Theory E. This focuses on the creation of economic value for shareholders. It involves formal systems and structures. The decision making process is top-down. Changes are carefully planned.
- Theory O. This focuses on a culture that develops employee commitment and takes note of a company’s ethical responsibilities. Change is emergent.

To be successful, a company must embrace both Theory E and Theory O and confront the tension between them.

The Beer and Nohria model is favourable to the adoption of green IT, because it encourages managers and employees to think of the bigger picture and not just focus on narrow financial considerations. In particular, it asks companies to take account of their ethical responsibilities. But the model does not ignore the practical exigencies of operating a successful business. For companies successfully to adopt green IT they must both have a vision and have the operational capability to realise it in the real world of business. The Beer and Nohria model provides a framework for constructively reconciling the conflicting pressures this creates.

3) Henderson and Venkatram’s Strategic Alignment Model: This model addresses IT strategy directly [36] [18]. It seeks to promote alignment between business strategy and IT strategy and also between the IT infrastructure and business operations. A key feature of this model that it provides for IT strategy influencing business strategy. This is visualized by the counter-clockwise arrow from top-right to bottom left in Figure 5. This is likely to be favourable to the adoption of green IT.

C. Operational Management Models

These models help a company to optimize operational process and activities.

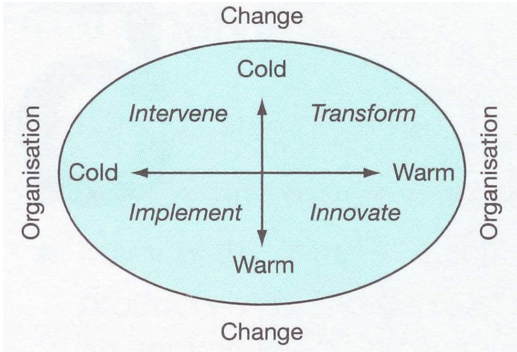


Figure 6. Change Quadrants

The *Change Quadrants* model is a tool to assist companies to effect a particular change [37] [18]. It analyses companies on two dimensions: whether they are "warm" or "cold"; and whether the key motivation for the proposed change is "warm" or "cold".

A warm organization is one where there is a shared sense of values and employees do not have a merely transactional relationship with the organization. It is rather like the "Clan Culture" in the Cameron and Quinn Competing Values model. A cold organization is one which is hierarchical and governed by rules, systems and procedures.

A warm motivation for a proposed change is driven by a shared sense of values across the company. A cold motivation is a response to a crisis such as the emergence of a dangerous competitor.

This produces the four quadrants in Figure 6. The change strategy should be tailored to the quadrant. A "warm organization that is willing" (the bottom right quadrant) will be open to change. It will be possible to develop a long-term vision bottom-up. A "cold organization that is obligated" (the top left quadrant) will have to drive change top-down; employees will only have a say in the implementational details. The key message of the model is that real transformation, such as is involved in the systematic adoption of green IT, requires a warm organization and a warm motivation for change.

V. General Conclusions about Management Models

Most of the older models are driven by relatively short-term bottom line considerations. These are likely to be unfavourable to green IT. More recent models, such as Beer and Nohria and Change Quadrants, tend to adopt a wider perspective on the responsibilities of companies and also take more note of "softer" people and ethical issues. They are more likely to be favourable towards green IT.

Managers need to be cautious about over-reliance on standard models, especially those which take a narrow view of corporate responsibilities. They should consider employing models which take account of wider issues, in particular those models which incorporate consideration of sustainability.

VI. Conclusion

This paper has considered the extent to which standard management models are likely to support the adoption of green IT. It explored strategic, tactical and operational management models. It was concluded that many management models are not favourable to the adoption of green IT, in particular many of the older standard management models which do not take a holistic view of corporate responsibilities. It is, therefore, incumbent upon managers not to place excessive reliance on such models.

There is a need for the development of new management models, which more explicitly integrate traditional bottom line considerations with the wider ethical responsibilities of companies, in particular those relating to sustainability.

Future research directions include empirical analysis of the impact of the use of management models on a sustainability culture within IT and consideration of the effect of operating within different cultures. There is also a need for development of more rigorous metrics for green IT.

References

[1] M. Menguc and L. Ozanne, "Challenges of the green imperative: a natural resource-based approach to the environmental-business performance relationship," Journal of Business Research, vol. 58, 2005, pp. 430–438.

[2] J. Elkington, "Towards the sustainable corporation." California Management Review, vol. Winter, 1994, pp. 90–100.

[3] ——, "Enter the triple bottom line," in The Triple Bottom Line: Does It All Add up?, A. Henriques and J. Richardson, Eds. Earthscan, London, 2004, pp. 1–16.

[4] G. H. Brundtland, "Our common future," in Report of the World Commission on Environment and Development. Oxford University Press, 1987.

[5] U. Nations, "Paris agreement," in Framework Convention on Climate Change. United Nations, 2016.

[6] T. Jenkin, J. Webster, and L. McShane, "An agenda for 'green' information technology and systems research," Information and Organization, vol. 21, 2011, pp. 17–140.

[7] K. Siegler and B. Gaughan, "A practical approach to green IT," Webinar.

[8] M. O'Neill, Green IT for Sustainable Business Practice. British Computer Society, 2010.

[9] T. Velte, A. Velte, and R. Elsenpeter, Green IT: Reduce your Information System's Environmental Impact While Adding to the Bottom Line. McGraw Hill, 2008.

[10] R. T. Watson, M. C. Boudreau, A. Chen, and M. H. Huber, "Green is: Building sustainable business practices," in Information Systems. Athens, GA, USA: Global Text Project, 2008.

[11] V. Dao, I. Langella, and J. Carbo, "From green to sustainability: Information technology and an integrated sustainability framework," Journal of Strategic Information Systems, vol. 20, 2011, pp. 63–79.

[12] A. Bokolo, M. Majid, and A. Romli, "Organizational culture and leadership: Preconditions for the development of a sustainable corporation," Sustainable and Applied Information Technology, vol. 95, no. 9, 2017, pp. 1875–1915.

[13] S. Murugesan and G. R. Gangadharan, Green IT: An Overview, S. Murugesan and G. R. Gangadharan, Eds. Wiley, 2012.

[14] Q. Deng and J. Shaobo, "Organizational green IT adoption: Concept and evidence," Sustainability, vol. 7, no. 12, 2015.

[15] B. Wernenfelt, "A resource based view of the firm," Strategic Management Journal, vol. 5, 1984, pp. 171–180.

[16] S. L. Hart, "A natural-resource based view of the firm," Academy of Management Review, vol. 20, no. 4, 1995.

[17] Where did FT500 chief executives go to business school? Financial Times. [Online]. Available: https://www.ft.com/content/3a63c054-b885-11e5-b151-8e15c9a029fb [retrieved: 7 August, 2017]

[18] M. VanAssen, Key Management Models: The 60+ models every manager needs to know. Prentice Hall, 2009.

[19] H. I. Ansoff, Corporate Strategy. Penguin Books, 1987.

[20] C. Christensen, The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail. Harvard Business Review Press, 1997.

[21] D. C. Hambrick, I. C. Macmillan, and D. L. Day, "Strategic attributes and performance in the BCG matrix," Academy of Management Journal, vol. 25, 1982, pp. 510–531.

[22] W. C. Kim and R. Mauborgne, Blue Ocean Strategy. Harvard Business School Press, 2005.

[23] J. Kay, Foundations of Corporate Success: How business strategies add value. Oxford University Press, 1993.

[24] K. Cameron and D. R. Ettington, "The conceptual foundations of organizational culture," in Higher Education: Handbook of Theory and Research, J. C. Smart, Ed. Norwell, Mass.: Kluwer, 1988, vol. 4.

[25] E. Schein, Organizational Culture and Leadership. Jossey-Bass, 1992.

[26] K. Cameron, D. Bright, and A. Caza, "Exploring the relationship between organizational virtuousness and performance," American Behavioral Scientist, vol. 47, 2004, pp. 766–790.

[27] G. Hofstede, Culture's Consequences. SAGE, 1980.

[28] R. Quinn and J. Rohrbaugh, "A spatial model of effectiveness criteria: Toward a competing values approach to organizational analysis," Management Science, vol. 29, 1983.

[29] K. Cameron and R. Quinn, Diagnosing and Changing Organizational Culture Based on the Competing Values Framework. Wiley, 2011.

[30] R. Quinn and G. Spreitzer, "The psychometrics of the competing values culture instrument and an analysis of the impact of organizational culture on quality of life," in Research in Organizational Change and Development, R. W. Woodman and W. A. Passmore, Eds. Greenwich Conn.: JAI Press, 1991, vol. 5.

[31] O. Richard, A. McMillan-Capehart, S. N. Bhuian, and E. C. Taylor, "Antecedents and consequences of psychological contracts: Does organizational culture really matter," Journal of Business Research, vol. 62, 2009, pp. 818–825.

[32] B. T. Gregory, S. G. Harris, A. A. Armenakis, and C. L. Shook, "Organizational culture and effectiveness: A study of values, attitudes and organizational outcomes," Journal of Business Research, vol. 62, 2009, pp. 673–679.

[33] W. M. Campbell, M. Ratcliffe, P. Moore, and M. Sharma, "The influence of culture on the adoption of green IT," in Green Services Engineering, Optimization and Modeling in the Information Age, X. Liu and Y. Li, Eds. Earthscan, London, 2015, pp. 25–60.

[34] W. M. Campbell, M. Ratcliffe, and P. Moore, "An exploration of the impact of organizational culture on the adoption of green IT," in Proceedings of the Green Computing and Communications Conference (Green-com), 2013, pp. 68–76.

[35] M. Beer and N. Nohria, Breaking the Code of Change. Harvard Business School Press, 2000.

[36] J. C. Henderson and N. Venkatram, "Understanding strategic alignment," Business Quarterly, vol. 55, no. 3, 1991, pp. 12–89.

[37] J. P. Cotter, Breaking the Code of Change. A Force for Change: How Leadership Differs from Management, 1990.

 ISBN: 978-1-61208-588-3

An Optimal Energy Conservation Measure (ECM) Decision Method based on Greenhouse Gas Reduction Target

Hong-Soon Nam, Tae-Hyung Kim and Youn-Kwae Jeong
IoT Research Division
Electronics and Telecommunications Research Institute
Daejeon, Korea
e-mail: {hsnam, taehyung, ykjeong}@etri.re.kr

Abstract— **This paper presents an optimal energy conservation measure (ECM) decision method, which is to find optimal ECMs to minimize the initial implementation cost of ECMs for satisfying the target of greenhouse gas (GHG) emission reduction. The method estimates energy savings by implementing ECMs and calculates GHG emission reduction considering both energy savings and carbon dioxide emission factors of each energy source. Then, it decides on the optimal ECMs whose initial implementation cost is minimal while meeting the target of GHG emission reduction. This paper modifies the knapsack algorithm to decide an optimal ECM combination to satisfy the target of GHG emission reduction and presents the simulation results including the optimal ECM list, the amount of GHG emission reduction and implementation cost to verify the optimal ECMs.**

Keywords- energy conservation measure (ECM); greenhouse gas (GHG); building retrofit; building energy.

I. INTRODUCTION

Globally, energy consumption in buildings makes up about 35% of the total energy consumption. Furthermore, the cost of building energy accounts for almost 30% of all building management costs [1]-[2]. For this reason, reducing the amount of building energy is an important issue in terms of global warming and the exhaustion of energy resources, as well as cost reduction. Energy saving in buildings can lead to both enormous cost savings and great greenhouse gas (GHG) emission reductions. Korea has planned to save GHG emission by 37% from business-as-usual (BAU) level by 2030 across all economic sectors. To cope with the twenty-first session of the Conference of the Parties (COP 21), the Korean government has announced a plan to new public buildings to be a zero energy building by 2020 and also expand the plan toward new private buildings by 2025. The number of public buildings in Korea is about 190,000. Thus, it is important to draw up a retrofit budget every year for a number of public buildings to meet the target of GHG emission reduction.

In existing buildings, on the other hand, much of energy consumptions can be safely reduced by the adoption of the adequate energy conservation measures (ECMs). ECMs are to reduce building energy consumptions by reducing operating time, improving energy efficiency and adopting new and renewable energy, which results in both energy cost saving and GHG emission reduction. However, the number of combinations of selectable ECMs is excessively large. Thus, it is difficult for building owners and project managers to select an optimal ECM combination in which the initial implementation cost is minimized while the target of GHG emission reduction is satisfied for a building retrofit.

Energy saving caused by implementing ECMs can be estimated by the measurement and verification (M&V) methodology [6]. The M&V methodology is to monitor and quantify the changes in the performance and operational parameters which are measured or calculated. The values of the parameters are needed to calculate energy savings associated with each ECM implementation. Thus, GHG reduction can be obtained from the energy savings. Recent researches in building retrofit include simulation tools based on various databases and standards [7]-[10].

This paper presents a method to decide the minimal initial implementation cost of a building retrofit for satisfying the target of GHG emission reduction. In this paper, we present how to analyze energy savings and GHG emission reduction in Section II. Afterwards, we describe how to decide the optimal ECMs whose implementation cost is minimized for satisfying the target of GHG emission reduction in Section III. Section IV describes the simulation environment and presents simulation results including an optimal ECM list, estimated GHG emission reduction and initial implementation cost, and finally, we conclude this paper in Section V.

II. GHG EMISSION REDUCTION ANALYSIS

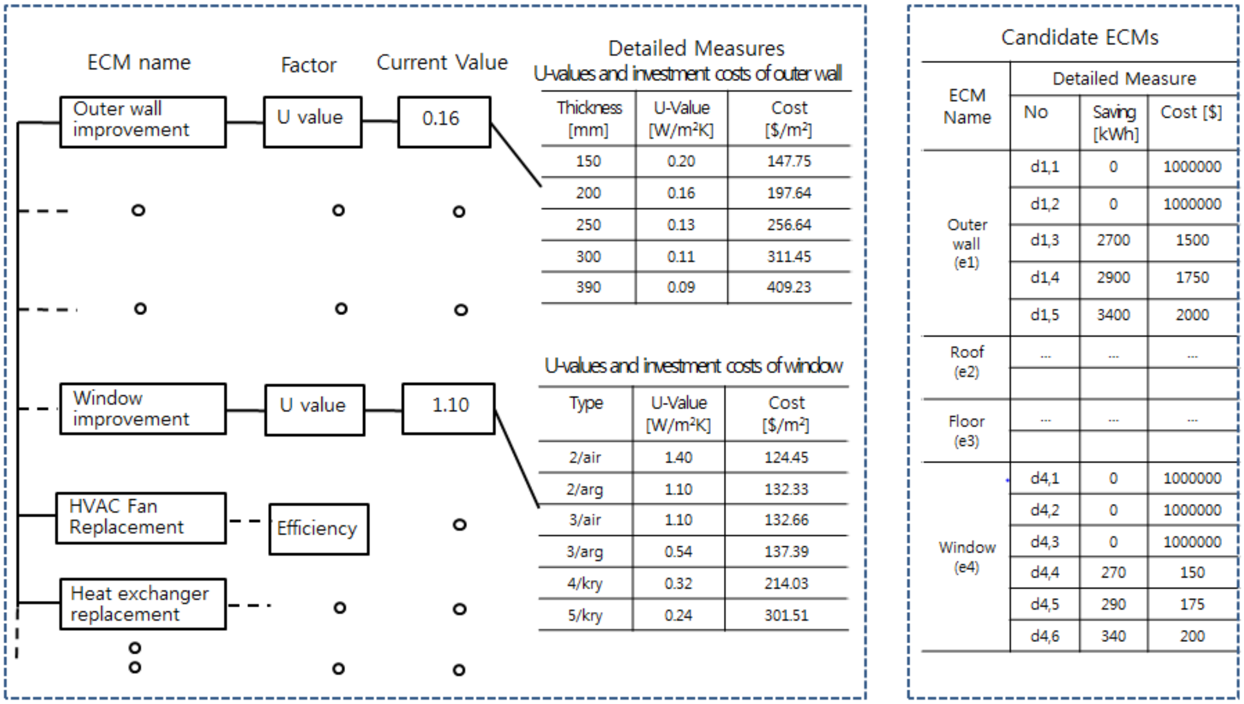
A. Energy Saving

The amount of energy saving can be estimated by comparing the difference of energy consumptions between before and after ECM adoption. In order to determine the energy saving, baseline energy E_{base} and post-installation energy E_{post} are firstly defined as the amount of energy that would be consumed without ECM implementation and the estimated or measured energy consumptions after the implementation, respectively [4]-[5]. Thus, the energy saving E_{save} is obtained as follows:

$$E_{save} = (E_{base} \pm E_{adj}) - E_{post} \quad (1)$$

where, E_{adj} represents adjustment energy to compensate for the changes in occupant behavior and weather condition, and for the difference of other factors between the baseline period and performance evaluation period.

 ISBN: 978-1-61208-588-3



(a) Typical ECM classification and detailed measures (b) Candidate ECMs

Figure 1. A typical ECM classification and cadidate ECMs of a building to be retrofitted

On the other hand, the energy saving E_{save} of an ECM also can be estimated depending on the ECM factor. A typical classification of ECMs and detailed measures (a) and candidate ECMs (b) are shown Figure 1. ECMs for building retrofit may include various types, such as construction, facility, lighting, and new and renewable energy. In figure 1(a), the ECM for outer wall improvement has five detailed measures with different U-values, thermal transmittance and costs per square meters. When the current thermal transmittance is 1.10, the detailed measures for improving thermal transmittance may be three detailed measures, $d_{1,3}$, $d_{1,4}$ and $d_{1,5}$. The others, $d_{1,1}$ and $d_{1,2}$, can be neglected. Candidate ECMs are composed of the ECMs that are capable of improving energy efficiency in figure 1(b). Candidate ECMs have energy savings and implementation costs of the corresponding building, which are calculated based on the ECM factors, weather condition and building information including location, size and occupant behavior.

A building retrofit includes one or more ECMs. If N ECMs are included in a building retrofit project, the energy savings of the project will be calculated as follows:

$$E_{save} = \sum_{i=1}^{N} E_{save,i} \cdot \tag{2}$$

where, E_{save} and $E_{save,i}$ denote the overall energy saving by ECMs and the estimated energy saving of ECM i, respectively, in a building retrofit project.

B. GHG Emission Reduction

Greenhouse gas, like CO2, results from the burning of fossil fuel sources including coal, petroleum and natural gas. Each fuel source has a different GHG emission factor, as shown in Table I [12].

TABLE I. CARBON DIOXIDE EMISSION FACTORS

Fuel	EIA Fuel Code	Factor [kg/MBtu]
Bituminous Coal	BIT	93.3
Distillate Fuel Oil	DFO	73.16
Jet Fuel	JF	70.9
Natural Gas	NG	53.07
Propane Gas	PG	63.07
Waste Coal	WC	93.3
Waste Oil	WO	95.25

* 1 [BTU/h] = 0.29307107 [W]

Then, the amount of GHG emission reduction, $GHG_{reduction}$ is obtained by

$$GHG_{reduction} = \sum_{i=1}^{N} E_{save,i} C_{fuel,i} \tag{3}$$

where, $C_{fuel,i}$ represents the GHG emission factor of the fuel source of ECM i.

 ISBN: 978-1-61208-588-3

III. Optimal ECM Decision based on GHG Emission Reduction Target

A problem of zero energy building is expensive. Likewise, a lot of money is required for existing building retrofit to meet the target of GHG emission reduction. Thus, the proposed optimal ECM decision method minimizes the implementation cost of ECMs for satisfying the target of GHG emission reduction, *T_GHG_R*. The implementation cost is obtained by

$$\text{Minimizes } \sum_{i \in T} Cost_i$$
$$\text{Subject to } \sum_{i \in T} GHG\_R_i \geq T\_GHG\_R \,. \quad (4)$$

where, GHG emission related parameters are shown in Table II.

TABLE II. GHG Related Paeameters

Parameter	Description
Cost	ECM impementation cost [$]
GHG_R	GHG emission reduction [ton/y]
T_GHG_R	Target of GHG emission reduction [ton/y]

From (3), the proposed algorithm tries to find an optimal combination of ECMs to minimize the implementation cost of the building retrofit, so the algorithm chooses the best possible outcome. We modified the Knapsack algorithm [11] to minimize the amount of implementation cost for the retrofit while still keeping the overall GHG emission reduction larger than or equal to its target. The pseudocode is shown in Figure 2.

We start with a set of candidate ECMs, whose implementation costs and energy savings can be obtained based on the corresponding building information including location, size and occupant behavior.

```
// c: cost;          r:GHG emission reduction;
// T: GHG emission target;     L: ECM list;

GHG_OptimalECM (c, r, n, T)
{
  for (c=0 to C) R[0, c]=0;
  for (i=0 to n)
    for (c=0 to C)
      if (c[i] ≤ c)
        R[i, c] = max{R[i-1, c], r[i] + R[i-1, c - c[i]};
      else
        R[i, c] = R[i-1, c};
      if (R[i, c] ≥ R[i-1, c])
        L[i, c] = L[i-1, c] ∪ {i};
      if (R[i, c] ≥ T)
        Cost = c; List = L[i, c]; Reduction = R[i, c];
        break GHG_optimalECM;
}
```

Figure 2. Psudocode for optimal ECM decision algorithm

The algorithm is to determine whether each ECM will be included in an ECM list for the building retrofit. Therefore, the total GHG emission reduction is larger than or equal to a given limitation while maintaining the total implementation cost is as low as possible.

IV. Simulation

To examine the proposed method, we set up an optimal ECM decision tool, as shown in Figure 3, which consists of building information, data base, GHG based optimal ECM decision engine and output unit. The building information includes building attributes, consumptions and energy diagnosis results. The current value of the attribute of each ECM can be obtained from the energy diagnosis results of the corresponding build. The database includes ECM, climate and energy price information.

The tool analyzes energy savings associated with each ECM for a building retrofit and organizes the candidate ECM table with energy saving and implementation cost. An example of a candidate ECM table is shown in Table III. Table IV represents the estimated optimal ECM list, initial implementation cost and GHG emission reduction. If the target of GHG emission reduction is 800 kilograms per year, the optimal ECM combination is {5, 6, 7}, the lowest implementation cost is 4800, and estimated GHG reduction is 820 kilograms per year, respectively.

From the results, we can see the budget required for a building retrofit to meet the target of GHG emission reduction and prioritize which buildings to be firstly retrofitted within an allowable budget.

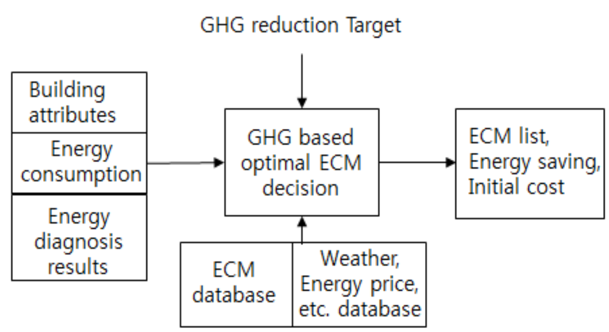


Figure 3. Architecture of the proposed method

TABLE III. An Example of Candidate ECMs with GHG Reduction and Cost of the Correponding Building

# of ECM	GHG reduction	Cost
1	300	2000
2	100	800
3	80	700
4	120	1000
5	200	1200
6	350	2100
7	270	1500
8	210	1400
9	150	1700

 ISBN: 978-1-61208-588-3

TABLE IV. OPTIMAL ECM AND IMPLEMENTATION COST FOR GHG REDUCTION TARGET

GHG reduction target	Estimated GHG reduction	Implementation cost [$]	Optimal ECM list
50	80	700	3
100	100	800	2
200	200	1200	5
300	300	2000	1
400	400	2800	1, 2
500	500	3200	1, 5
600	600	4000	1, 2, 5
700	700	4300	3, 6, 7
800	820	4800	5, 6, 7
900	910	5700	3, 6, 7, 8

V. CONCLUSIONS

In this paper, an optimal ECM decision method has been presented, which determines an optimal ECM combination for building retrofit. The optimal ECM combination is a subset of ECMs to minimize the initial implementation cost for satisfying the target of GHG emission reduction. ECMs can reduce building energy so building owners and project managers can decide an optimal ECM combination both to reduce GHG emission and to save energy cost. The proposed method modified the knapsack algorithm to decide an optimal ECM combination among numerous combinations of ECMs. The presented method provides building owners and project managers with the optimal ECM list, estimated implementation cost and estimated GHG emission reduction.

Much of building energy can be reduced with a suitable ECM adoption, which leads to enormous GHG reduction. However, the number of ECM combinations increases exponentially with the number of ECMs and each ECM requests an initial implementation cost. For this reason, it is difficult for building owners and project managers to select a suitable ECM combination for their building retrofits to reduce GHG emission within a limited budget. To estimate the initial implementation cost for satisfying the target of GHG emission reduction, this paper set up an optimal ECM decision environment and performed simulations on GHG emission reduction. The analysis results provide the optimal ECM combination to meet the target of GHG emission reduction while the implementation cost is minimal.

Further studies are necessary to get more data on detailed measures for ECMs and verify simulation results of the optimal ECM decision method for various buildings.

ACKNOWLEDGMENT

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea. (No. 20152010103180)

REFERENCES

[1] S. H. Lee, T. Hong, and M. A. Piette, “Review of existing energy retrofit tools,” LBNL, July 2014. [Online]. Avialable from: http://eetd.lbl.gov/publications/ 2017.07.

[2] Technavio, “Global Building Energy Software Market,” 2015.

[3] E. Pikas, M. Thalfeldt, and J. Kurnitski , “Cost optimal and nearly zero energy building solutions foroffice buildings,” Energy and Buildings, Vol. 74, pp. 30–42, 2014.

[4] ISO 13790 (2008), “Energy performance of buildings – Calculation of energy use for space heating and cooling,” 2008.

[5] US DoE, ”Energy Efficiency Program Impact Evaluation Guide,” 2012. [Online]. avialable from: https://www4.eere.energy.gov/ 2017.07.

[6] US DoE, ”M&V Guidelines: MeasurementAnd Verification for Performance-Based Contracts Version4.0, ”Federal Energy Management Program.(2015). [Online]. Available from: www1.eere.energy.gov/femp/, 2017.07.

[7] P. L. Luis, J. Ortiz, and C. Pout, “A review on buildings energy consumption information,” Energy and Buildings, vol. 40, pp. 394-398. 2008.

[8] B. R. Champion and S. A. Gabriel, “An improved strategic decision-making model for energy conservation measures,” Energy Strategy Reviews vol. 6, pp. 92-108, 2015.

[9] C. Baglivo, P. M. Congedo, D. D'Agostino, and I. Zaca, “Cost-optimal analysis and technical comparison between standard and high efficient mono-residential buildings in a warm climate,” energy Vol(83), pp. 560-575, 2015.

[10] H. S Nam, J. T Kim, T. H Kim, Y. K Jeong, and I. W Lee, “Economic Impact Analysis of Energy Conservation Measures for Building Remodeling,” EMERGING 2016, pp. 59-62, 2016.

[11] S. Martello and P. Toth, “Knapsack Problem: Algorithm and computer implementations,” Avialable from: http://www.or.deis.unibo.it/knapsack.html, 2017.07.

[12] IEA, “Carbon Dioxide Uncontrolled Emission Factors,” [Online]. avialable from: https://www.eia.gov/electricity/annual/html/epa_a_03.html, 2017.07.

 ISBN: 978-1-61208-588-3

Advanced Metering Infrastructure Data Driven Phase Identification in Smart Grid

Wenyu Wang, Nanpeng Yu, Zhouyu Lu

Department of Electrical and Computer Engineering
University of California, Riverside
Riverside, California 92521
Email: wwang032@ucr.edu, nyu@ece.ucr.edu, zlu044@ucr.edu

***Abstract*—Many important distribution network applications, such as load balancing, state-estimation, and network reconfiguration, depend on accurate phase connectivity information. The existing data-driven phase identification algorithms have a few drawbacks. First, the existing algorithms require the number of phase connections as an input. Second, they can not provide accurate results when there is a mix of phase-to-neutral and phase-to-phase connected smart meters, or when the distribution circuit is less unbalanced. This paper develops an advanced metering infrastructure (AMI) data driven phase identification algorithm that addresses the drawbacks of the existing solutions in two ways. First, it leverages a nonlinear dimensionality reduction technique to extract key features from the voltage time series. Second, a constraint-driven hybrid clustering (CHC) algorithm is developed to dynamically create smart meter clusters with arbitrary shapes. The field validation results show that the proposed algorithm outperforms the existing ones. The improvement in the phase identification accuracy is more pronounced for distribution feeders that are less unbalanced. In addition, this paper discovers that more granular voltage time series leads to higher phase identification accuracy.**

***Keywords*—AMI; density-based clustering; phase identification; smart grid; t-SNE.**

I. Introduction

It is estimated that electric utilities around the world will spend $10.1 billion on advanced metering infrastructure (AMI) data analytics solutions through 2021 [1]. The boom in the development and implementation of AMI data analytics is driven by two trends. First, electric utilities which have already deployed or plan to deploy the AMI are looking for new value streams to justify the business case of the AMI projects. Second, the advent of distributed energy resources (DERs) on the edge of the distribution grid is creating significant challenges and opportunities for the electric utilities and third-party aggregators.

The phase identification problem is defined as identifying the phase connectivity of each smart meter and structure in the power distribution network [2]. It is a critical AMI data analytics application due to two reasons. First, the rise of DERs requires the distribution system operators to actively manage the distribution grid to coordinate the operations of the DERs. However, most electric utilities in the world do not have accurate records of the phase connectivity of their distribution networks to enable advanced control strategies. Second, it is labor and capital intensive to perform phase identification using field validation tools. Therefore, conducting phase identification with AMI data driven analytics can provide another useful justification for the deployment of AMI projects.

In this paper, an AMI data driven machine learning algorithm is developed to solve the phase identification problem. The proposed algorithm leverages voltage magnitude data recorded by the AMI to identify the phase connection of each smart meter and structure. A nonlinear dimensionality reduction technique is first used to extract key features from the voltage time series. A constraint-driven hybrid clustering (CHC) algorithm is then developed to separate smart meters/structures into various clusters. Finally, the phase connection of each cluster can be identified by performing field validations on the phase connections of very few smart meters. Comprehensive case studies are conducted on 5 distribution circuits, which went through detailed field validations. The AMI data driven machine learning algorithm has yielded high accuracies on all circuits. In addition, this paper discovers that more granular voltage readings will lead to even more accurate phase identification results.

Compared to the existing data-driven phase identification algorithms, the proposed method has the following advantages:

1) The proposed algorithm does not require prior knowledge about the number of phase connections in the distribution system. Most of the existing AMI data driven methods need the number of phase connections as an input parameter.
2) The proposed algorithm works well with distribution feeders that have both phase-to-neutral and phase-to-phase connections. Most of the existing techniques are only capable of identifying the phase connections in distribution feeders with only phase-to-neutral connections or phase-to-phase connections.
3) The accuracy of the proposed phase identification algorithm is not very sensitive to the level of unbalance in a distribution feeder.

Currently, most electric utilities conduct phase identification using special phase meters [3][4]. Typically, two phase meters/units are used. One unit is located at the substation to serve as the reference. The other is called the field unit and is located at the smart meter/structure of interest in the distribution feeder. The working mechanism of these special phase meters is

 ISBN: 978-1-61208-588-3

very similar to that of the phasor measurement units except that the phase meters are mobile. With GPS time, the phase angle difference between the reference point and the field structure can be accurately measured, which then determines the phase connectivity of the field structure. Although phase meters provide highly accurate phase identification results, this solution is very time consuming and labor intensive, which make it unsuitable for large-scale deployment.

The existing data-driven algorithms leverage electric load and voltage magnitude measurements from the AMI to identify the phase connections of the smart meters and structures in the distribution network. These data-driven algorithms include supply and consumption balancing [5][6], linear regressions and correlation analysis [7][8], and constrained k-means clustering algorithm (CK-Means) [2]. However, the existing data-driven algorithms have three drawbacks. First, all of these methods assume that the number of phase connections are known. Second, the existing methods can not provide accurate phase identification results when there is a mix of phase-to-neutral and phase-to-phase connected smart meters and structures. Third, the existing methods are quite sensitive to the level of unbalance in a distribution feeder. The proposed AMI data driven phase identification algorithm addresses these drawbacks by leveraging a nonlinear dimensionality reduction technique to extract hidden features from voltage time series and using the CHC algorithm to dynamically create smart meter clusters with arbitrary shapes. The field validation results show that the proposed algorithm outperforms the existing methods in all of the 5 distribution feeders.

The rest of this paper is organized as follows. Section II studies the drawbacks of the existing data-driven phase identification algorithms. Section III describes the proposed phase identification algorithm in detail. Section IV presents the case studies on multiple distribution feeders to validate the proposed phase identification algorithm. Section V provides the conclusions.

II. Drawbacks of the Existing Data-driven Phase Identification Methods

Three main drawbacks of the existing phase identification methods are studied in detail below. As the CK-Means method is the most promising algorithm among the existing data-driven phase identification methods, it will be used as an example in the performance evaluation. A comprehensive study is conducted on 5 distribution feeders and 18 data sets to analyze the impact of unbalance level and the mix of phase connection types on the phase identification accuracy for the CK-Means method.

The general descriptions of the 5 distribution feeders and 18 data sets are shown in Table I. The feeder and smart meter data is provided by the Pacific Gas & Electric Company and Southern California Edison. The number of customers, feeder voltage level, proportion of the major phase connection types, and feeder peak load are listed in the second column of the table. A distribution feeder can have 3 possible phase-to-neutral connections, AN, BN, and CN, and/or 3 possible phase-to-phase connections, AB, BC, and CA, where A, B, C, and N denote the three phases' wires and the neutral wire. 2 months of smart meters' voltage data with 5-minute granularity is gathered from feeder 1, 2, and 3. 6 months of smart meters' voltage data with hourly granularity is gathered from feeder 4 and 5.

TABLE I. Descriptions of the Distribution Feeders

Feeder	Number of Customers, Feeder Voltage, and Peak Load	Month	Data Set
1	3200 customers (99.8% phase-to-neutral), 12.47 kV, 4.4 MW.	Nov 2016	s_1
		Dec 2016	s_2
2	4800 customers (98.8% phase-to-neutral), 12.47 kV, 8.3 MW.	Nov 2016	s_3
		Dec 2016	s_4
3	4000 customers (97% phase-to-neutral), 12.47 kV, 6.4 MW.	Nov 2016	s_5
		Dec 2016	s_6
4	1500 customers (100% phase-to-phase), 12.47 kV, 5.2 MW.	Aug 2015	s_7
		Sep 2015	s_8
		Oct 2015	s_9
		Nov 2015	s_{10}
		Dec 2015	s_{11}
		Jan 2016	s_{12}
5	2400 customers (84% phase-to-phase), 12.47 kV, 8.5 MW.	Aug 2015	s_{13}
		Sep 2015	s_{14}
		Oct 2015	s_{15}
		Nov 2015	s_{16}
		Dec 2015	s_{17}
		Jan 2016	s_{18}

In feeder 1, 2, and 3, some meters have missing voltage readings at different time intervals, making up 9%, 21%, and 18% of the total customer population respectively. The missing readings are filled in using the k-nearest neighbor (k-NN) imputation method. A meter's missing readings are imputed using the average values of the five nearest neighbor meters' corresponding readings. The distance between meters are measured by the Euclidean distance of the voltage time series of the corresponding meters.

To make the results comparable, the hourly average voltage magnitudes are calculated for feeder 1, 2, and 3. The hourly average voltage magnitudes are used as inputs in this section. Each of the 18 data sets includes one month of voltage magnitude data from a feeder. The drawbacks of the existing data-driven phase identification algorithms are explored in the next three subsections.

A. Number of Phase Connections

In order to solve the phase identification problem, the supply and consumption balancing approach [5][6] requires the number of phase connections in the distribution feeder as an input. In fact, the problem formulation in [5][6] only allows the identification of phase-to-neutral connections where the number of phase connections is 3. In the linear regression and correlation analysis [7][8], the number of phase connections in the feeder is also a mandatory input. In fact, both linear regression and correlation analysis work well when there are only three phase-to-neutral connections. The k-means clustering algorithm is used in the CK-Means method [2], where the number of phase connections/clusters needs to be known as prior knowledge. When applying the CK-Means method to identify the phase connections of the 5 distribution feeders, the number of clusters is set to be 3 for feeders 1 to 4, given that over 97% of the smart meters in these feeders only have 3 connection types. The number of clusters is set to be 6 for feeder 5.

B. Impact of Unbalance Level on the Phase Identification Accuracy

This subsection evaluates the impact of the distribution feeder's unbalance level on the phase identification accuracy of the CK-Means algorithm. The CK-Means algorithm works as follows: The voltage magnitude measurements are first standardized. Linear features are then extracted by using principal component analysis (PCA) and the top d components are selected. To provide a fair comparison with the proposed phase identification algorithm in Section IV, the number of principal components is set to 30. Next, the data points in the low-dimensional space are clustered by using a constrained k-means clustering algorithm. Must-link constraints are derived from the distribution feeders' connectivity information, which is typically available from the Geographical Information System (GIS). The must-link constraints state that if some smart meters are connected to the same lateral or transformer, then they must be linked together and grouped into the same cluster. To identify the phase of each cluster, field validations are performed on a must-link group of at least 20 smart meters that has the least mean squared distance to the cluster center.

The CK-Means algorithm is applied on the 18 voltage time series from the 5 distribution feeders. The phase identification accuracy is calculated based on independent field validations conducted by the electric utility companies. To measure the level of unbalance of a distribution feeder, define $u(t)$ as the level of unbalance of a feeder at time interval t:

$$u(t) = \frac{|I_A(t) - I_m(t)| + |I_B(t) - I_m(t)| + |I_C(t) - I_m(t)|}{3I_m(t)} \tag{1}$$

where $I_m(t) = \frac{1}{3}(I_A(t) + I_B(t) + I_C(t))$ is the mean of the distribution substation line currents of the three phases. $u(t)$ can be interpreted as the ratio of the average three-phase current deviation to the mean. The average level of unbalance for a distribution feeder $u(t)$ over a month is calculated for each data set.

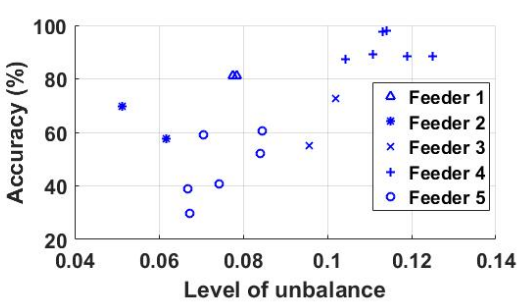


Figure 1. The phase identification accuracy of the CK-Means method under different levels of unbalance.

Figure 1 plots the phase identification accuracy against the level of unbalance. It shows that the CK-Means algorithm is very accurate for the highly unbalanced data sets. As the level of unbalance decreases, the phase identification accuracy drops quickly. This result is very intuitive. Imagine there is a perfectly balanced distribution feeder whose three phase wires have the same load distribution all the time. In this case, the level of unbalance should be zero. Therefore, it is impossible to distinguish the phase connections of the smart meters on the three phases with unsynchronized voltage magnitude measurements.

C. A Mix of Phase-to-Neutral and Phase-to-Phase Connections

In general, the existing data-driven phase identification algorithms do not perform well for the distribution feeders with a mix of phase-to-neutral and phase-to-phase connections. For example, Figure 1 shows that the phase identification accuracy is the lowest for feeder 5. This is because feeder 5 not only has a lower degree of unbalance, but also has all 6 possible phase connections types, AN, BN, CN, AB, BC, and CA. In this case, the default phase identification accuracy is only 16.7% instead of 33.3% for the distribution feeders with only three possible phase connections.

III. TECHNICAL METHODS

The overall framework of the proposed phase identification algorithm is illustrated in Figure 2. The phase identification methodology involves three stages. In stage 1, voltage magnitude measurements are collected from the smart meters. Each smart meter's readings are centered and normalized by their standard deviation. Key features are then extracted from the preprocessed voltage time series with a nonlinear dimensionality reduction method. In stage 2, the CHC algorithm is leveraged to cluster the low-dimensional data points generated in stage 1. In stage 3, the phase connection of each cluster

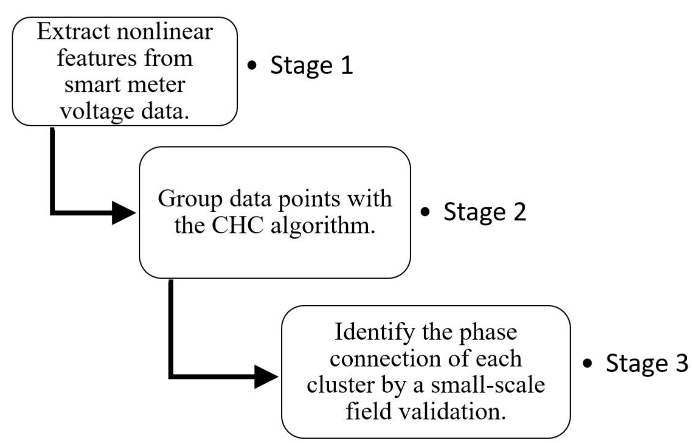


Figure 2. The overall framework of the proposed phase identification algorithm.

is identified by performing field validations on a very small number of smart meters. The three stages are explained in detail below.

A. Stage 1: Feature Extraction from Voltage Time Series

It is undesirable to directly work with raw voltage readings, which are high-dimensional and noisy. Therefore, in the first stage, dimensionality reduction techniques will be applied to extract key features from the raw voltage time series. The extracted features will then be fed into the CHC algorithm in stage 2.

Dimensionality reduction techniques can be divided into two categories, linear dimensionality reduction methods and nonlinear ones. Linear dimensionality reduction techniques, such as PCA, are restricted to learning only linear manifolds. However, high-dimensional data typically lies on or near a low-dimensional, nonlinear manifold [9]. Furthermore, it is very difficult for linear mappings to keep the low-dimensional representations of very similar points close together. This explains the lower accuracy of the phase identification algorithm using linear features for less unbalanced feeders. To address this problem, we turn to nonlinear dimensionality reduction methods. Many nonlinear dimensionality reduction techniques have been proposed, e.g., Sammon mapping [10], curvilinear components analysis (CCA) [11], Isomap [12], and t-distributed stochastic neighbor embedding (t-SNE) [9]. This paper adopts t-SNE, because it has been shown to work well with a wide range of data sets and captures both local and global data structures. t-SNE improves upon SNE [13] by 1) simplifying the gradient calculation with a symmetrized version of the SNE cost function and 2) adopting a Student's t-distribution rather than a Gaussian distribution to compute the similarity between two points in the low-dimensional space [9].

The basic idea of t-SNE is to convert the high-dimensional Euclidean distances between data points into joint probabilities and represent the data points in a low-dimensional space, so that similar joint probabilities are preserved. Suppose we need to map a high-dimensional data set $X = \{x_1, x_2, ..., x_n\}$ to a low-dimensional data set $Y = \{y_1, y_2, ..., y_n\}$. Define p_{ji} as a joint probability of X. p_{ji} is a symmetric approximation of the conditional probability that x_i would pick x_j as its neighbor. The neighbors are picked in proportion to their probability density under a Gaussian distribution centered at x_i with a variance σ_i. p_{ji} is calculated as $p_{ji} = p_{ij} = (p_{j|i} + p_{i|j})/2n$, where $p_{j|i}$ is calculated as:

$$p_{j|i} = \frac{\exp(-\|x_i - x_j\|^2/2\sigma_i^2)}{\sum_{l \neq i} \exp(-\|x_i - x_l\|^2/2\sigma_i^2)} \tag{2}$$

In the same way, define q_{ji} as a joint probability in Y, but under a Student's t-distribution with one degree of freedom. Then q_{ji} can be calculated as:

$$q_{ji} = q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{l \neq m}(1 + \|y_l - y_m\|^2)^{-1}} \tag{3}$$

Then given X, the mapping Y is found by minimizing the Kullback-Leibler divergence between joint probability distribution P, in the high-dimensional space, and the joint probability distribution Q, in the low-dimensional space:

$$C = D_{\text{KL}}(P\|Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{4}$$

The t-SNE algorithm requires three input parameters: 1) the output dimension d_{out} (typically selected to be either 2 or 3); 2) the initial dimension d_{in}, which is the dimension that the original data set is reduced to by PCA before performing t-SNE; 3) perplexity p, which is a measure of effective number of neighbors and controls σ_i. Since the objective function (4) is minimized using a gradient descent optimization that is initiated randomly, each run of t-SNE produces a slightly different mapping result. In practice, it is recommended to run t-SNE multiple times and select the result with the lowest cost function value in (4). More details of the t-SNE algorithm can be found in [9].

B. Stage 2: Group Data Points with the CHC Algorithm

After the preprocessed voltage time series are mapped to a 2-dimensional or 3-dimensional feature space through t-SNE, they need to be grouped into clusters. Three features of the phase identification problem need to be considered when designing the clustering algorithm. First, many electric utility companies do not know the number of phase connections for each of their distribution feeders. Second, the customers with the same phase connection in the low-dimensional feature space do not necessarily form a convex-shape cluster, which is very common in t-SNE applications [9][14][15]. Third, valuable distribution network connectivity information which defines the mapping between smart meters and laterals/transformers should be incorporated into the clustering algorithm.

In order to leverage the features of the phase identification problem, the CHC algorithm is developed and applied to solve the smart meter clustering problem. The proposed CHC framework synergistically combines the merits of an unsupervised density-based clustering algorithm and a supervised classification algorithm. This paper selects the density-based spatial clustering of applications with noise (DBSCAN) [16] as the unsupervised clustering algorithm in the CHC framework, because it naturally incorporates the first two features of the phase identification problem. Unlike centroid-based or medoid-based methods, DBSCAN does not need the number of clusters as an input parameter. In addition, DBSCAN is capable of discovering clusters with arbitrary shapes.

DBSCAN separates data points into different clusters and noise/outliers. The noise/outliers do not belong to any cluster. However, in the phase identification application, all smart meters must have a particular phase connection. To mitigate this drawback, k-nearest neighbor (k-NN) classification is adopted as the supervised machine learning algorithm in the

 ISBN: 978-1-61208-588-3

CHC framework to assign these outliers and points in the low-density region into one of the existing output clusters from DBSCAN. At last, the must-link constraints defined by the feeder connectivity model will be considered in reassigning smart meters connected to the same lateral/transformer to the same cluster.

1) Review of DBSCAN: A brief review of DBSCAN is provided here. DBSCAN is one prominent example of density-based clustering approach with high computational efficiency. The good efficiency of DBSCAN is crucial for deploying phase identification algorithms in electric utilities with t-housands of distribution feeders. The DBSCAN algorithm defines clusters and outliers based on four key concepts: ϵ neighborhood of a point, directly density-reachable, density-reachable, and density-connected. The algorithm requires two parameters: ϵ, the radius of neighborhood, and *MinPts*, the minimum number of data points in an ϵ neighborhood. The ϵ neighborhood of a point p is defined as the set of points in the data set with a distance to p less than ϵ. A point p is a core point if it has at least *MinPts* neighbors within the radius ϵ. These neighbors are directly density-reachable from p. A point q is density-reachable from p if there is a path $p, p_1, p_2, ..., p_m, q$ such that each point is directly reachable from the previous point. Two points are considered density-connected if they have a distance of less than ϵ. These four definitions allow us to define the transitive hull of density-connected points, forming density-based clusters. The points on the border of the clusters are called border points. Any point(s) not reachable from a core point is counted as an outlier or noise.

2) The CHC Algorithm: The framework of the CHC algorithm is shown in Algorithm 1. It requires four input parameters, α, k, ϵ, and *MinPts*. α is a threshold parameter used to filter out very small clusters. k is the parameter in the k-NN algorithm representing the number of nearest neighbors.

1: Run the DBSCAN algorithm on a preprocessed data set D with n data points with parameters ϵ and *MinPts*.
2: Define a threshold coefficient $\alpha \in (0, 1)$. Given the output of step 1, keep the data points from the clusters of size greater than or equal to αn as the training data set. Suppose there are c clusters kept. All the data points outside these clusters are “un-clustered” data points.
3: Assign all un-clustered data points to one of the c clusters with the k-NN algorithm.
4: With must-link constraints, the data set D can be divided into N groups $D_1, ..., D_N$. If a data point has no links to others, it forms a group itself. In each group D_i, the data points may have been assigned to different clusters. To enforce the constraints, assign all data points in group D_i to the cluster that contains the largest number of data points in D_i.
5: Return the final clustering result.

Figure 3. Algorithm 1: the CHC algorithm

The CHC algorithm has 5 steps. Step 1 runs the DB-SCAN algorithm on features extracted by the t-SNE algorithm. Depending on the distribution of data points in the low-dimensional feature space, the DBSCAN output may include large clusters, small clusters, and noise/outliers. Step 2 filters out the points in the small clusters and noise/outliers and only keeps the large clusters as the training data set for the next step. Step 3 classifies the points from small clusters and noise/outliers with k-NN algorithm using the training data points from the large clusters. Step 4 enforces the must-link constraints by assigning all smart meters connected to the same lateral/transformer to the same cluster. The final clustering results will be returned in step 5.

Note that researchers have proposed alternative approaches, such as C-DBSCAN [17] to integrate constraints into density-based clustering algorithms. In the C-DBSCAN algorithm, the data points from different clusters involved in a must-link constraint are simply forced to merge together. However, when the preprocessed voltage time series are mapped to the low-dimensional space, we often encounter cases where a very small number of meters connected to one phase are spread over two clusters representing two phases. To address this issue, in step 4 of the proposed CHC algorithm, we only reassign all the data points connected by a must-link constraint to the same cluster without affecting the grouping of other data points.

C. Stage 3: Phase Identification for Clustered Customers

The final stage identifies the phase connection of the clusters determined in stage 2. This can be accomplished by performing field validations on a small number of samples of smart meters with phase measurement tools [3][4]. The cost associated with the field validation is minimal as the number of customers that require phase measurement is quite small. To achieve the highest accuracy, the small sample of customers should be chosen as close to the clusters' centers as possible. Depending on the availability of must-link constraints, two sampling strategies can be implemented:

1) If there are no must-link constraints, then in each cluster choose m smart meters that are closest to the cluster center. Field validations can then be performed on these m smart meters. The most frequent phase connection of these m meters is selected as the phase connection of all the customers in the cluster.
2) If must-link constraints are available, then in each cluster choose the group D_g that contains at least w customers and has the least mean squared distance to the cluster center. Field validations will be performed on any of the smart meters in group D_g. The phase connection of the group is selected as the phase connection of all the customers in the cluster.

IV. Case Studies

A. Experimental Design

Two types of experiments are designed below to 1) examine the performance of the proposed phase identification algorithm and 2) explore the impact of smart meter data granularity on the phase identification accuracy.

The first set of experiments compare the performance of the constrained k-means clustering algorithm with linear dimensionality reduction [2] and the CHC algorithm with nonlinear dimensionality reduction proposed in this paper. The constrained k-means clustering algorithm with linear dimensionality reduction is referred to as "CK-Means" method. Both methods are evaluated over 18 hourly voltage time series gathered from 5 distribution feeders as described in Table I.

The second set of experiments evaluate the impact of smart meter sampling frequency on the accuracy of the proposed phase identification algorithm. The experiments are conducted over 6 voltage time series gathered from 3 distribution feeders. The smart meters on distribution feeder 1-3 were configured to record voltage magnitudes every 5 minutes. The average voltage magnitudes with hourly, 15-minute, and 5-minute granularity are used as inputs.

B. Parameter Selection

A few parameters need to be set up in order to run the proposed phase identification algorithm. In the feature extraction stage, three parameters from the t-SNE algorithm need to be selected. The dimensionality of the PCA output and t-SNE input d_{in} is set to be 30. The perplexity p is set to be 100. Note that these two parameters can be tuned by running the optimization several times on a data set and picking the parameters that yield the best map [9]. The dimensionality of the t-SNE output d_{out} is typically set to be 2 or 3. For better visualization, we set d_{out} to 2. In fact, the case study results with $d_{\text{out}} = 2$ and $d_{\text{out}} = 3$ are very similar.

In the proposed CHC algorithm, three key parameters *MinPts*, ϵ, and α need to be tuned first. The typical ranges for the three parameters are 8 to 20 for *MinPts*, 1 to 3 for ϵ, and 0.005 to 0.01 for α. When tuning these parameters, the aim is to see the data points in the t-SNE space being clustered appropriately. For example, assume we select some initial settings for *MinPts*, ϵ, and α, and get the clustering results as shown in Figure 5. Intuitively, cluster 11 and 15 should be two separate clusters. If the initial parameter setting merges these two clusters, then the parameters need to be tuned so that they are separated in the clustering results. In this particular case, we should decrease ϵ and/or increase *MinPts* to separate cluster 11 and 15. Note that ϵ is the radius of neighborhood and *MinPts* is the threshold for determining if a point p is a core point or a border point in a cluster. The parameter α controls the number of output clusters. If the value of α is too large, then the phase identification accuracy will be lower. However, if the value of α is too small, then a large number of meters need to be field validated, which increases implementation costs. k, the parameter of the k-NN, can be selected to be equal to *MinPts*. At last, in the field validation, choose the must-link group with at least $w = 20$ customers.

C. Performance of the Proposed Phase Identification Algorithm

The phase identification accuracies of the CK-Means method and the proposed phase identification algorithm are calculated based on field validation results. For the proposed algorithm, 30 runs of t-SNE are conducted. 10 runs with the lowest cost function values are kept. The average accuracy over the 10 runs are reported in Table II and Figure 4. As shown in the table, the proposed phase identification algorithm significantly outperforms the CK-Means method with all the data sets in terms of accuracy. On average, the proposed phase identification algorithm improves the identification accuracy by 19.81% over the CK-Means algorithm. Figure 4 shows that the improvement in phase identification accuracy varies by the unbalance level of the distribution circuit. The improvement is more significant for periods when the distribution feeder is less unbalanced.

The combinations of phase connections in the 5 testing feeders include 3 phase-to-neutral connections, 3 phase-to-phase connections, and a mix of all 6 possible connections. The accuracy of the proposed phase identification algorithm is very high under most cases. s_{13}, s_{14}, and s_{15} have relatively lower accuracy, because they have lower levels of unbalance and they have all 6 possible connections, which is more difficult to identify than other feeders. When the level of unbalance is higher, the accuracy is greatly improved in s_{16}, s_{17}, and s_{18}, whose accuracies are very decent for a feeder with all the 6 possible phase connections. Figure 5 illustrates the clustering result of data set s_{18} in the 2-dimensional t-SNE map, using the proposed phase identification algorithm. In the figure, each dot represents a smart meter. Figure 6 depicts the actual phase connection of each smart meter. By comparing

TABLE II. Phase Identification Accuracies

Feeder	Data Set	Level of Unbalance	CK-Means Accuracy (%)	Proposed Algorithm Accuracy (%)
1	s_1	0.0785	81.21	93.06
	s_2	0.0776	81.18	93.62
2	s_3	0.0514	69.67	87.55
	s_4	0.0617	57.51	87.79
3	s_5	0.0956	54.91	83.94
	s_6	0.1019	72.78	82.83
4	s_7	0.1109	89.29	98.60
	s_8	0.1141	97.82	98.94
	s_9	0.1131	97.79	99.63
	s_{10}	0.1190	88.42	99.66
	s_{11}	0.1043	87.49	99.88
	s_{12}	0.1250	88.34	99.65
5	s_{13}	0.0673	29.80	73.18
	s_{14}	0.0668	38.80	73.32
	s_{15}	0.0705	59.07	67.01
	s_{16}	0.0742	40.56	88.19
	s_{17}	0.0846	60.49	87.11
	s_{18}	0.0842	52.02	89.84

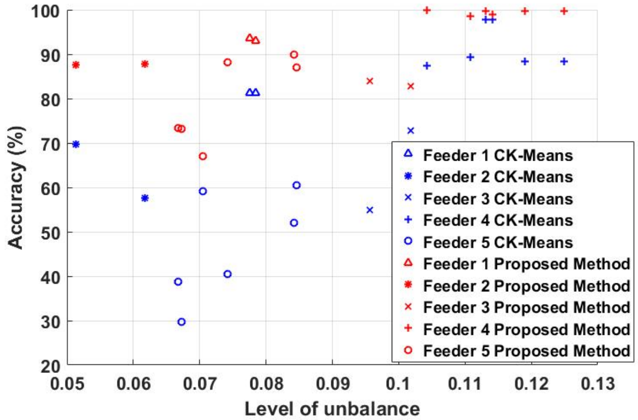


Figure 4. The phase identification accuracy with CK-Means and proposed algorithm.

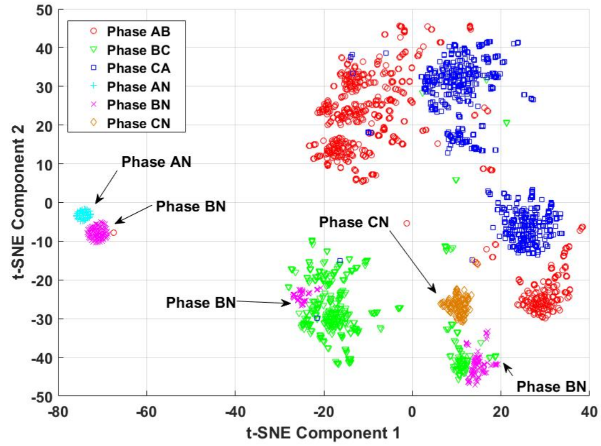


Figure 6. Field validated phase connections of data set s_{18} in the 2-dimensional t-SNE map.

Figure 5 and Figure 6, it is shown that the proposed phase identification algorithm not only groups phase-to-phase meters accurately, but also groups phase-to-neutral meters with a high accuracy. Cluster 2, 11, 12, 13, and 15 each represents one of the phase-to-neutral connections AN, BN, and CN, as indicated by the arrows in Figure 5 and Figure 6.

As a comparison, Figure 7 shows the distribution of smart meters from data set s_{18} in the 2-dimensional PCA map. The data points are not well separated according to phase connection. From Figure 7 and Figure 6, it is clear that the nonlinear dimensionality reduction technique, t-SNE, does a much better job in extracting hidden features from the voltage time series during a less unbalanced period for the feeders.

As shown in Figure 5, the clusters are in different sizes and shapes. Some of the clusters are non-convex. The proposed CHC algorithm has a great advantage in identifying clusters with such data point distributions. Figure 5 also shows how the must-link constraints could improve the phase identification accuracy. In the top right cluster 8, a few data points are linked together. Although a small number of the data points are located in cluster 14, they will eventually be assigned to cluster 8 due to the must-link constraint. From Figure 6, these data points should belong to cluster 8, which is connected to phase CA instead of phase AB.

D. Impact of the Smart Meter Sampling Frequency on the Phase Identification Accuracy

The phase identification accuracies of the proposed algorithm under 3 different meter reading granularity levels are calculated and summarized in Table III. It shows that as the granularity of meter readings increases from hourly to every 15 minutes and then 5 minutes, the phase identification accuracy increases. The average increase in the phase identification accuracy over the 3 distribution circuits is 3.36% when the meter reading granularity increases from hourly to 5 minutes. More granular voltage readings allow extractions of features/patterns

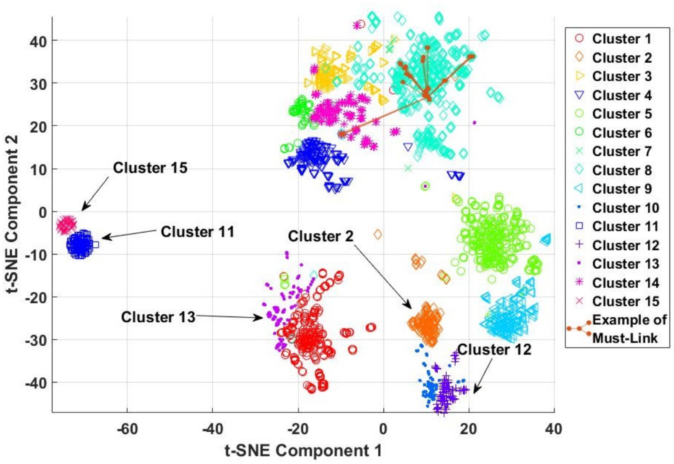


Figure 5. The clustering result in the 2-dimensional t-SNE map on data set s_{18}.

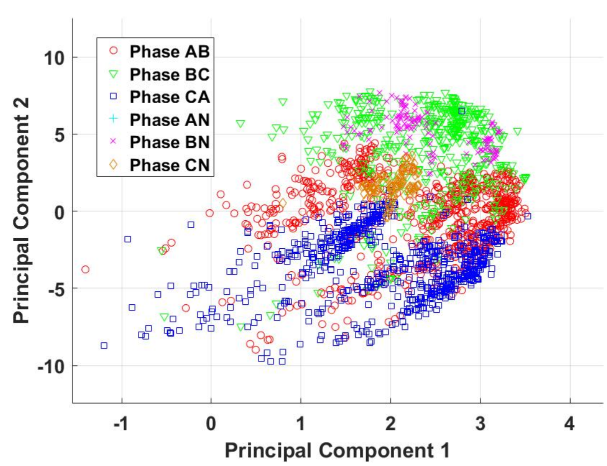


Figure 7. Field validated phase connections of data set s_{18} in the 2-dimensional PCA map.

TABLE III. IMPACT OF SAMPLING FREQUENCY ON THE PHASE IDENTIFICATION ACCURACY

Feeder	Data Set	Granularity of Meter Readings		
		1 hour	15-minute	5-minute
1	s_1	93.06%	93.93%	93.88%
	s_2	93.62%	94.32%	94.40%
2	s_3	87.55%	88.86%	92.03%
	s_4	87.79%	90.47%	89.93%
3	s_5	83.94%	90.02%	91.56%
	s_6	82.83%	84.51%	87.16%

that may not be present in coarse data sets. However, it should be noted that there are additional costs associated with gathering more granular smart meter data. Note that the phase identification accuracy decreases slightly for data set s_1 and s_4 when the sampling frequency increases from 15-minute to 5-minute. This is partly due to the randomness of the t-SNE mapping.

V. CONCLUSION

This paper develops an AMI data driven phase identification algorithm that addresses the drawbacks of the existing solutions. Compared to the existing solutions, the proposed algorithm has three main advantages. First, the proposed algorithm does not require prior knowledge about the number of phase connections in the distribution system. Second, the proposed algorithm works well with distribution feeders that have both phase-to-neutral and phase-to-phase connections. Third, the accuracy of the proposed phase identification algorithm is not very sensitive to the level of unbalance in a distribution feeder. Comprehensive field testing results on 5 distribution feeders show that the proposed algorithm significantly outperforms the existing methods. In addition, we discover that more granular voltage time series leads to higher phase identification accuracy.

In the proposed CHC algorithm, a few parameters need to be tuned manually. To implement the proposed AMI data driven phase identification algorithm on thousands of distribution feeders, we plan to develop an automatic parameter tuning algorithm.

ACKNOWLEDGMENT

The authors would like to thank Austen D'Lima, Joshua Davis, and Tom Martin from Southern California Edison and the Pacific Gas and Electric Company for fruitful discussions and supplying AMI, network connectivity, and field validation data.

REFERENCES

[1] A. Gupta and B. Kellison, "Utility AMI analytics at the grid edge: Strategies, markets and forecasts," GTM Research, Tech. Rep., 2016.
[2] W. Wang, N. Yu, B. Foggo, J. Davis, and J. Li, "Phase identification in electric power distribution systems by clustering of smart meter data," in *Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on*. IEEE, Dec. 2016, pp. 259–265.
[3] L. A. Pomatto, "Apparatus and method for identifying the phase of a three phase power line at a remote location," Apr. 23 1996, US Patent 5,510,700.
[4] W. S. Bierer, "Long range phasing voltmeter," Oct. 5 2010, US Patent 7,808,228.
[5] M. Dilek, "Integrated design of electrical distribution systems: Phase balancing and phase prediction case studies," Ph.D. dissertation, Virginia Polytechnic Institute and State University, 2001.
[6] V. Arya *et al.*, "Phase identification in smart grids," in *Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on*. IEEE, Oct. 2011, pp. 25–30.
[7] T. A. Short, "Advanced metering for phase identification, transformer identification, and secondary modeling," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 651–658, Jun. 2013.
[8] W. Luan, J. Peng, M. Maras, J. Lo, and B. Harapnuk, "Smart meter data analytics for distribution network connectivity verification," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1964–1971, Jul. 2015.
[9] L. V. D. Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, Nov. 2008.
[10] J. W. Sammon, "A nonlinear mapping for data structure analysis," *IEEE Transactions on Computers*, vol. 100, no. 5, pp. 401–409, May 1969.
[11] P. Demartines and J. Hérault, "Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets," *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 148–154, Jan. 1997.
[12] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.
[13] G. E. Hinton and S. T. Roweis, "Stochastic neighbor embedding," in *Advances in Neural Information Processing Systems*, 2003, pp. 857–864.
[14] E. Z. Macosko *et al.*, "Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets," *Cell*, vol. 161, no. 5, pp. 1202–1214, May 2015.
[15] A. Frome *et al.*, "Devise: A deep visual-semantic embedding model," in *Advances in Neural Information Processing Systems*, 2013, pp. 2121–2129.
[16] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *KDD*, vol. 96, no. 34, Aug. 1996, pp. 226–231.
[17] C. Ruiz, M. Spiliopoulou, and E. Menasalvas, "C-DBSCAN: Density-based clustering with constraints," in *International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing*. Springer, May 2007, pp. 216–223.

 ISBN: 978-1-61208-588-3