

# Phase Identification in Smart Grids

V. Arya, D. Seetharam, S. Kalyanaraman, K. Dontas  
IBM Research, India  
{vijay.arya, dseetharam, shivkumar-k, kedontas}@in.ibm.com

C. Pavlovski, S. Hoy  
IBM GBS, Australia  
{chris\_pav, steve.hoy}@au1.ibm.com

J. R. Kalagnanam  
IBM Research, NY USA  
jayant@us.ibm.com

**Abstract**—Electricity is distributed throughout the electrical power network in 3-phase voltage. This power reaches households as a single-phase voltage, generally 115vac or 240vac. This is achieved by allocating households with either phases A, B, or C of the final 3-phase power distributed to the street through a low voltage transformer. A present problem confronting the electrical power industry is identification of which particular phase a household is connected to. This information is often not tracked and the mechanisms for identifying phase require either manual intervention or costly signal injection technologies. Phase information is important as it is a foundation for the larger problem of balancing phase loads. Unbalanced phases lead to significant energy losses and sharply reduced asset lifetimes.

In this paper we propose a new approach to compute household phase. Our techniques are novel as they are purely based upon a time series of electrical power measurements taken at the household and at the distributing transformer. Our methods involve the use of integer programming and solutions can be retrieved using branch and bound search algorithms implemented by MIP solvers such as CPLEX. Furthermore, as the number of measurements increase, continuous relaxations of integer programs may also be used to retrieve household phase efficiently. Simulation results using a combination of synthetic and real smart meter datasets demonstrate the performance of our techniques and the number of measurements needed to uniquely identify household phase.

## I. INTRODUCTION

The majority of electrical power is generated at large power plants as a 3-phase AC voltage. The electricity is initially produced at a very high voltage and injected into a transmission and distribution network. A distribution grid typically consists of high voltage circuits that supply power to transformer substations that step down voltage progressively until delivering this power to the household or building. When finally distributing power to the house, each low voltage transformer supplies power to about 50-200 homes. However the electricity consumed by a household is single-phase, and each home is connected to one of the three phases of a low voltage transformer; either phase A, phase B, or phase C.

A key problem faced by energy distributors world wide is the ability to maintain an accurate record of which house is on which phase. In some cases the phase may be recorded when connecting a new house, however this is not always possible due to the inaccessibility of the distribution transformer. Furthermore, in the case this may be recorded the information typically deteriorates over time due to maintenance and repair. While there exist manual techniques for identifying phase and solutions based on signal injection techniques, these approaches have thus far not been adopted due to the challenges in costs and effort required. Knowing customer phase is important for a number of different reasons, most important of which is phase balancing - the loads on the

three phases of a transformer must be balanced for grids to be efficient. By identifying which house is on which phase, the load may be evenly rebalanced between the phases.

Automated monitoring and control in grids has traditionally been done in the high voltage transmission network portion of the grid. More recently the medium and low voltage networks have begun to be instrumented with intelligent monitoring and control devices that report in real-time on the electrical behavior of the network. In addition, energy distributors have commenced upgrades of manually-read analogue household meters with automated smart meters that communicate meter readings with greater frequency back to the distributors. Collectively these initiatives form the basis of many smart grid transformations that energy distributors are undertaking.

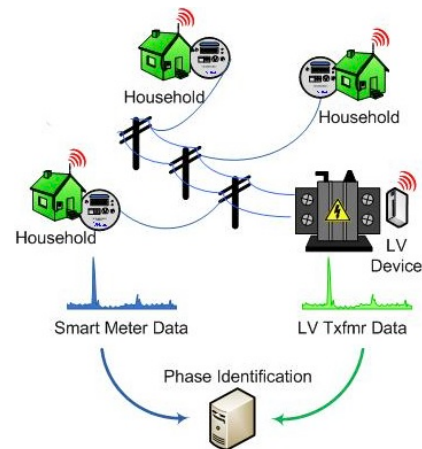


Fig. 1. System Overview

In this paper we propose a novel technique [1] for household phase identification that takes advantage of the recent drive for digitization of power grids [2]. Our approach derives a solution using a time series of discrete power measurements taken at the households and at the distributing transformer. The measurements are used to set up a system of linear equations based upon the principle of conservation of electric charge *i.e.*, energy supplied by a feeder must be equal to the energy consumed by all the households connected to that feeder plus errors. The errors arise due to imperfect synchronization of measurements at homes and transformers, and unknown and time-varying line-loss conditions. The equations are analyzed to determine an assignment of homes to phases that optimally fits the measurements. As illustrated in Fig. 1, data collected from the transformer and household meters is transported to a server that executes phase identification algorithms. We view the main contributions of this paper as follows:

1. A new technique for household phase identification is proposed based upon mathematical optimization.
2. We develop integer programming formulations for noiseless and noisy variants of the phase identification problem. We propose continuous relaxations of these integer programs that may be used with increasing amounts of measurements to retrieve household phase efficiently.
3. The methods, tools, and implementation using mathematical programming are outlined together with our discussion of errors, uniqueness of solution, and experimental results.

The rest of the paper is organized as follows. Section II presents motivation and related work. The phase identification approach based on a time series of power measurements is introduced in section III. Section IV presents the mathematical models and solutions techniques based on optimization. Section V discusses experimental results. We conclude in section VI with key observations and opportunities for further work.

## II. MOTIVATION AND RELATED WORK

While utilities usually have knowledge of the grid topology at its core, accurate information regarding the connectivity at the customer edge of the network is variable. There is considerable motivation to ensuring an accurate record of phase per household. These are based on improving the efficiency of the electrical network, extending lifetime of assets, and facilitating the infusion of renewable energy within the grid.

In a 3-phase system, three feeders carry three alternating currents which reach their instantaneous peaks at different times. These phases must be balanced - the magnitudes of current and voltage must be equal for all three phases - for the grids to be efficient. Unbalanced feeders not only increase power losses and the risk of overload, but also affect power quality and electricity prices. Imbalances also lead to overheating and consequently, shorten the lifespan of the grid assets such as transformers [3]. Typically the operator may manually compute the load endured by a transformer as a collective 3-phase load. This unfortunately does not highlight that one phase may be experiencing a significantly higher load in comparison to the other phases. Depending on customer demands and how they are assigned to different phases, loads on the three phases of a transformer can constantly remain unbalanced. By identifying their phase, households may be reassigned to a different phase so that load is evenly balanced between the phases, thus reducing power loss and improving the operational efficiency of the network.

A further motivation for phase identification is to facilitate introduction of distributed energy generation [4] at the households. The excess energy generated at the households can be injected back into the network over one of the three phases. Hence the need to determine phase is important to ensure a balanced infusion of power into the grid. The reasoning for balancing this is similar to the distribution of energy in the other direction - from the grid to the households.

The literature on phase identification at the household level is limited. Caird [5] discloses a system and method for phase

identification with suitably enhanced automated meters that can detect phases based upon a unique signal injected into the phase line. The disadvantage of signal injection methods is that they require enhanced hardware to transmit and receive special signals at different points of the grid, increasing capital and maintenance costs.

Our approach on the other hand, doesn't require any additional hardware other than household and transformer meters. Moreover, there is no requirement for interventions through signal injection or physical access to record measurements.

Dilek's [6] work on phase prediction in circuits is similar in spirit to our approach. The author employs a Tabu search on power flow measurements to determine the phase of various loads. However there are a number of differences. Unlike [6], our work discusses different types of errors, optimal solutions for both noiseless and noisy versions of the problem, uniqueness of solutions, and relaxations that can be used to obtain solutions efficiently with increasing number of measurements. Whereas the approach of [6] is tested only on a few loads, we present experimental results for larger number of homes.

## III. DESIGN METHODS AND TOOLS

### A. Phase Identification Approach

In our approach, household phase is determined using a time series of synchronized measurements collected from homes and the transformer. The principle of conservation of electric charge implies that during any time interval, the total load on a phase equals the sum of demands of customer households on that phase. Since these demands vary with time and across different households, customer phases can in fact be recovered by analyzing home and transformer load measurements over several time steps.

Time Interval	Households (Wh)						Phases (Wh)		
	H1	H2	H3	H4	H5	H6	A	B	C
(0, 10]	2	3.5	1	2.5	4.5	4	5.5	1	11
(10, 20]	3	3	10	5.5	7	1.5	6	10	14
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

The above table shows an example time series of power measurements taken from homes and the distributing transformer over intervals of  $\Delta t = 10$  minutes each. For instance, the first row shows that during the first 10 minutes home H1 consumed 2 watt-hours (Wh) while the total power supplied by phase A is 5.5 Wh. We wish to determine household phases from the above measurements. Now observe that, at each time step, loads of homes 'H1' and 'H2' sum up to the load on phase 'A'. Similarly, columns 'H4', 'H5', and 'H6' sum up to column 'C'. Column 'H3' is same as column 'B'. Thus we conclude that {H1, H2} must have been connected to Phase A, {H3} to B, and {H4, H5, H6} to C.

In the above example, at each time step, the total load on a phase is *exactly* equal to the sum of loads of homes connected to that phase. In practice however, due to line losses, synchronization errors, and other errors discussed next, the load measurement at a phase is only *approximately* equal to the sum of load measurements of homes on that phase.

## B. Measurement Setup and Errors

Consumer smart meters can record and report periodic measurements of power consumed in watt-hours (Wh) over small time intervals of  $\Delta t = 15$  or 30 minutes, as setup by the utility (For e.g. Itron smart meters used by CenterPoint Energy in US record Wh over 15 min intervals). Government regulations require that the watt-hours reported by consumer meters be accurate, typically of the order of 99.5% accuracy. A meter records readings based on its internal clock and this clock may be out of synch with respect to the true clock. For e.g. if a meter reports that 75Wh were consumed from 10:00:00 to 10:15:00 AM and its clock lags the true clock by 1 sec, in reality the 75Wh were consumed from 10:00:01 to 10:15:01 AM. Therefore even if all consumer meters are setup to report over the same time intervals, each may suffer from a different clock drift and report Wh consumed over a slightly different time interval.

Meters deployed at transformers are much more complex devices. Unlike consumer meters, they measure several fine-grained parameters needed to monitor a transformer, such as voltage, power factor, fault analysis parameters, etc. Typically the meters publish average values of parameters over small time intervals. Therefore the watt-hours computed from these parameters for each phase are estimates of the actual watt-hours supplied and may contain errors. (The *real* power (Wh) supplied by a phase can be computed as a product of rms voltage, rms current, and power factor measurements). In addition, similar to a consumer meters, clock synchronization problems may also occur at the transformer meter.

Other sources of errors include line losses and unmetered loads. Since power lines connecting transformer phases to homes possess a certain amount of electrical resistance, some of the transferred energy is lost as heat. These losses vary with ambient temperature, load, age of the feeder, etc. Transformers may also have unmetered loads such as street lights which affect measurements taken during the night.

## IV. PHASE IDENTIFICATION TECHNIQUE

### A. Mathematical Model and Problem

The phase identification problem is modeled as follows. Let  $m$  be the number of measurements taken over time and let  $n$  be the number of homes. Let  $C = \{1, \dots, n\}$  be the set of indices for customer homes,  $J = \{a, b, c\}$  the set of indices for phases, and  $K = \{1, \dots, m\}$  the set of indices for measurements. Let  $x_{ij}$  be an indicator variable which determines the connectivity of home  $i$  to phase  $j$  i.e.,

$$x_{ij} = \begin{cases} 1 & \text{if home } i \text{ is connected to phase } j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where  $i \in C$  and  $j \in J$ . Since each home is connected to exactly one of the three phases, we have

$$x_{ia} + x_{ib} + x_{ic} = 1 \quad \forall i \in C \quad (2)$$

Let  $h_{ki}$  denote the (possibly erroneous) measurement at home  $i$  in the  $k$ th time step. Let  $p_{kj}$  denote the (possibly erroneous) measurement at phase  $j$  at the  $k$ th time step. Then the principle of conservation of electric charge implies the following relationship:

$$\sum_{i=1}^n h_{ki} x_{ij} + e_{kj} = p_{kj} \quad \forall k \in K \quad \forall j \in J \quad (3)$$

where  $e_{kj} \in \mathbb{R}$  is the error in the  $k$ th measurement corresponding to the summation for phase  $j$ .  $e_{kj}$  compensates for the difference between the sum of home measurements and their phase measurement arising due to errors discussed in III-B. Thus the model allows errors to vary across measurements.

We will use the following matrix form to express constraints (2) and (3). Let  $H = [h_{ki}]_{m \times n}$  be the matrix of home measurements. Let  $X_j = [x_{1j}, \dots, x_{nj}]^T$ ,  $P_j = [p_{1j}, \dots, p_{mj}]^T$ ,  $e_j = [e_{1j}, \dots, e_{mj}]^T$ . Define  $M$ ,  $X$ ,  $P$ , and  $E$  as follows.

$$M = \begin{bmatrix} H & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & H & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & H \end{bmatrix} X = \begin{bmatrix} X_a \\ X_b \\ X_c \end{bmatrix} P = \begin{bmatrix} P_a \\ P_b \\ P_c \end{bmatrix} E = \begin{bmatrix} e_a \\ e_b \\ e_c \end{bmatrix} \quad (4)$$

Let  $D_3 = [I_n \ I_n \ I_n]$  where  $I_n$  is an  $n \times n$  identity matrix. Let  $\mathbf{1}_n$  denote an  $n \times 1$  vector of all 1's. Then (2), (3) give

$$D_3 X = \mathbf{1}_n \quad (5)$$

$$M X + E = P \quad (6)$$

The *phase identification problem* is to determine the unknown binary phase assignment vector  $X \in \{0, 1\}^{3n}$  given  $H$ ,  $P$ , and unknown  $E$  from (5) & (6). The following sections discuss noiseless and noisy variants of the above problem and show how they could be solved.

### B. The Noiseless Problem and Solution

With no errors, the loads on phases exactly match the sum of loads of respective homes, i.e. the vector  $E = \mathbf{0}$  in (6).

1) *Single Measurement*: For one measurement from a single time step, if the meter readings from homes and transformers can be converted to integers without loss of accuracy, the problem reduces to a variant of *Subset-sum* problem [7]. Given the Wh for each of the three phases, we are interested in finding three disjoint subsets of homes, such that the Wh of homes within each subset sum up to the Wh of a different phase. The subset-sum problem is NP-hard and can be solved in pseudo-polynomial time using a dynamic program. However, one measurement may not yield a unique solution.

2) *Multiple Measurements*: In the general case when we have a series of measurements, (5) & (6) can be combined as:

$$A = \begin{bmatrix} M \\ D_3 \end{bmatrix} \quad B = \begin{bmatrix} P \\ \mathbf{1}_n \end{bmatrix} \quad (7)$$

Thus we have a system of constrained linear equations

$$\begin{aligned} (\text{ILP1}) \quad & AX = B \\ & x_{ij} \in \{0, 1\}, \quad \forall i \in C, \forall j \in J \end{aligned} \quad (8)$$

(8) is a 0-1 integer linear program (ILP) with zero objective function. It can have multiple solutions, especially when the number of measurements (i.e. constraints) is low. CPLEX's [8] MIP (mixed integer programming) solver can be used to obtain a solution to (8). However, ILPs are NP-hard and therefore some instances of (8) may require exponential time. We now

propose two relaxations that can be used to retrieve the unique solution to (8) in polynomial time given sufficient number of measurements.

i) *Linear systems relaxation:*

$$AX = B$$

$$\cancel{x_{ij} \in \{0, 1\}} \quad x_{ij} \in \mathbb{R}, \forall i \in C, \forall j \in J \quad (9)$$

(9) is an unconstrained relaxation of (8) wherein we have dropped the integrality constraints on  $X$ . It can be solved using linear algebra provided  $A$  has full rank. When this holds, the linear equations (9) and the ILP (8) both yield the same unique binary solution. This is because the ground truth that generates these measurements is in fact binary and the full rank condition implies that the system has a unique solution.

The matrix  $A$  has  $3n$  columns and  $3m + n$  rows. However only  $2m + n$  rows are independent. This is because it suffices to find the phase assignment corresponding to any two phases, the unassigned homes would have to be connected to the third phase. Therefore  $\text{rank}(A) = \min\{3n, 2m + n\}$ . This implies that when  $m = n$ ,  $A$  has full rank so that phases of all homes can be recovered simply as  $X = A^{-1}B$ .

ii) *Linear programming (LP) relaxation:* First we transform the binary variables in (8) from  $\{0, 1\}$  to  $\{-1, 1\}$  using the transformation  $Y = 2X - \mathbf{1}_{3n}$ . Note that  $X \in \{0, 1\}^{3n} \Leftrightarrow Y \in \{-1, 1\}^{3n}$ . Therefore we have the following ILP that is equivalent to (8):

$$AY = 2B - A\mathbf{1}_{3n}, \quad Y \in \{-1, 1\}^{3n} \quad (10)$$

To solve the above, we use the following LP relaxation:

$$(LP1) \quad \text{Min} \quad \|Y\|_1$$

$$\text{s.t.} \quad AY = 2B - A\mathbf{1}_{3n}$$

$$\cancel{Y \in \{-1, 1\}^{3n}} \quad Y \in [-1, 1]^{3n} \quad (11)$$

The objective function of LP1 is not strictly linear due to  $L_1$  norm, but it can be linearized using standard LP methods. The following lemma relates (11) to (10).

*Lemma 1:* If (11) returns an integer solution, that solution is the unique integer solution of both (11) and (10).

We omit the proof due to space constraints. The proof follows from the fact that the fractional solution  $\in [-1, 1]^{3n}$  has a lower  $L_1$  norm than an integer solution  $\in \{-1, 1\}^{3n}$ .

Note that the converse of Lemma 1 is not true. (11) can return a fractional solution even if (10) has a unique integer solution. However, in [9], authors show that the probability that a system of the form (11) returns a unique integer solution is high. In particular this probability rapidly approaches 1 for  $m > n/2$ . In section V, we show the same using experiments.

*Solution Space:* Fig. 2 shows the solution space for the noiseless case. Given one measurement, the problem reduces to a variant of subset-sum. Given a few measurements, we would need to solve the ILP (8) that is NP-hard and may yield multiple solutions. When the number of measurements exceeds  $n/2$ , it is highly likely that we can use the LP relaxation (11) to retrieve the unique integer solution in polynomial time. When  $m = n$ , we can retrieve the unique solution simply by solving the system of linear equations (9).

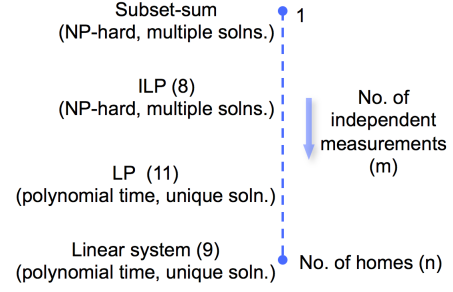


Fig. 2. Noiseless Problem: As the number of measurements increases, the phase identification problem can be solved efficiently.

### C. The Noisy Problem and Solution

This is the general setting wherein errors can vary across measurements due to loss of synchronization, line losses, and factors discussed in III-B. We now propose two approaches to retrieve a solution to the noisy variant of phase identification problem (5)-(6). We determine a phase assignment solution vector  $X$  that optimally fits the measurements by minimizing either the  $L_1$  or  $L_2$  norm of the error vector  $E$ . As long as errors do not grow significantly with measurements, the approaches below will retrieve the true underlying phase assignment solution with increasing number of measurements.

*Approach 1) Min  $L_1$  norm: Integer Linear Program (ILP)*

$$(ILP2) \quad \text{Min}_X \quad \|P - MX\|_1 \quad (\text{Linear function})$$

$$\text{s.t.} \quad D_3 X = \mathbf{1}_{n \times 1} \quad (\text{Linear Constraints})$$

$$X \in \{0, 1\}^{3n} \quad (\text{Integer constraints}) \quad (12)$$

*Approach 2) Min  $L_2$  norm: Integer Quadratic program (IQP)*

$$(IQP1) \quad \text{Min}_X \quad \|P - MX\|_2^2 \quad (\text{Quadratic function})$$

$$\text{s.t.} \quad D_3 X = \mathbf{1}_n \quad (\text{Linear constraints})$$

$$X \in \{0, 1\}^{3n} \quad (\text{Integer constraints}) \quad (13)$$

Solutions to ILP2 and IQP1 can be obtained using CPLEX's MIP solver. The choice of IQP1 vs. ILP2 depends on errors and computational performance. For gaussian errors, IQP1 yields the maximum likelihood estimate (MLE) of the true underlying solution. In terms of computation, IQP1 is quadratic with  $3n$  integer variables whereas ILP1 is linear, however with  $3n + 3m$  variables. The additional  $3m$  continuous variables are needed to linearize the  $L_1$  norm in the objective function of ILP1.

1) *Multiple Solutions:* For a fixed number of measurements, ILP2 and IQP1 may yield multiple optimal solutions and the problem of checking if a given optimal solution is unique is NP-hard [10]. Also when the error is high, ILP2 and IQP1 may yield optimal solutions with objective function value lower than that of the true underlying solution. Therefore a practical approach is to find all phase assignment solutions that lie in the close vicinity of the optimal. Let  $g(X)$  denote the objective function of ILP2/IQP1 and let  $X^*$  be an optimal solution. Then one could find all solutions  $X$  with a relative optimality gap of  $\alpha = \{g(X) - g(X^*)\}/g(X)$ .  $\alpha$  can be set close to 0, for e.g 5%. CPLEX's MIP solver allows the retrieval of multiple solutions within a given optimality gap.

2) *Continuous Relaxations LP2 and QP1* : We now present relaxations that may be used to retrieve the phase assignment solution efficiently with increasing number of measurements. We replace the integer variables  $X \in \{0, 1\}^{3n}$  with their continuous counterparts  $X' \in [0, 1]^{3n}$  in (12) and (13) to obtain relaxations denoted by LP2 and QP1 respectively. LP2 is a linear program and QP1 is the constrained least squares problem [11]. Both LP2 and QP1 can be solved efficiently. Given a fractional solution  $X' \in [0, 1]^{3n}$  obtained via LP2/QP1, we round it back to an integer solution  $X \in \{0, 1\}^{3n}$  as follows. Let  $j_1$  be such that  $x'_{ij_1} = \max_{j \in C} \{x'_{ij}\}$ . We set  $x_{ij} = 1$  for  $j = j_1$  and  $x_{ij} = 0 \forall j \neq j_1$ . This is repeated  $\forall i \in C$ . The rounding method ensures that every home is assigned to exactly one phase.

We demonstrate via experiments that as the number of measurements increase, rounded solutions of relaxations LP2/QP1 essentially coincide with optimal solutions of ILP2/IQP1.

## V. EXPERIMENTS

We summarize the main results of our experiments. We conducted several Monte-Carlo simulations using three different datasets: *dataset1* of  $n = 250$  homes, with meter readings generated uniformly at random, *dataset2* of  $n = 140$  homes, with meter readings constructed from actual consumption profiles of Canadian homes collected every  $\Delta t = 15$  min by [12], and *dataset3* of  $n = 100$  homes with anonymous meter readings collected every  $\Delta t = 30$  min by an Australian utility. In each dataset, homes are randomly assigned to different phases in order to generate phase measurements. The mathematical programs of Sec. IV are solved by invoking CPLEX [8] from within MATLAB [13]. We compare the optimal solutions output by these programs with the true underlying phase assignment solutions as a function of number of measurements. Results are plotted over multiple runs of experiments.

1) *Noiseless Problem*: In this case, we test the success rate of relaxation LP1 (11), *i.e.* how often LP1 yields the true integer phase assignment solution with increasing amounts of measurements. Fig. 3(a) shows the benchmark results using *dataset1* over 100 runs of experiments for  $n = 50$  and 250 homes. The blue curves (left) plot the results when homes are evenly assigned to each of the 3 phases with probability  $1/3$  while the red curves (right) plot the results for uneven phase assignments. We observe that when the number of measurements  $m < n/2$ , LP1 often yields fractional solutions and therefore the success rate is low. However when  $m > n/2$ , LP1 very often yields the true integer solution and its success rate rapidly approaches 1. These results concur with those of [9].

2) *Noisy Problem*: We introduce three types of errors in the datasets: (a) Gaussian, (b) Gaussian random walk, and (c) Clock skew errors. (b) & (c) mimic clock synchronization errors while (a) represents errors due to a combination of factors discussed in Sec. III-B, *e.g.* line losses, synchronization, etc.

(a) *Gaussian errors*: In this case, every home and phase meter reading  $r$  is made erroneous by introducing a gaussian error whose standard deviation is proportional to the size of

the reading, *i.e.*, the erroneous reading  $r' \sim \mathcal{N}(\mu = r, \sigma = fr)$ ,  $f \in [1, 5]\%$ . A large load therefore implies a large error.

For integer programs, we observed that IQP1 (13) is much more slower than ILP2 (12). Table I shows benchmark results for ILP2 using *dataset3* over 25 runs of experiment. In

m	Optimal = True phase assignment?			Avg. time (sec)		
	Error: 1%	2%	3%	1%	2%	3%
2n	100%	96%	76%	10	50	263
3n	100%	100%	100%	13	29	116

TABLE I

GAUSSIAN ERRORS 1-3%: PERFORMANCE OF ILP2, *dataset3*,  $n=100$  each case ILP2 yields a unique optimal solution. When no. of measurements is twice the no. of homes ( $m = 2n$ ) and the error is high (3%), for 24% of all runs, the optimal solution had a lower objective function value than that of the true underlying phase assignment solution. However as the number of measurements increase to  $3n$ , the optimal solution always coincides with the true underlying phase assignment solution. The right column shows the mean time taken by CPLEX on a linux system (Intel T9400 2.53GHz processor, 3GB RAM). Since measurements constrain the set of possible solutions, fewer measurements with high error result in a larger search space and hence more search time.

In case of relaxations, both QP1 & LP2 (Sec. IV-C2) yield solutions without any perceptible delay. However when measurements become really large (*e.g.*  $10n$ ), LP2 is a slower compared to QP2 since LP2 has more variables. Fig. 3(b) & (c) plot the quality of rounded solutions output by QP1 for 50 and 250 homes respectively using *dataset1*. The blue curves show the success rate, *i.e.* the fraction of times the rounded solution coincides with the true underlying phase assignment, for different error rates as a function of increasing number of measurements. The red curves show the average fraction of homes assigned the true phase over multiple runs. As the measurements increase, the rounded solutions coincide with the true phase assignment solution. Since each meter reading is made erroneous, as the number of homes increases, so does the total error. Therefore both more homes and more errors require larger number of measurements to retrieve the true household phase. Fig. 3(d) plots the results for LP2 with  $n=250$  homes. We observe that LP2 takes more measurements than QP1 to retrieve the true solutions. This is not surprising given that IQP1 is the MLE for gaussian errors. Fig. 3(e) & (f) show similar benchmark results for *dataset2* and *dataset3*. In both cases,  $7n$  measurements were sufficient to retrieve the true phase assignments with up to 3% gaussian errors.

(b) *Gaussian random walk*: To simulate synchronization errors, we assume that each meter's clock has errors behaving as a gaussian random walk. Instead of clocking the load after every  $\Delta t$  units, the  $k$ th measurement clocks the load for the interval  $[T_{k-1}, T_k]$  where  $T_k = T_{k-1} + \mathcal{N}(\mu = \Delta t, \sigma = f\Delta t)$ ,  $f \in [0.5, 1.5]\%$ . For *e.g.*, for  $\Delta t = 15$ min &  $f = 0.5\%$ ,  $\sigma = 4.5$ sec. This is a very high error; for real clocks  $\sigma$  is a few milliseconds. Fig. 3(g) shows the benchmark results for QP2 using *dataset2*. Despite the errors, as measurements increase, QP2 eventually retrieves the true phase assignments.

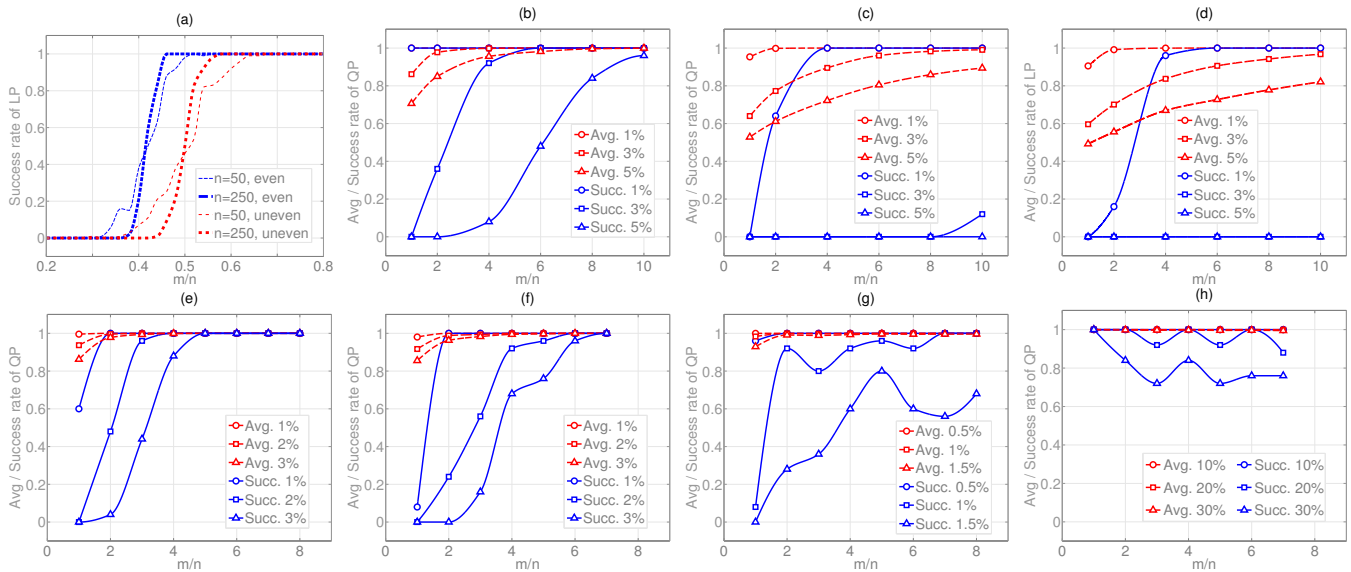


Fig. 3. Performance of mathematical programs: (a) No Errors, LP1, dataset1, (b) Gaussian errors, QP1, dataset1,  $n=50$ , (c) Gaussian errors, QP1, dataset1,  $n=250$ , (d) Gaussian errors, LP2, dataset1,  $n=250$ , (e) Gaussian errors, QP1, dataset2,  $n=140$ , (f) Gaussian errors, QP1, dataset3,  $n=100$ , (g) Gaussian random walk errors, QP1, dataset2,  $n=140$ , and (h) Clock skew errors, QP1, dataset3,  $n=100$ .

(c) *Clock Skew*: Clock skew results when a clock’s frequency differs from that of the true clock. To compute the frequency of a meter’s clock with skew, we assume that the total time difference between the meter’s clock and true clock after  $10n$  measurements lies uniformly in the interval  $[-f\Delta t, f\Delta t]$ ,  $f \in [10, 30]\%$ . For e.g., for  $\Delta t = 30\text{min}$ ,  $n = 100$  homes,  $10n$  measurements take  $\sim 21$  days. If  $f = 10\%$ , after 21 days the true and meter clocks could differ by as much as 3 min. This is a very high skew error; after tens of days, the time difference between a true clock and a real clock with skew is in the order of milliseconds. With these errors, Fig. 3(g) shows the benchmark results for QP2 using dataset3. For errors as high as 10%, correct phase assignments are retrieved. But for higher errors, we see a reverse trend: QP2 retrieves correct solutions for less measurements and incorrect solutions for more measurements. This reverse trend occurs when errors grow significantly with measurements. As long as errors remain bounded with increasing number of measurements, correct solutions are eventually retrieved.

## VI. DISCUSSIONS AND FUTURE WORK

In this paper, we presented a novel approach for phase identification in smart distribution grids. The measurements from meters at homes and the distributing transformer are collected into a linear system of equations; and an inverse solution is desired. We present a collection of integer programs and their continuous relaxations, which can be used to compute household phase with increasing number of measurements. We systematically explore the behavior of these mathematical programs as a function of different types of errors. Our experiments demonstrate that as long as the errors do not grow significantly with measurements, household phase can eventually be retrieved given sufficient measurements.

Future work will investigate the performance of our ap-

proaches using real metering data obtained from both transformers and homes as well as explore the performance of MIP solvers with larger number of households. A related issue is to determine confidence levels for rounded solutions of relaxations in terms of number of measurements. We also wish to handle the case of missing measurements. Lastly, phase identification is the first step towards the larger problem of phase balancing and we shall investigate this problem taking into account various costs associated with phase re-balancing.

## REFERENCES

- [1] V. Arya, S. Kalyanaraman, D. Seetharam, V. Chakravarthy, K. Dontas, J. Kalagnanam, and C. Pavlovski, “Systems and Methods for Phase Identification,” Patent Pending 13/036,628, Feb. 2011.
- [2] T. Standish, “Digitization of the electric grid: The enabler of energy efficiency and economic gain,” <http://www.centerpointenergy.com/static/files/CNP/Common/SiteAssets/doc/>.
- [3] J. Zhu, M. Y. Chow, and F. Zhang, “Phase Balancing using Mixed-Integer Programming,” *IEEE Transactions on Power Systems*, vol. 13, no. 4, pp. 1487–1492, November 1998.
- [4] D. Roberts, “Local Power: Tapping Distributed Energy in 21st-Century Cities,” *Scientific American*, June 2010.
- [5] K. Caird, “Meter Phase Identification,” US Patent Application 20100164473, January 2010, Patent No. 12/345702.
- [6] 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.
- [7] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, “35.5: The subset-sum problem,” in *Introduction to Algorithms*. MIT Press, 2001.
- [8] IBM ILOG, “CPLEX: High-performance software for Mathematical Programming and Optimization,” <http://www-01.ibm.com/software/integration/optimization/cplex-optimizer>.
- [9] O. L. Mangasarian and M. Ferris, “Uniqueness of Integer Solution of Linear Equations,” *Optimization Letters*, pp. 1–7, 2010.
- [10] P. M. Pardalos and S. Jha, “Complexity of Uniqueness and Local Search in Quadratic 0-1 Programming,” *Operations Research Letters*, vol. 11, no. 2, pp. 119–123, 1992.
- [11] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], 1981.
- [12] “Integration of Micro-Generation and Related Energy Technologies in Buildings,” <http://iea-annex54.org/annex42/data.html>.
- [13] MATLAB, version 7.8.0.347 (R2009a). The MathWorks Inc., 2009.