# DECLUTTERING GEOGRAPHIC DATA VIEW DISPLAYS

Esa M. Rantanen and Limor Hochberg[1], Rochester Institute of Technology, Rochester, NY
Mingyang Di[2] and Diego Klabjan, Northwestern University, Evanston, IL
[1]Now at UL-Wiklund, Concord, MA. [2]Now at American Express

To reduce clutter on wide-area geographic data view displays of electric power systems, substations in geographically compact areas should be spread out and line overlaps and intersections minimized. Such patterns optimized with respect to given constraints can be modeled as a multicommodity flow problem. Due to the size of the developed model, we developed two clustering-based iterative algorithms to decompose the global network into smaller regions and then iteratively solve the subproblem in each region by readjusting some values. Three solutions were selected for experimental evaluation. The experimental results were somewhat inconclusive, due to naïve participants, simple task, and limitations of the eye tracker. Nevertheless, two of the decluttered maps appeared to have distinct advantage over the baseline map, where substations were simply connected by straight lines, in terms of performance time and workload.

## INTRODUCTION

Following the August 14, 2003 Northeast U.S. blackout there has been increasing demand to broaden the views of power system control centers to also encompass neighboring control areas. The rationale behind wide-area displays is to allow operators earlier warnings of possibly cascading events that may originate outside their area of responsibility and disrupt the grid over large geographical areas (Klump, Dooley, & Wu, 2003). These system visualizations are known as geographic data views (GDVs). With the GDV approach, power system visualizations can be dynamically created using power system information along with geographic information embedded in the system model.

Although advantageous in many ways, dynamic GDVs suffer from some disadvantages. For example, when the task switches from monitoring to corrective control or analysis, it can be difficult to design *a priori* a single display, or even a set of displays, that contain all the information needed to make effective, corrective, control decisions. Displaying such information on an existing GDV designed for system monitoring could result in slow display performance and a cluttered appearance. Hence, there is a continual need to develop novel display solutions to meet the operators' needs.

Two main problems that degrade the readability of displays and that are to be avoided in display design are visual "noise", which may result from careless use of animation and color, and clutter from displaying too much information in too small a display space (Fig. 1). We define clutter as areas of high information density within a display, to the extent that it impairs readability. These present-day challenges are further aggravated by the future need to display predicted values and visualizing dynamic information.

The GDV displays show power transmission lines and substations laid on a geographic map; consequently, certain areas (e.g., in and around cities) become extremely cluttered. To reduce clutter, closely located substations could be spread out (distorting the true geographic relationships in a carefully controlled manner) and transmission lines between them drawn, for example, in rectilinear patterns, so as to minimize line overlaps and crossings. Such patterns optimized with respect to given constraints can be created by a number of operations research (OR) techniques.

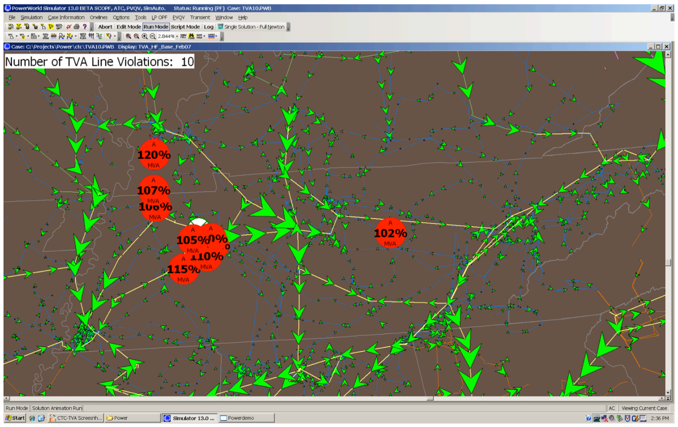

*Figure 1*. An example of a cluttered dynamic GDV display. In this simulator-screen capture, the arrows move to indicate power flow and are proportionally sized to display the amount of current the line is carrying. The red pie charts show line overloads.

### Display Design Challenges

Minimizing clutter will create room for other information to be displayed in innovative ways without overwhelming the human operators using the display. Given the number of buses and other devices, substations, and transmission lines, the design of wide-area displays presents a daunting challenge. Another challenge is the large geographic areas covered by individual control centers. This dual challenge makes design of wide-area displays layouts by hand nearly impossible. Therefore, we sought to solve the problem computationally.

We developed a linear integer programming model for the task. Due to the size of the problem, however, a commercial solver could not solve it to a satisfactory quality within a reasonable computational time. As a result, further research was required to develop effective and tailored solution methodologies for solving the model. Different Lagrangian relaxation techniques were examined. Another option was to reformulate the model by using so-called path variables. Such a reformula-

Copyright 2018 by Human Factors and Ergonomics Society. DOI 10.1177/1541931218621375

tion requires a branch-and-price algorithm due to an excessive number of variables.

To our knowledge, mathematical optimization of power systems GDV display layouts has never been done before. Optimization of display layouts can nevertheless be seen as fundamental to all other display innovations and operators' ability to effectively use the displayed information.

### Human Factors Considerations

Geographic relations are very important to power systems operators. Many critical system parameters depend on geographical distances between generators and substations and intervening in cascading failures necessitates mastery of systemwide geography. The question therefore is how to best combine geographic and other relevant information for operators so that it is available to them at all times and for all possible situations.

Our search for relevant, existing research on optimal layouts for wide-area GDV displays resulted in very few published works. To the best of our knowledge, the problem has never been researched in operational electric power control context. The operators' task solving line problems is very much like maze-solving. Results from a few laboratory studies that have used maze-solving tasks converge to suggest design criteria for GDV displays.

Fixation durations have been shown to increase as a linear function of the length and number of turns in the path segment between the current and the upcoming fixation points, suggesting that much mental processing relates to the length and turns in that segment (Crowe et al., 2000). Chafee et al. (2000) found a positive linear correlation between number of turns in a maze-solving task and response time. In yet another experiment (Crundall, Cole, & Underwood, 2008) the length of a curved line segment rather than the Euclidean distance between two targets determined the response time of comparing two targets along the line.

These results suggest that power lines depicted on GDV displays should not have turns in them, for turns make their visual tracing more difficult and time-consuming. Other potential source of confusion while visually tracing lines is the number of lines crossing each other. Self-evidently no lines or substation symbols should overlap, which would make it impossible to differentiate between them on a visual display. We also concluded that preservation of the spatial *relationships* between substations was paramount, even if the distances between them would necessarily be distorted to declutter displays around densely populated areas.

### Basic Decluttering Algorithm

A map of electric transmission lines was generated and the lat./long. coordinates of substations (transmission lines are straight lines between these) transformed so that the following criteria were met:

(1) The lines should be either vertical or horizontal so that (a) intersections between lines and (b) turns in them are minimized. The conditions may be mutually exclusive, in which case they may need to be defined hierarchically.

(2) Different substation symbols and lines must not overlap. In other words, all display elements must be visible to the operators at all times regardless of the display scale (zoom).

(3) The relative positions between substations must be preserved; if there is a substation A and another substation B north and east of it, after the algorithm is done moving the substations around, B should still be north and east (or, above and to the right) of A.

(4) The magnitude of positional change for the substations from natural geographical to the schematic coordinates per the above constraints must be minimized. This problem may be solved iteratively by a computer.

## DECLUTTERED MAP CREATION

The NYISO State Electrical System Map represents the network of existing and proposed stations and transmission lines connecting them. Most stations are transmission stations; the others are generating stations. Transmission lines depicted were above 115 kV. The PDF copy of the system map was converted to TIFF format via a software package. It was then uploaded into another software application, GetData Graph Digitizer. Using this program, coordinates were logged for each existing transmission and generating stations.

The decluttering problem turned out to be more difficult than we anticipated and its size, even for just one state (NY) area, required breaking it down to smaller pieces (a grid) manageable by the computer. A cluster-based modeling approach was used to decompose the problem into smaller regions in the grid. An iterative algorithm was implemented to solve the problem in each region by readjusting selected parameters.

We created 8 different solutions. The parameters varied to create these solutions were penalization for turns (T), line crossings (C), overlapping line segments (O), and stations deviating from their geographical coordinates (D). See Table 1 for the parameter values for the solutions and Figures 2 and 3 for the baseline (original) and a decluttered map.

Table 1.
*The Objectives (Parameters: Line Turns = T, Line Crossings = C, Overlapping Line Segments = O, and Deviation from True Geographic Coordinates = D) and Their Values (Penalties) for Producing Alternative Decluttering Solutions.*

| | Parameter | | | |
|---|---|---|---|---|
| Solution | T | C | O | D |
| 1 | 1 | 1 | 1 | 1 |
| 2 | 100 | 1 | 1 | 1 |
| 3 | 1 | 100 | 1 | 1 |
| 4 | 1 | 1 | 100 | 1 |
| 5 | 1 | 1 | 1 | 100 |
| 6 | 1 | 1 | 1 | 100 |
| 7 | 1 | 1 | 1 | 0.01 |
| 8 | 1 | 1 | 1 | 100 |

*Note:* Solution 6 was produced with a finer grid than the other solutions, as well as by using a heuristic algorithm.

Since complexity drops dramatically in a heuristic approach, a finer (larger) grid may be created to improve the quality of the solution, but that trades off against a more accurate algorithm. Both will improve the quality of the solution. However, due to the computational power currently available, both of them cannot be achieved simultaneously.

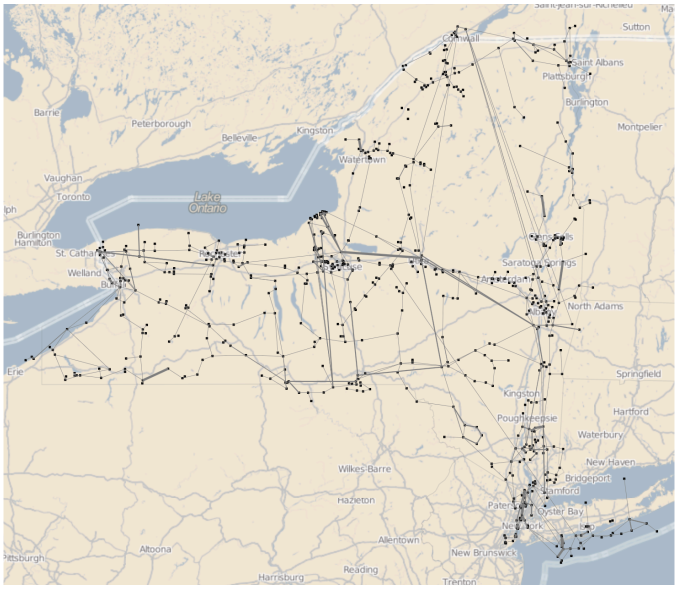

*Figure 2*. The baseline map was created by simply connecting the true geographic locations of the substations with straight lines representing transmission lines between substations. That the substations on Long Island are "off" is due an uncorrected error caused by the Earth's curvature when the grid is overlaid on a geographic map.

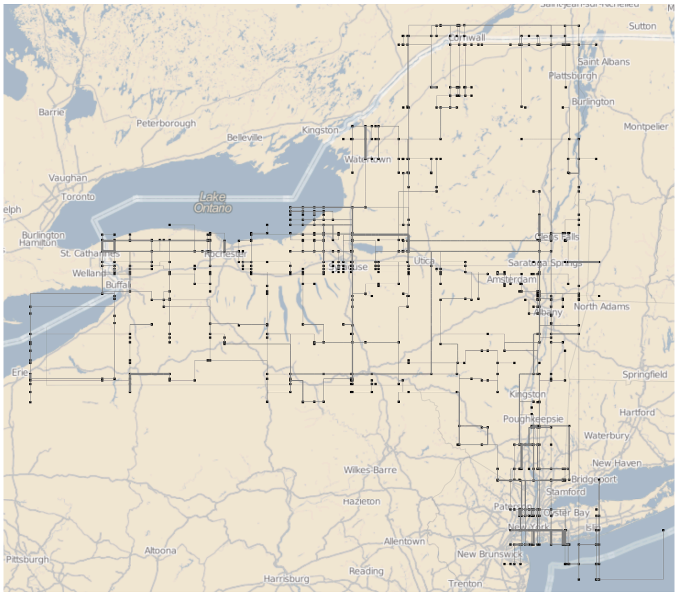

*Figure 3*. Solution 1, created using an accurate algorithm and a sparse grid. All objectives, line turns (T), line crossings (C), line segment overlaps (O), and deviations (D) from true geographic coordinates are equally penalized: T=C=O=D=1.

Within this research, solutions 1–5 used the more accurate algorithm (with solutions 2–5 penalizing different objectives) while solutions 6 and 7 create a larger (finer) grid and used a heuristic. Solution 8 used a less accurate heuristic algorithm and a smaller grid. Thus, it is just a benchmark. See Di, Klabjan, and Rantanen (2014) for further details.

## EXPERIMENTAL EVALUATION: METHOD

### Participants

A total of 20 participants, recruited from the student population at RIT, 15 male and 5 female, with a mean age of 24.5 (SD = 6.23) years, yielded good data for analysis.

### Apparatus

We used a SensoMotoric Instruments (SMI) RED250 remote eye tracking system coupled with a Dell 2210 TFT LCD 22-inch diagonal monitor. The viewable area of the monitor measures 474 × 296 mm and has a resolution of 1680 × 1050 pixels (a pixel pitch of 0.282 mm/pixel). The display was driven at full resolution at 60 Hz. Observers were seated at a viewing distance of approximately 70 cm in a chair with fixed position to minimize body movement. Mean angular subtense of the full monitor at mean viewing distance was 37.4 × 23.9 degrees of visual angle, yielding an angular resolution of approximately 45 pixels/degree. The system recorded gaze at 250 Hz with a spatial accuracy of approximately 1 deg. Participants could make natural head movements within a 20 × 20 × 10 cm head volume, and the system quickly re-acquires gaze if the participant moves out of the volume or looks away from the stimulus display. Prior to the experiment, an automated 9-point calibration was run for each participant.

### Experimental Materials

Out of the 8 solutions created, we chose solutions 1, 3, and 4, as well as the baseline map with straight lines connecting substations for the experiment. We wanted to limit the maximum time participants would perform the task without becoming fatigued. Solution 1 was another "baseline" solution with equal penalizing of all parameters. We skipped solution 2 as it appeared very confusing, and in our judgment line crossings were more detrimental than turns. Solutions 5–8 were not chosen because the task was abstracted, deviations were not relevant to the naïve participants.

The experimental maps were created by a Python program that read the substation and line coordinates as translated by the decluttering algorithm and the different penalties assigned to the parameters. The experimental maps contained a central section the New York state and had 1680 × 1050 pixel resolution for the eye tracker display. We did not expect the participants to recognize the geography from these maps, and geography was irrelevant to the experimental task in any case. Figure 4 depicts one experimental display.

### Experimental Task

The participants were presented an experimental map with a green circle and a blue square, as well as one line segment highlighted in red. They were instructed to visually trace a path from the green circle to the blue square *without* going through the red line segment and noting where two line segments were connected (at substations marked by black dots) and where they were not (no dots at line crossings).

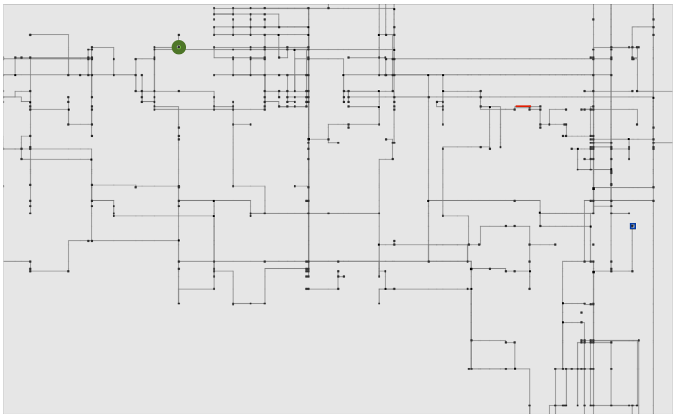

*Figure 4*. An experimental map based on solution 1, where turns, line crossings, line overlaps, and substation (marked with black dots) deviations from their true geographic locations were equally penalized. The participants' task was to visually trace a path from the green circle to the blue square through connected line segments while avoiding the line segments highlighted in red.

### Experimental Design

Independent variable was the weight of penalties for turns, crossings, and overlaps of transmission lines on the experimental maps, compared to the baseline map with straight lines between substations. We measured several dependent variables for analyses:

(1) The time to complete the task (i.e., visually tracing the path between the two highlighted points with- out going through the line segment highlighted in red); the shorter the time, the easier the map is to read.

(2) The time participants to orient themselves on the map before starting the experimental task; shorter times suggest an easier map to read.

(3) Amplitude (length) of saccades; longer saccades indicate an easier map to read (Rayner, 2009).

(4) Peak velocities of saccades; higher saccade peak velocities indicate an easier map to read (Di Stasi et al., 2010).

(5) Durations of fixations (or dwell times); shorter dwell times indicate an easier map to read (Rayner, 2009).

(6) Pupil dilation is an indicator of mental workload\; smaller pupils indicate an easier map to read.

(7) Subjective ratings of the difficulty of the task using each map.

This was a within-subjects design, each participant performing the experimental task on each of the four maps, plus a similar task (different only by the start and end points and the line segment marked "out") on a replicate for each map. Analyses of the resulting data compared the four maps along several dependent measures.

### Procedure

The participants were given 3 practice trials with a map that was different from the experimental maps (solution 5). After both the participant and the experimenter were satisfied that the participant understood the instructions and was able to perform the task, experimental maps were presented in a counterbalanced order. Each map was replicated once, with different highlighted nodes and the "open" line segment. After each trial and blank screen was displayed offering the participants a rest break before continuing. At the end of the experiment, a short questionnaire was administered to the participants and their responses recorded.

## EXPERIMENTAL EVALUATION: RESULTS

We performed linear mixed effects (LME) analyses on all dependent variables (Winter, 2013). All graphics were created and analyses performed with R, a free software environment for statistical computing and graphics (R Core Team, 2013); the package *lme4* provided functions for fitting and analyzing mixed models (Bates et al., 2014). Tukey contrasts were used for multiple comparisons of means (Hothorn, et al, 2008).

### Orientation and Task Times

We used a natural logarithm (ln) transformation on the data to restore normality of the distributions before performing analyses on these data. Arranged by the median orientation time no map appeared to be superior to others in terms of the time needed to find the highlighted nodes and the "open" line segment on the map and to begin the experimental task. The task performance times show some differences. Map 1 appears to have the shortest task performance time and map 0 the longest The LME analysis confirmed this. The effect of map was statistically significant, $\chi^2(6) = 22.46$, $p < 0.001$, and pairwise comparisons indicated that Map 1 had significantly lower task performance time than Map 0 ($p < 0.001$) and Map 3 ($p < 0.001$).

### Saccade Amplitude and Velocity, Pupil Diameter

Longer saccades indicate an easier map to read. Although Map turned out to have had a statistically significant effect, $\chi^2(6) = 27.78$, $p < 0.001$, the only significant difference was between Map 0 and Map 3, and this difference was very small. Saccade peak velocity and pupil diameter did not yield statistically significant results.

### Fixation Durations

Shorter fixation durations would indicate an easier map to read. The advantage seems to go to Map 4, which had consistently (also smallest SD) smaller fixation durations than other maps. However, although the effect of map was significant in the LME analysis, $\chi^2(6) = 35.102, p < 0.0001$, only Map 3 was significantly different from Map 0 ($p = 0.04$) and Map 1 ($p = 0.037$). These effects are quite weak, too.

Table 2
*Summary Table of Experimental Results.*

| Variable | Advantage | Effect Size | Sig. |
|---|---|---|---|
| Task Performance Time | Map1 | 4s[1] | Yes |
| Saccade Amplitude | None | < 0.1 deg[2] | No |
| Saccade Peak Velocity | Map0,1 | 7deg/s[3] | No |
| Fixation Duration | Map4 | 24ms[4] | No |
| Pupil Diameter | Map4 | 0.01 mm[5] | No |

*Notes*: [1]The difference between the median task time for Map 1 and the second fastest task time (Map 3). [2]The maximum difference in median degrees of visual angle between any two maps. [3]The difference between saccade peak velocities for Maps 0 and 1 (0.55 deg/s apart) and the next closest map (3). [4]The difference in median fixation duration between Map 4 and the next smallest median duration with Map 0. [5]The difference with the map with smallest pupil diameter and the map with next smallest pupil diameter (Map 0).

## Limitations

Our participants were mostly students at RIT who had little or no experience in reading wiring diagrams and no knowledge of power systems. Therefore, the experimental task had to be abstract and simple, so that naïve participants could perform it without much training. Although the task was simple, it was still too complex for the capabilities of the eye tracker we used. In other words, even in a simple line-tracing task the participants' eyes moved in very complex patterns across the display of the experimental maps and the resulting variability in the data likely masked subtle differences in their performance on the different maps.

## DISCUSSION

Table 2 summarizes the experimental results. It is difficult to draw firm conclusions from the eye movement data, but qualitatively some maps appear to have been more advantageous than others in our experimental task. In particular, Map 1, where line turns, line crossings, overlapping line segments, and deviations from true geographic coordinates, were equally penalized seemed to have had perhaps the most robust advantage in terms of time to complete the task, which in turn was probably the best measure of performance. In a short task, an average of 4 s time advantage is a reasonably large effect. All the experimental (decluttered) maps seem to have had the same advantage over the baseline maps in terms of task time. This result was corroborated by the participants' responses to the post-experiment survey, where the majority preferred the more rectilinear structure of the experimental maps to the baseline map with straight lines connecting the nodes, which they thought to be visually confusing. Map 4, which penalized overlap of map elements (nodes and lines) appeared to have had small advantage over the other maps in terms of fixation duration and pupil diameter. This seems reasonable as Map 4 appears the "clearest" of all maps generated with most space between the map elements.

The goal of this project was to provide an optimized layout for geographic data view (GDV) displays and electric system visualizations. The optimization algorithm we developed indeed achieved this goal, despite the fact that the task turned out to be much more difficult (and as such, also much more interesting) than we anticipated, requiring substantial computational resources. Future research on this topic should be able to produce a truly optimal GDV display layout for electric power systems control applications. However, a successful outcome will require close collaboration between researchers and experienced operators serving as subject matter experts (SMEs). A more sophisticated simulation capability than what was available to us will also be required to thoroughly test the GDV layouts and other visualization solutions in realistic settings and with realistic problems.

## ACKNOWLEDMENTS

This research was supported by a grant from the New York State Energy Research and Development Authority (NYSERDA). Mr. Michael Razanousky was the project manager. The views expressed in this paper are those of the authors and do not necessarily reflect those of NYSERDA. We also thank Dr. Jeff Pelz at RIT for his invaluable assistance with the experiment.

## REFERENCES

Bates, D. Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7. http://CRAN.R-project.org/package=lme4.

Chafee, M. V., Averbeck, B. B., Crowe, D. A., & Georgopoulos, A. P. (2002). Impact of path parameters on maze solution time. *Archives Italiennes de Biologie*, *140*(3), 247-251.

Crowe, D. A., Averbeck, B. B., Chafee, M. V., Anderson, J. H., & Georgopoulos, A. P. (2000). Mental maze solving. *Journal of Cognitive Neuroscience*, *12*(5), 813-827.

Crundall, D., Cole, G. G., & Underwood, G. (2008). Attentional and automatic processes in line tracing: Is tracing obligatory?. *Attention, Perception, & Psychophysics*, *70*(3), 422-430.

Di Stasi, L. L., Renner, R., Staehr, P., Helmert, J. R., Velichkovsky, B. M., Cañas, J. J., ... & Pannasch, S. (2010). Saccadic peak velocity sensitivity to variations in mental workload. *Aviation, Space, and Environmental Medicine*, *81*(4), 413-417.

Di, M., Klabjan, D., & Rantanen, E. (2014, November). Power grid visualization by means of optimization. In *INFORMS Annual Meeting*, San Francisco, CA.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical J., 50*(3), 346–363.

Klump, R., Dooley, G., & Wu, W. (2003). Displaying aggregate data, interrelated quantities, and data trends in electric power systems. *Proc. 36th Hawaii Int'l Conference on System Sciences* (HICSS'03). IEEE.

R Core Team (2013). R: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of ExperimentalPpsychology*, *62*(8), 1457-1506.

Winter, B. (2013). *Linear models and linear mixed effects models in R with linguistic applications* (arXiv:1308.5499). http://arxiv.org/pdf/1308.5499.pdf