How can the effectiveness of marketing 'Airbnb Seattle' be improved? – dataset of 2016

LIST Nicole , LÖHR Tim,
BOHNSTEDT Timo, PANG Tsz Ching, and BAL Kiran Jeniffer

***Abstract*—This paper presents a theoretical analysis of harmonically-terminated high-efficiency power rectifiers and experimental validation on a class-C single Schottky-diode rectifier and a class-F^{-1} GaN transistor rectifier. The theory is based on a Fourier analysis of current and voltage waveforms which arise across the rectifying element when different harmonic terminations are presented at its terminals. An analogy to harmonically-terminated power amplifier theory is discussed. From the analysis, one can obtain an optimal value for the DC load given the RF circuit design. An upper limit on rectifier efficiency is derivedce-pull measurement of a Schottky diode rectifier with short-circuit terminations at the second and third harmonic are presented. A maximal device rectification efficiency of 72.8 impedance for self-synchronous rectification. Measurements of conversion efficiency and output DC voltage for varying gate RF impedance, DC load and gate bias are shown with varying input RF power at the drain. The rectifier demonstrates an efficiency of 85% for a 10 W input RF power at the transistor drain, with a DC voltage of 30 V across a 98 Ω resistor.**

***Index Terms*—kaggle, machine learning, business, data analytics, airbnb**

I. Project background and motivation

THE dataset we are going to analyze contains several information about the renting of apartments via the platform Airbnb in Seattle. By analyzing the data, we want to find possibilities to improve the marketing of Airbnb optimized for Seattle. Mentioning the word 'marketing' most people immediately think about advertisement. Indeed, this is a very important part of marketing but it is far not enough to cover the whole meaning of marketing. By definition it rather is about the firm's effort to address customer needs as well as their expectation and to orient their products according to the requirements of customers. So, it is also about the tradeoff of 'evoking' needs and expectation of consumers on a level that the product can satisfy and is able to compete with competitive products. To fulfill this, applying marketing instruments like the marketing mix can be helpful. The marketing mix consists of four policies: Product, Price, Promotion and Place. In the following the tasks and aims arising in these four areas will be described and explained how our data analysis can help to improve tasks in these areas aiming for a better overall marketing.

As Airbnb is an agent for lessors, who want to rent their apartment to tourists, it ears its money by receiving a commission for every rented apartment. Therefore, Airbnb should aim for a high booking rate to increase their own profit. That is one reason, why Airbnb should care which apartments are offered on their platform and how they are presented (e.g. by the description, price, ...). So, it can make sense for Airbnb to give lessors some suggestions how to promote and present their accommodation to achieve a maximal booking rate. Deducted of this assumption, we want to provide some suggestions for lessors to promote their apartments in the perfect manner but also want to give some suggestions, what point in time is best for Airbnb to release a marketing campaign to advertise some apartments in Seattle.

A. Product

As already mentioned before, marketing is about fulfilling customer needs and expectations. Within product policy the aim is to understand one's market and be able to figure out which needs and wants the customers have. In general, one can say that the main need of travelers is to find an accommodation but nowadays it is not only about finding accommodations but even more about discovering the right accommodation. It is not only having a nice and clean room bathroom, with white and clean towels and bead sheets. The surrounding and flair of the accommodation becomes more and more important. This issue Airbnb has already addressed in its advertising spot , so Airbnb is aware of the wants of travelers and is responsive to this in its advertisement. Thinking a step further, it is not enough to just show the customer that renting Airbnb apartments is a nice way to 'really live' there instead of just 'go there'. When the customer has been attracted by Airbnb to search for on apartment on their website it is important to present the apartment in a good way. Therefore, the description of the apartments should mention all aspects the customer considers as important. By our data analysis we would like to support the lessor in creating a good and appealing description. By analyzing the reviews of customers and filtering the 50 most mentioned words, we can conclude that these mentioned

Manuscript received July 10, 2012. This paper is an expanded paper from the IEEE MTT-S Int. Microwave Symposium held on June 17-22, 2012 in Montreal, Canada. This work was funded in part by the Office of Naval Research under the Defense Advanced Research Projects Agency (DARPA) Microscale Power Conversion (MPC) Program under Grant N00014-11-1-0931, and in part by the Advanced Research Projects Agency-Energy (ARPA-E), U.S. Department of Energy, under Award Number DE-AR0000216.

M. Roberg is with TriQuint Semiconductor, 500 West Renner Road Richardson, TX 75080 USA (e-mail: michael.roberg@tqs.com).

T. Reveyrand is with the XLIM Laboratory, UMR 7252, University of Limoges, 87060 Limoges, France (e-mail: tibault.reveyrand@xlim.fr).

I. Ramos and Z. Popovic are with the Department of Electrical, Computer and Energy Engineering, University of Colorado, Boulder, CO, 80309-0425 USA (e-mail: ignacio.ramos@colorado.edu; zoya.popovic@colorado.edu).

E. Falkenstein is with Qualcomm Inc., 6150 Lookout Road Boulder, CO 80301 USA (e-mail: erez.falkenstein@gmail.com).

words are important to customers and therefore lessors should mention them in their description. As Seattle is a huge city with lots of different areas with different style and flairs, we conduct word clouds for every neighborhood. This offers the advantage that we can better address and select certain customer groups. A neighborhood, with lots of parks, is probably better for nature lover than for reveler. So, nature lovers will mention the parks in their reviews very often and according to our word cloud (which is based on the reviews), lessors in this neighborhood will focus on the parks around their accommodation. If a person, who prefers partying at night, reads this description his or her interest will not be raised and therefore the person will search for apartments in another neighborhood. This creates and 'automatic' selection and increase the chance that a customer finds an apartment in a neighborhood, which fits his/ her interest. So, a first question we would like to answer is: Which facts lessors need to address in their description to raise the interest of potential customer, who 'fit' the vibe of the neighborhood and set their focus on the same aspects as former customers of these apartments? Customers' needs and wants do not only reflect in the feedback they give but also of the degree of booking of an apartment the customers preferences can be deducted. As a second question we would like to answer: Which factors did influence the degree of booking of former rented apartments? To answer this, we compute the correlation between the different attributes and the degree of booking of an apartment. According to our results lessors can see, which attributes are of great importance, when renting out an apartment in a certain neighborhood, and can design their accommodation according to our suggestions.

B. Price

The price of a product is a very important aspect regarding the marketing of a product. It, kind of reflects the customers expectation and needs to be determined on the right level. By analyzing some data, it will be a lot easier to set the right price for an apartment as in this market segment (and as we consider Airbnb) there are a lot of objects of comparison. Within our data analysis we want to screen the dataset for accommodations with a high rate of booking. Based on the information provided by these apartments, we would like to train a decision tree, that helps lessors to classify their apartment into a certain price level according to the attributes it holds. As selection of attribute we take the attributes, which show the highest correlation with the degree of booking according to our analysis within the product policy.
Furthermore, we would like to indicate the price trend of the rented apartments. So, lessors can see during which season prices increase. We also want to predict the price trend in order to give even more indication how prices should be adapted during the season.
According to the price policy we would like to answer the following two questions: Which price can be charged for an apartment with certain characteristics? When can lessors increase the price per night for their apartment and when should they lower it?
By constructing a decision tree and also predicting the price changes over the year, lessors can classify their apartment in order to find out how much they can earn by renting their apartment. By considering price variability over the year, lessors can yield an optimal return and increase their booking rate as their neither too expensive nor too cheap. This also secures the existence of Airbnb apartments can be secured as lessors do not quit to rent their home due to too low prices or too less customers, because of a too high rent.

C. Promotion

For Promotion Policy it is important to find out when advertisement should issue a marketing campaign and which content. So far, the data we analyzed mainly bring the advantage to make some suggestions to (former) lessors, which characteristics and which price their apartments should have in order to effectively rent them. But these results can become important for a marketing campaign of Airbnb. As already mentioned, the word clouds and the attributes with a high correlation with the booking rate, represent the features customers value the most and therefore a marketing campaign should address these issues.
The Promotion Policy does not only focus on the content of the marketing campaign but also when the campaign will be most effective. In order to give an indication regarding that issue, we would like to analyze the booking rate and predict it for the next year. By this it becomes clear when there will be a phase with a low booking rate. Shortly before that phase the campaign should be started in order to motivate people to book their apartment on Airbnb. *So, the question of interest is: What is a good point in time to start a marketing campaign?* According to the Promotion Policy a second issue is to decide where/ via which tools the marketing campaign should be distributed and be presented to potential customers. As our dataset does not give any information on the fact how customers got to know of Airbnb or for which reason they decided to book their accommodation on Airbnb, we decided to neglect the aspect of Promotion Policy as we can not yield any results or deduce some information, which would be helpful to decide on the distribution channel of our marketing campaign.

D. Place

The Place Policy considers where customers get in touch with the product and consume it, in order to find a suitable retail location that is accessible for customers. The first contact between lessor and renter happens on Airbnb but the final 'purchase' of the product, takes places in Seattle. As Airbnb is a platform, which acts as agent between lessor in Seattle and renter, we do not need to care about this in our data analysis, as this fact is fixed and can not be changed.

II. Data description with visualization

A. General

Our dataset consists of three excel sheet: 'listings.csv', 'calendar.csv' and 'reviews.csv'.

Listings consists of 92 attributes with 3,818 data entries. Every single row represents one apartment in Seattle that has been offered for rent via the platform of Airbnb. Calendar consists of 1,393,570 data entries and 4 columns. The dataset connects a certain time period with an apartment and indicates whether it has been rented out or been available during that period. The last dataset 'reviews.csv' contains all reviews former visitors have handed in for an apartment, it contains the reviewer's Id, name comments, the date as well as the house ID.
So, if the excel sheets are combined it can be deducted the following information: the key facts about an apartment (like the size, prize, number of beds, usable facilities, . . .), furthermore one can get a general idea about, what the apartment looks like by the description (written by the lessor) as well as by the reviews (written by former guests) and it is known when the apartment has been available or rented.

B. Data preprocessing

First of all, we computed how many different values certain attributes can have. Therefore, we computed a table that listed the attribute name as well as how many unique attribute values exist for that attribute. In a following step we replace all na-values by suitable possible values in order to be able to use the data in the following steps and analyze it in a proper way. To be able to fit our models, we replaced some textual values by numerical ones and split the dataset into a training dataset and a test dataset. The training dataset we will use to 'create' or models in order to check our created models and be able to see how good for example predictions will be and which error we need to expect we need the test data.

C. More detailed dataset description

1) Listings: The listing dataset consists of 92 columns, with attributes that describe different characteristics of the rented apartments. It describes 3818 apartments that are allocated in 79 neighborhoods. Combined with the review dataset it represents a dataset, which can be used to find some attributes, valued as most important by customers. If it is combined with the calendar dataset, a data analysis can help to determine the factors which are important to, not only attract visitors, but also to finally rent it out successfully. Especially we will use this combination to find out which attributes of the apartments did influence the price charged the most.

2) Reviews: In order to have an insight whether the reviews are written by satisfied or rather unsatisfied visitors, we computed a bar chart which shows the rating-number of reviews ratio. (Figure 1) 1 shows that most of the reviews give the highest value possible (10) and there are hardly no reviews that give a rating value lower than 8. So, in our further analysis we assume that the reviews are written in a rather positive tone and we need to keep in mind that the reviews have been written by overall satisfied tourists. In order to make sure that there is no correlation between the number of reviews and the review score of apartments in a certain neighborhood, we generated a figure which shows that relationship (Figure 2). In this figure it becomes visible that neighborhoods, whose apartments in total got 50 to 100 reviews score a mean rating of about 9.2 to 9.7. Approximately the same mean rating is also achieved by district where the apartments are rated 125 to 300 times. There is only one outlier ('University District'), which scored al pretty low mean rating value compared to all other districts. As this is the only outlier, we will neglect this in our following assumptions. To sum up, we can conclude that both figures (2 and 1) show that the mean rating for every district is relatively high and a high value is not dependent on the number of reviews for a certain district.

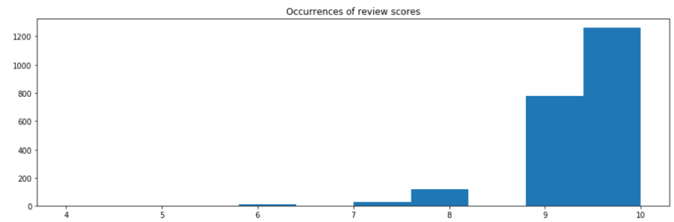

Fig. 1. score and number of reviews ratio

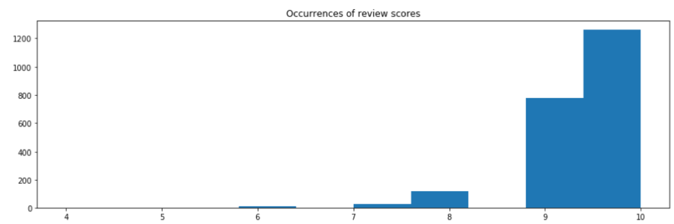

Fig. 2. mean rating and number of reviews ratio (per neighborhood)

3) Calendar: This part of the data shows, when certain apartments are available and when they are rented out. Within our analysis we assume, that an apartment, which is listed as 'not available' in the 'calender.csv' is occupied by a visitor. If it is 'available' we assume that the owners would have liked to rent the apartment out but there has not been any visitor renting it. In figure 3 the ratio of occupied and available apartments in 2016 is presented.

While in January only half of the apartments have been occupied, the number of rented apartments raised until April when a sudden drop in rented apartments appeared. The same procedure repeats in July. After that second drop the number of occupied accommodations raises again. Nevertheless, it has to be mentioned that the percentage of free apartments never exceeds 40% after February anymore but also never drops below 20%. At this point the question, whether there is a correlation

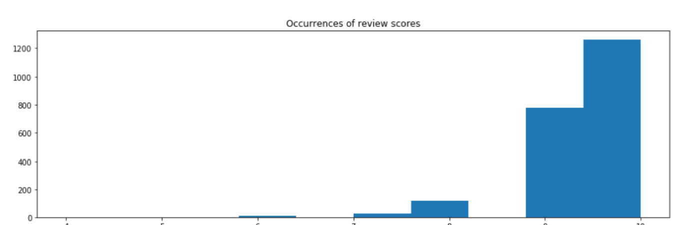

Fig. 3. occupied ratio of appartments

between the degree of booking and a ‘perfect’ description exists and therefore the percentage of free apartments can be decreased by mentioning the right aspects in the description.

III. Models and analysis

A. *Which facts lessors need to address in their description to raise the interest of potential customer, who ‘fit’ the vibe of the neighborhood and set their focus on the same aspects as former customers of these apartments?*

In order to know what words the lessors are supposed to use in the description, we take a look at the reviews with a wordcloud. The bigger the words in the wordcloud, the more often the word was used. For example, if the customers often write the word walk in the reviews due to informing what was in walking distance of the Airbnb, the work walk will be big in the wordcloud. To answer this question the six most reviewed cities Capitol Hill, Ballard, Queen Anne, Belltown, Minor, Wallingford were analyzed. First, you can see that all cities have the words walk, park and restaurant written in big. The word “downtown” is also used often. For minor, the word “lake” is very important. The rest of the words indicate that words like minute, shop, market are used often by customers. Interestingly, the word “Washington” is important for the cities Minor and Capitol Hill.

B. *Which factors did influence the degree of booking of former rented apartments?*

In order to answer this question, the Extra Tree Classifier was used. With the help of the Extra Tree Classifier one can see the feature importance. According to the feature importance the number of reviews influences the degree of booking of former rented apartments the most. Customers also care about how long the host has been on the platform and extra people. Almost equally as important are the number of reviews per month. This underlines how important the reviews are. The amount of host listings, the response rate and what is included for the guests influences the results of ones booking almost on the same level. The host should also care about the response rate and the number of beds as the next important thing on the list. Less important are the cancellation policy, bedrooms, bathrooms and if the host is verified on Airbnb. The Room type, host acceptance rate, phone verification and if the host has a profile picture are the least important factors influencing the degree of booking the apartments.

C. *5.3 Which price can be charged for an apartment with certain characteristics?*

First of all, we created a Linear Regression model in order to predict the price influenced by all other attributes. For this prediction we achieved a Mean Absolute Error of 44.28, which is, taking into account that the minimum rent is 10 dollar and maximum rent per night 1,650$, very low. Therefore, we can assume that the listed attributes have a significant impact on the price of an accommodation and it is possible to estimate the rental price by these attributes and therefore it makes sense to use a classification tree, which takes these attributes into

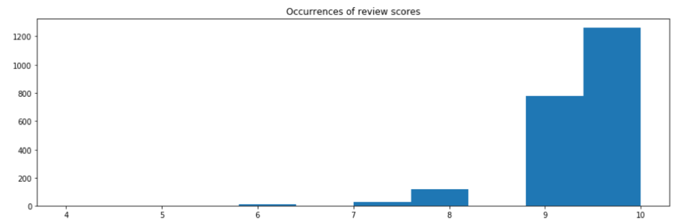

Fig. 4. correlation attribute-price

account. Figure Linear regression In the following, we use the Extra Tree Classifier in order to find out, which attributes have a strong correlation with the price and therefore influence the price of an apartment the most. A classification tree uses the values of attributes to create groups, in which all the apartments have the same (or at least similar) values for the attributes. According to our analysis, the attributes with the highest influence on the price have a correlation coefficient in a range of 0.01 and 0.2. The highest influence on the price is the number of reviews an apartment has. This attribute has a correlation of approximately 20%, which is almost twice as much as the second important attribute has by showing 0.112 as correlation coefficient. This attribute indicates how long the lessor is already registered as host on Airbnb. As third important factor, with a correlation coefficient of a little bit less than 0.1 the number of extra people influences the price. Other attributes which influence the price are for example the number of listings a host has/ had, number of beds, number of bath- and bedrooms. The least significant attribute in our visualization is, whether the guest needs a profile picture (with a coefficient of about 0.01). (Figure 4) In order to check that all the attributes only determine the price and do not have an influence on the price via another attribute, we also compute a heatmap to see the correlation of attributes among each other. This visualizes that host-response-time, host-response-rate and host-acceptance-rate correlate with each other and also the number of bedrooms, bathrooms and beds. So, these attributes groups influence each other and their impact on the price should be considered in a group. Furthermore, we can mention that the number of bathrooms, bedrooms and beds influence each other. So, a lessor should also take into account that if an apartment has for example a high number of bedrooms then also the number of bathrooms is high. According to this a lessor can also get some indication for the price he can charge. If his apartment has many bedrooms but only one bathroom, he can assume that other accommodations with the same number of bedrooms has more bathroom and therefore can also charge a higher price. So, this correlation is interesting for the lessors to get some price indication. (Figure 5) So, if a lessor wants to (re)determine the price for his accommodation, he first of all needs to consider how many reviews he got so far for the apartment. (Note: As we noted before, the reviews all give a high rating, so they are all positive. Taking this into account, it is very likely that the lessor needs to check the number of positive reviews for determining the price.) Furthermore, the lessor needs to consider how long he is already registered as host and how many people can stay in the apartment. If he

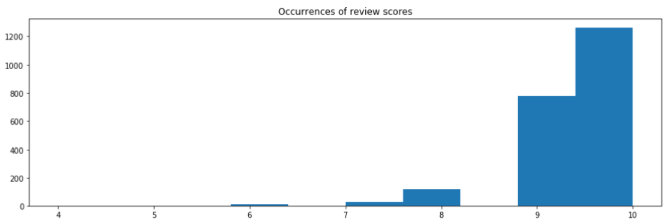

Fig. 5. correlation attribute-price

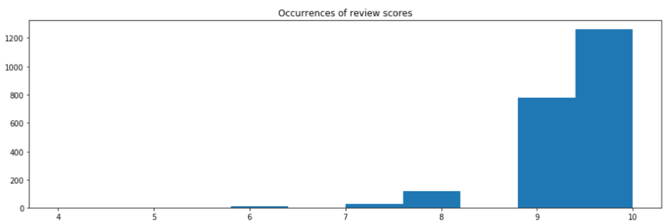

Fig. 6. number-of-bedroom-price-correlation

considers the impact of the number of bedrooms, bathrooms and beds, he should also take into account, the correlation of the three attributes among each other. As we already saw in figure 4, the number of beds, bedrooms and bathrooms do not have the highest impact on the price although somebody might expect, that the price is strongly influenced by these factors. By having a closer look on the correlation between the number of bathrooms and the price as well as the number of bathroom and price, it can be recognized that an increasing number of bed- and bathrooms does not increase the entry level price within this group but only the maximal price increases. So, it is possible to find an accommodation with a different amount of bath-/ bedroom for a price of 50 dollar per night. This changes if the number of bedrooms exceeds 5. Then there is also an increase of the entry level price visible. We can conclude that the price of an apartment is probably more influenced by the fact whether it has less or more than 5 bedrooms or not. (Figure 6) The same change can also be seen for the correlation of bathrooms and price but the change of the entry level is not that big. Therefore, the ‘strong’ change in the entry level price becomes already visible for 4 bathrooms. (Figure 7)

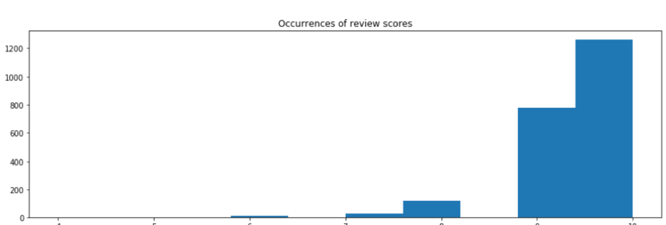

Fig. 7. number of bathrooms price correlation

D. 5.4 When can lessors increase the price per night for their apartment and when should they lower it?

To answer this question, we want to predict the price changes for 2017. First of all, we started by checking out the change of mean flat prices in 2016 (Figure 4). Prices increase from January to July form about 120 to 150 dollar and decrease until November to 135 dollar. During the last month prices slightly start to increase again. It can be seen that the correlation between charged prices and the occupation ratio is rather low, as the first drop of the occupation ratio in April does not reflect in a price drop. Nevertheless, the second decrease in the number of rented apartments does reflect in the charged prices as also the price curve decreases after July. But during the month of December, prices increase as well as the number of free apartments decrease. This irregularity might be due to the fact that many people travel during Christmas time and therefore their willingness to pay is slightly increased.

E. 5.5 What is a good point in time to start a marketing campaign?

In order to find the right point in time to start a marketing campaign, we need to know, during which periods the number of visitors decreases. To get some good results, we plotted a graph showing the number of visitors according to the time. In Figure 5.2 it can be seen that the number of visitors fluctuates very strong also within a month. This might be caused by changes within a week, as usually on weekends more people consider to travel and therefore more apartments are rented as well as public holidays influence the number of tourists. As exactly reflecting this fluctuation in a model would lead to a very complex model, which is expensive to create and the risk of a model overfit becomes very high, we use Linear Regression to reflect the trend of 2016 and in a second step predict the numbers of visitors for 2017. We can see that the number of visitors decreases between January and March and increases between March and July. After that there is a decrease until December. Nevertheless, the fluctuation is about 200 visitors. Compared to a total number of over 2400 visitors this change is quite small. The predicted number of visitors can be seen in Figure 5.3.

For 2017 it is predicted that the number of visitors raises from January to March and decreases until July. For the rest of the Year the number of visitors will raise again. So, the best point in time to start a marketing campaign would be during February till March. As the number of visitors will fall from March to July, the marketing campaign should prevent this. A campaign usually needs some time until its effectiveness can be seen so it should be started earlier than the actual decrease takes place and when the predicted decrease should take place, the effect of the campaign should keep the number of visitors on the level of March or even increase it. The campaign could be stopped in August, as its effect will last a little bit longer and between September and October the number of visitors will be on the level of March again without any campaign.

IV. CONCLUSION

In summary, this paper addresses high-efficiency power rectifiers designed with harmonic terminations at the RF input, in analogy to high-efficiency power amplifier design with harmonic terminations at the output. The applications of such power rectifiers include wireless power beaming [?], recycling power in high-power circuits [?] and ultra-fast switching integrated DC-DC converters with no magnetics [?].

The theory for an ideal rectification element is based on Fourier analysis and establishes the basic design parameters such as the relationship between output DC resistance and impedance at the fundamental frequency at the rectifier input which optimizes efficiency. The analysis also predicts the time-domain waveforms at the terminals of the rectification element and the efficiency as a function of on-resistance and DC output resistance. Specific results are derived for class-C and class-F^{-1} classes of operation, as they are defined for power amplifiers. These two cases are chosen for experimental validation with a 2.45 GHz diode and 2.14 GHz transistor rectifier, respectively. It is straightforward to repeat the derivation for other classes of operation, such as class-F as shown in detail in [?].

The experimental results show that good agreement can be reached between theory and experiment with a Schottky-diode single-ended rectifier with finite class-C harmonic terminations, resulting in 72.8% efficiency for input power levels in the mW range, intended for wireless power harvesting detailed in [?], [?]. A GaN pHEMT class-F^{-1} power rectifier achieved 85% efficiency with 40 dBm input power across 98-Ω DC load with a DC output voltage $V_{DC} = 30V$. The efficiency and output voltage of the self-synchronous rectifier are shown to depend on the input power at the drain, the impedance at the gate port and the DC load at the output drain bias line, but not on the gate bias.

Time-domain large-signal measurements of a class-F^{-1} power amplifier configured as a rectifier show that one can accomplish the same rectifier efficiency as the amplifier drain efficiency in self-synchronous mode without external gate RF drive. This is somewhat surprising, and to the best of our knowledge, the first time this type of high-efficiency rectifier has been demonstrated.

Acknowledgment

The authors would like to thank Dr. David Root and Dr. Jean-Pierre Teyssier at Agilent Technologies for the loan of the time-domain nonlinear measurement equipment and TriQuint Semiconductor for the donation of the transistors.