|  |
| --- |
| Identifying Wireless Users via Transmitter Imperfections |

|  |  |
| --- | --- |
| Authors: | Adam C. Polak, Sepideh Dolatshahi, and Dennis L. Goeckel, Fellow, IEEE |
| Publication: | IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 29, NO. 7, AUGUST 2011 |
| Speaker: | Ju-Sung Kang |

**Short summary**:

**Variations in the RF chain of radio transmitters can be used as a signature to uniquely associate wireless devices with a given transmission. Previous approaches, which have varied from transient analysis to machine learning, do not provide verifiable accuracy, which is essential for admissibility of the methods in the court. Here we detail a first step toward a model based approach, which uses statistical models of RF transmitter components that are amenable for analysis. Algorithms based on statistical signal processing methods are developed to exploit non-linearities of wireless transmitters for the purpose of user identification in wireless systems. The decision rules are derived and their performance is analyzed. In order to establish the viability of the proposed approach, the practical variations of transmitter chain components are analyzed based on simulations, measurements and manufacturers’ specifications. Results show that the proposed identification methods can be effective, even for short data records and relatively low signal-to-noise ratios, when exploiting imperfections of commercially used RF transmitters.**

# Introduction

This work concentrates on breaking criminals’ anonymity in wireless systems.

 By using the Radio Frequency fingerprints of wireless users, find the criminal’s individual devices, which cannot be identified at MAC layers.

Two main approaches of Radio Frequency fingerprinting efforts.

1) Using the channel information

- In a rich multipath environment, because of rapid path decorrelation, user can be almost uniquely identified.

- The received signal strength information (RSSI) method

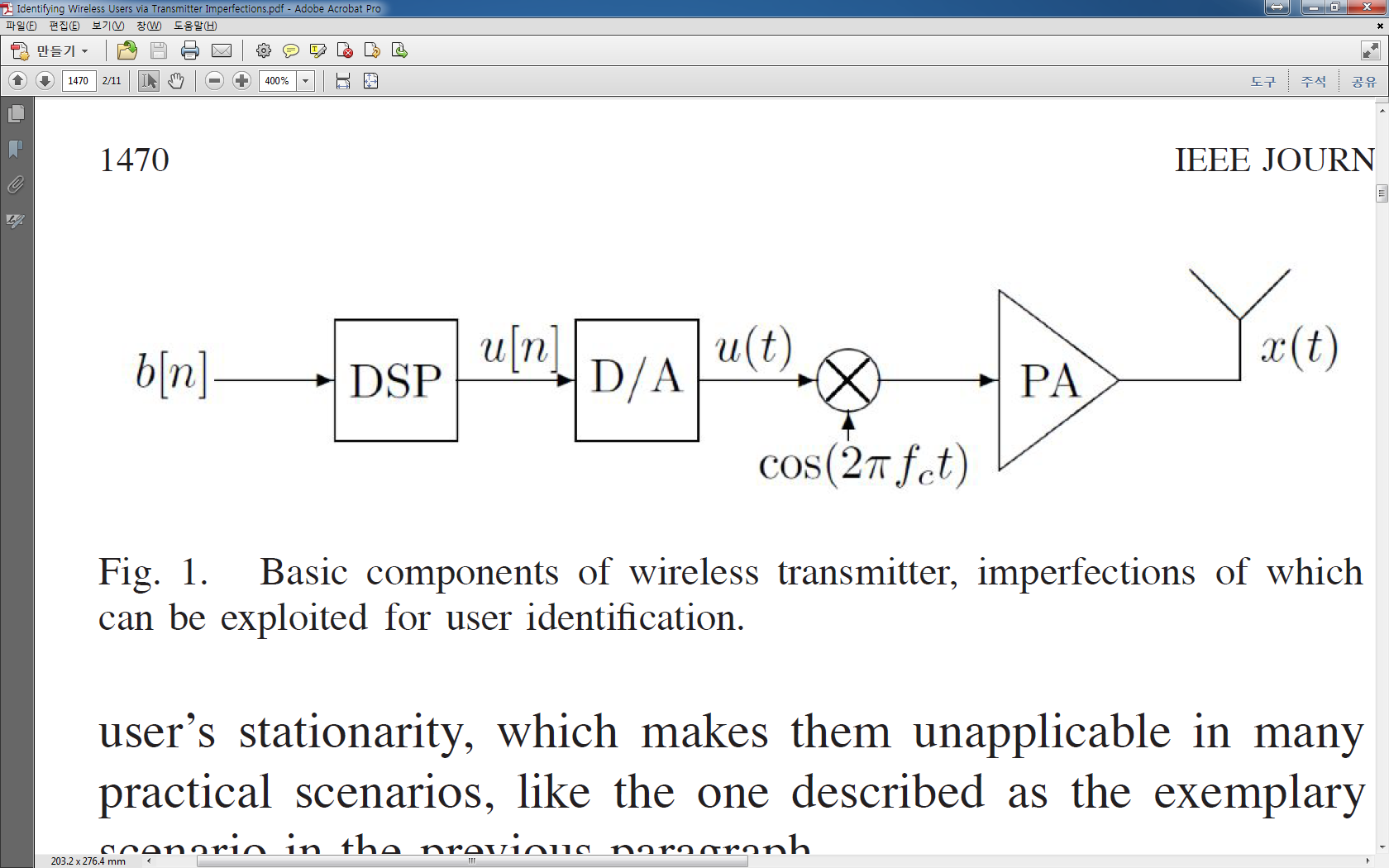
 Strong assumption : User’s stationarity.

2) Using H/W imperfections of RF devices.

- Transient signal analysis.  Too short to detect and describe.

- Other approach : Clock difference, demodulation data, etc.

The things which authors want to do is :





The approach here is focused on a comprehensive understanding and exploitation of the phenomena being expoited for node identification.  based on statistical modelling method.

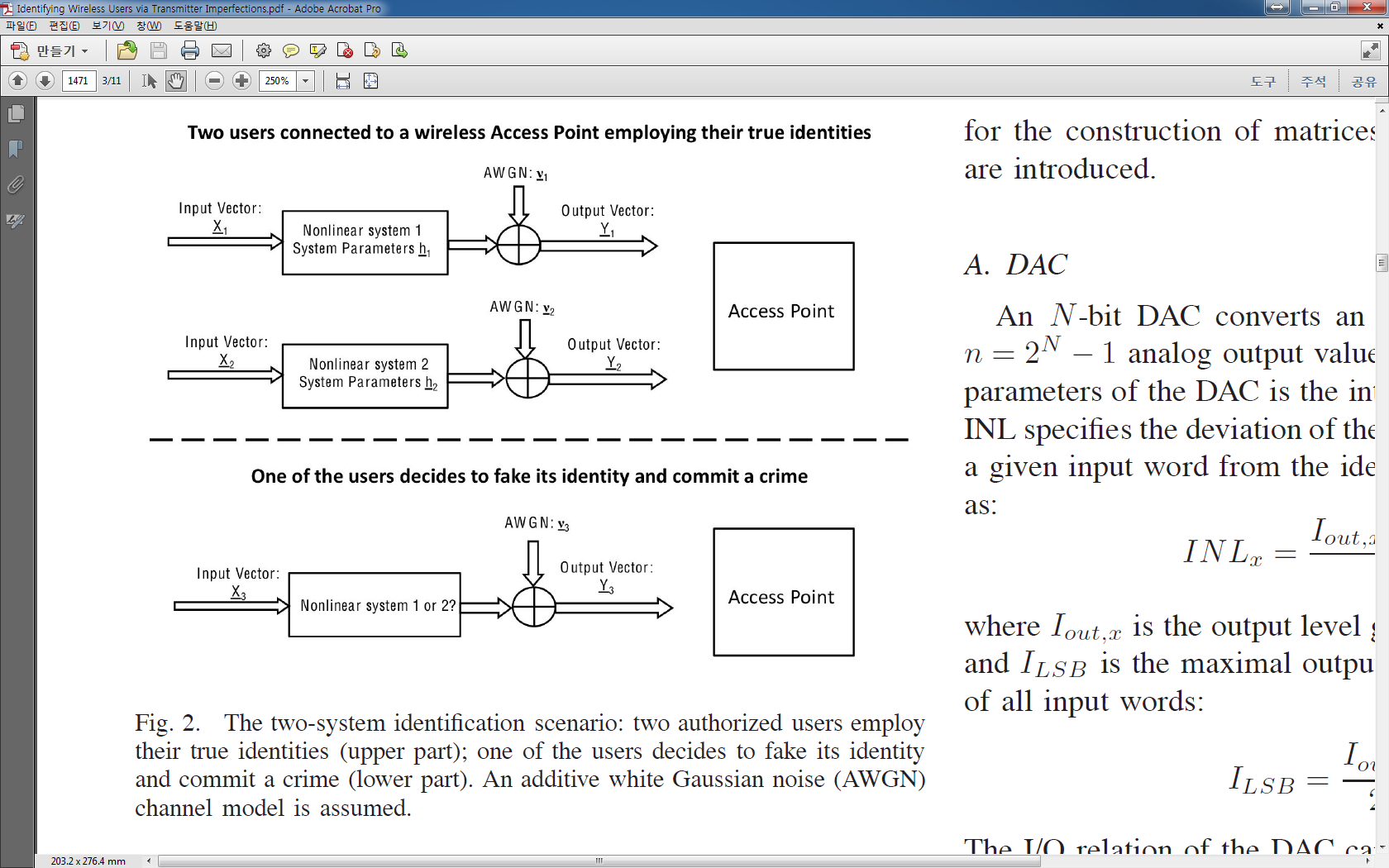
Only with the variations of communication standards, it is enough to allow the identification.

The non-idealities of digital-to-analog converter (DAC) and power amplifier (PA) are considered here.

# Problem Statement

The nonlinearities of DAC and PA can be significant across the devices.

Because of these variations, each user in a multiple user network can be characterized with a group of parameters that uniquely describe input/output (I/O) characteristics of its transmitter components.



To simplify the exposition, this work considers the two-user case, but the generalization to the case of n users is straightforward (using -hypothesis testing techniques).

Each of the users is characterized with a parameter vector . The input vectors of a considered component of users are denoted as  and , and its output are denoted as  and .  and  are the measured input and output data.

Now, the problem is ‘When vector  is transmitted so that it resulted in , which nonlinear system pass through the vector ?’  binary hypothesis problem.

# Modeling Transmitter Components

Both considered components display nonlinearities of their I/O characteristics.

In general the I/O characteristic of a given transmitter component can be described with a matrix equation of the form :

 (2)

 : a matrix, elements of which are nonlinear functions of elements of the input vector . These function are determined by the model adopted for a given type of component and are the same for all devices of that type.

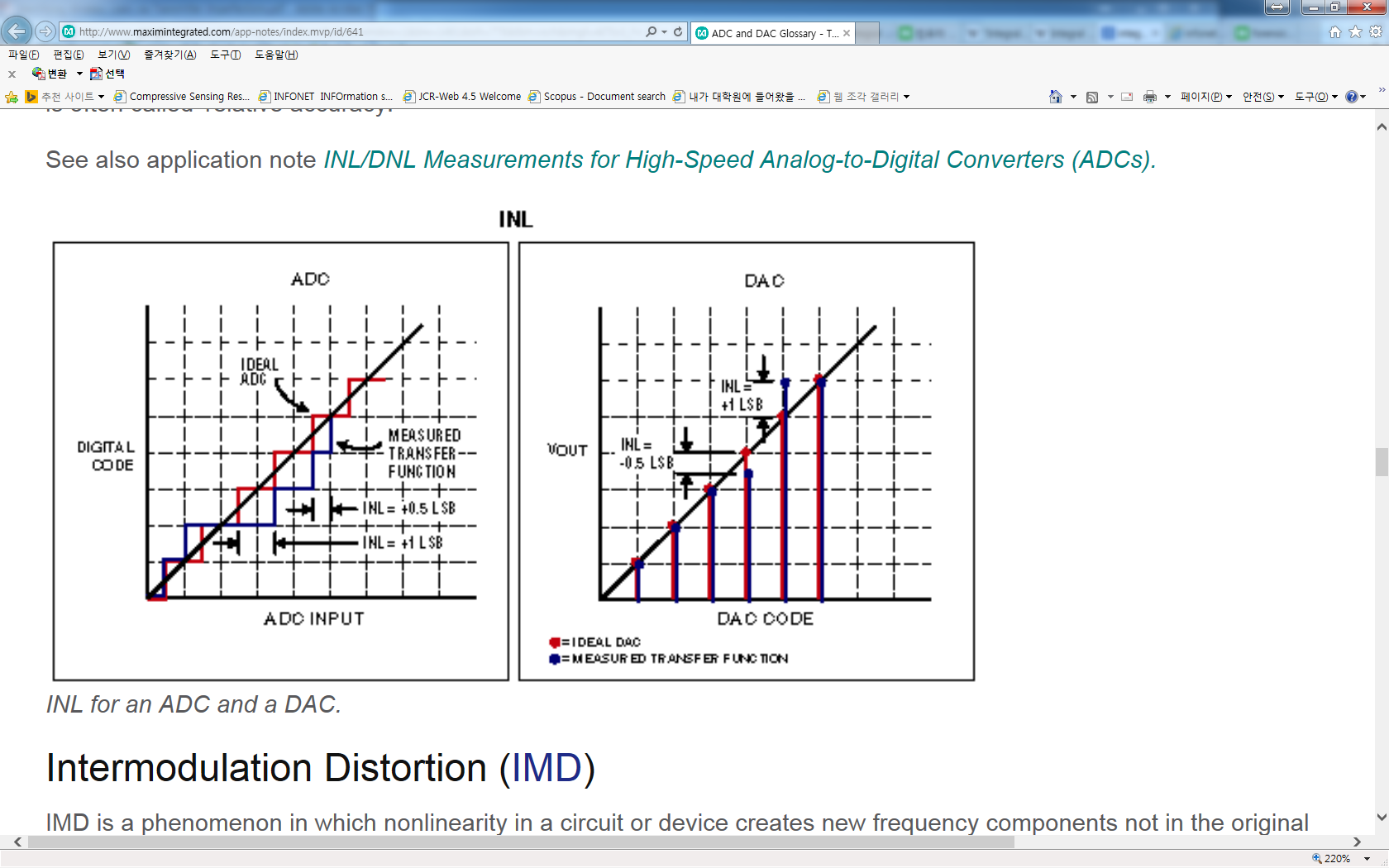
 : a column vector that contains the unique component parameters of user 

 : a AWGN vector 

## DAC

An -bit DAC converts an -bit input word to one of  analog output values.

* Integral nonlinearity (INL) : the deviation of the actual DAC’s output level for a given input word from the ideal output level



 (3)

 (4)

 : the output level generated by an input word x

 : is the maximal output level divided by the number of all input words

The INL is caused by production inaccuracies that cause output levels of the DAC to vary around their nominal values.

If individual DAC analog sources are modeled as independent normally distributed random variables with standard deviation , then the INL of DAC can be modeled with a discrete Brownian Bridge random process BB [19] :

(all analog sources have identical nominal values and each increase of the input word by one causes activation of an additional source)

 (6)

For large ,  becomes very large, and the discrete BB random process can approximated as continuous BB random process :

 (7)

*  : Wiener random process, value equal to zero for , its increments are normally distributed random variables with variance equal to the argument difference. i.e. (8).

Using the Karhunen-Loeve theorem, a continuous BB random process can be represented with its eigen-functions and eigen-values found as solutions of the integral equation [20]:

  (9)

 (10)

 (11)

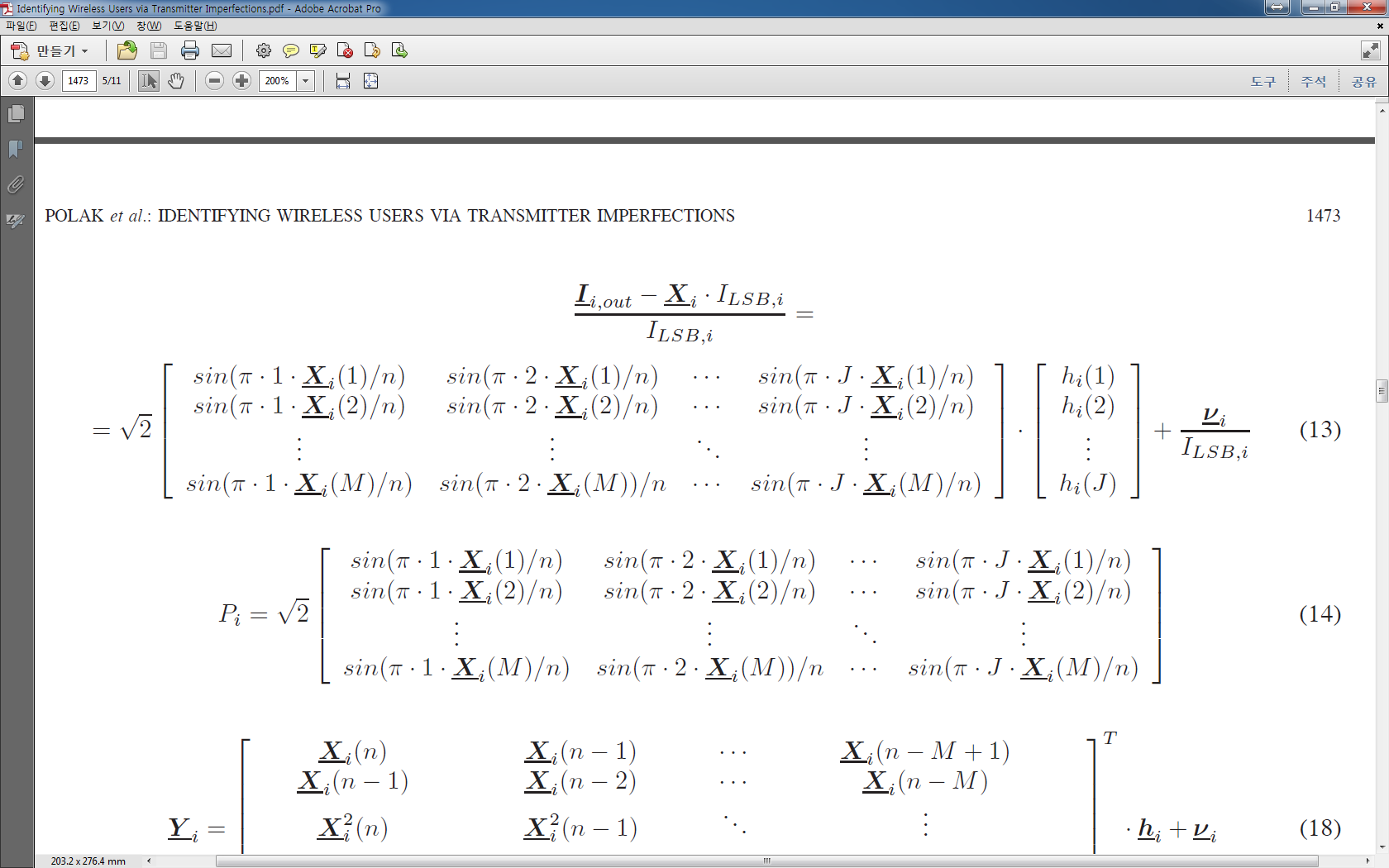
 : i.i.d. normal random variables .

Now, the I/O characteristic of DAC can be described as ()

 (12)

The equation (12) means the I/O characteristic of DAC when Access Points observe that. So, we can conclude that :  (15)

Now, we can induce the equation (13) ,(14) and finally we can obtain the form of equation (2).



 : realizations of random variables , eigenvalues of the BB random process, that uniquely describe the single INL path of user .

Because of the knowledge of ,  and relationship of nonlinearity of DAC, from the equation  the only parameters that receiver need to estimate are elements of the vector 

## Power Amplifier (PA)

Power amplifiers are attractive for digital forensics purposes in that they are the last elements of the transmitter chain and thus are the most difficult for a user to modify via software or even baseband control.

In this work the nonlinear characteristics of power amplifiers are modeled with Volterra series representations, as well-established in the microwave literature (see Chapter 4 of [21]).

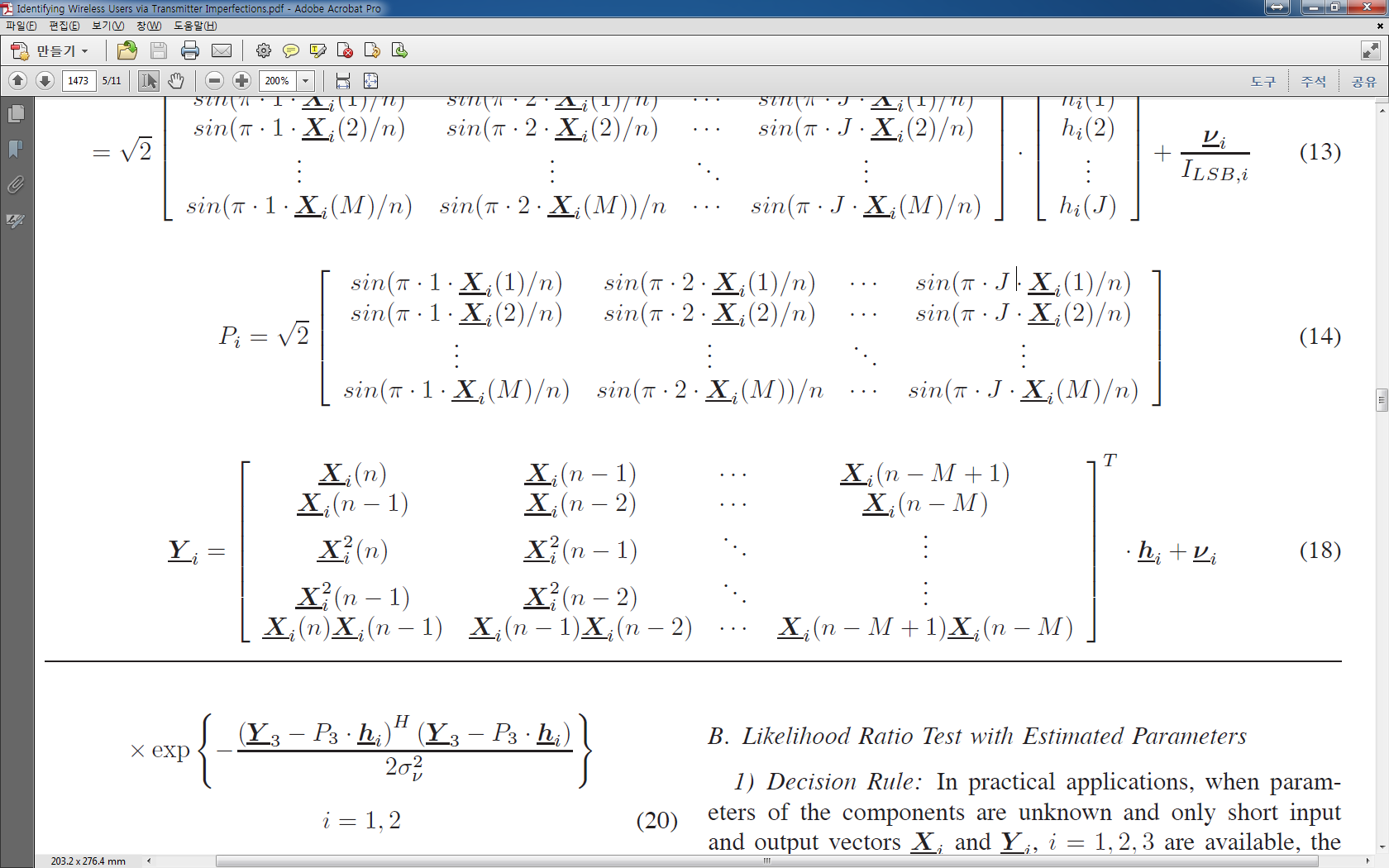
* Example : One memory, an order is two (linear quadratic system)



 : the second order Volterra kernel.

 : Volterra coefficients vector capturing the I/O characteristic of amplifier .

Now, we can induce the equation (18) which is similar with equation (2).



This problem is also simplified as estimation problem of 

# Algorithms

1) The parameter vectors  and  are exactly known in order to find an upper bound on the identifiability in the noisy channel.

 estimation method with long period,

 optimal method are known for solving the hypothesis testing problem.

2) The parameters are unknown and only short observations are available

 multiple approaches are considered.

## Likelihood Ratio Test with Known Parameter Vectors.

1) Decision rule : In this case, the probability of error of the receiver is minimized by a likelihood ratio test (LRT). :  - the measured user is user one,  - the measured user is user two.

 (19)

 is the likelihood function.

In the AWGN channel :

 (20)

Which allows to simplify the decision rule (19) to :

 (21)

2) Algorithm Performance : The probability of error is the probability that the algorithm decides for a different user than the one that decided to fake its identity and commit the crime.

 (22)

With the equal probable hypotheses and symmetry property, we can obtain :

 (23)

With (21),  can be expressed as :

(24)

With :  (25)

Under  simplifies to :

(26)

## Likelihood Ratio Test with Estimated Parameters

1) Decision Rule : When parameters are unknown and only short input and output vectors ,  are available, the decision rule from (21) with the estimated parameters is not optimal.

However it is still reasonable to use, since the result converges to the optimal rule when the parameter estimates become more and more accurate.

 (27)

And the probability of error  is :

 (28)

Where  and  are estimates of the parameter vectors  and 

Other approach : from norm solution of  to decision rule.

2) Algorithm Performance : The expected value of the left side of (28) is :

 (37)

Where  is a eigenvalue matrix,  is the covariance matrix of ,  is a egenvector function matrix.

Motivated by the transmitted signal in OFDM systems, elements of the input vectors are assumed to be realizations of zero-mean normal random variables with standard deviation .

With this assumption, example equation (16) comes like this

 (38)

*:* a weighted sum of the components of the distance vector .

Eq. (38) shows how the importance of different Volterra coefficients changes with the standard deviation of the elements of the input vectors.

For large values of , the elements of the Volterra representation are more important. This is intuitively correct since the increase of the input power beyond the linear range of the PAs should allow for better exploitation of the differences in the nonlinearities of the considered units.

## Generalized Likelihood Ratio Test(GLRT)

GLRT is another algorithm for unknown parameter. In the case of GLRT, the receiver does not estimate the parameters, but rather builds and compares the maxima of the likelihood functions over the unknown parameter vectors.

1) Decision Rule : For GLRT, the decision rule is :



With AWGN channel, we can get this relation :



Because  is the norm solution problems, we can get the unique solution and finally we can get the decision rule

 (47)

2) Algorithm Performance : from (47), we can get :

 (48)

And with the , we can simplify the (48) to (51) : 

Now, from the assumption of zero-mean normal random input, expectation of (51) becomes :

 (52)

With the equation (52) and (37), shows that probability of error for user identification based on the PA’s imperfections does not only depend on the Euclidean distance between the parameter vectors of the considered units, but also on the range of the input signal driving them.

# Simulations and Measurements

The influence of parameters is analyzed : the power of the input signal (), the SNR on the prob. Of error.

Normally, INL information of DAC is included in the data sheets, but Pas information is not included in data sheets.

## Exploitation of Digital-to-Analog Converter Nonlinearities for User Identification



Note that a relatively high SNR was required for user identification at these short input lengths

in this case.

## Exploitation of Power Amplifier Nonlinearities for User Identification

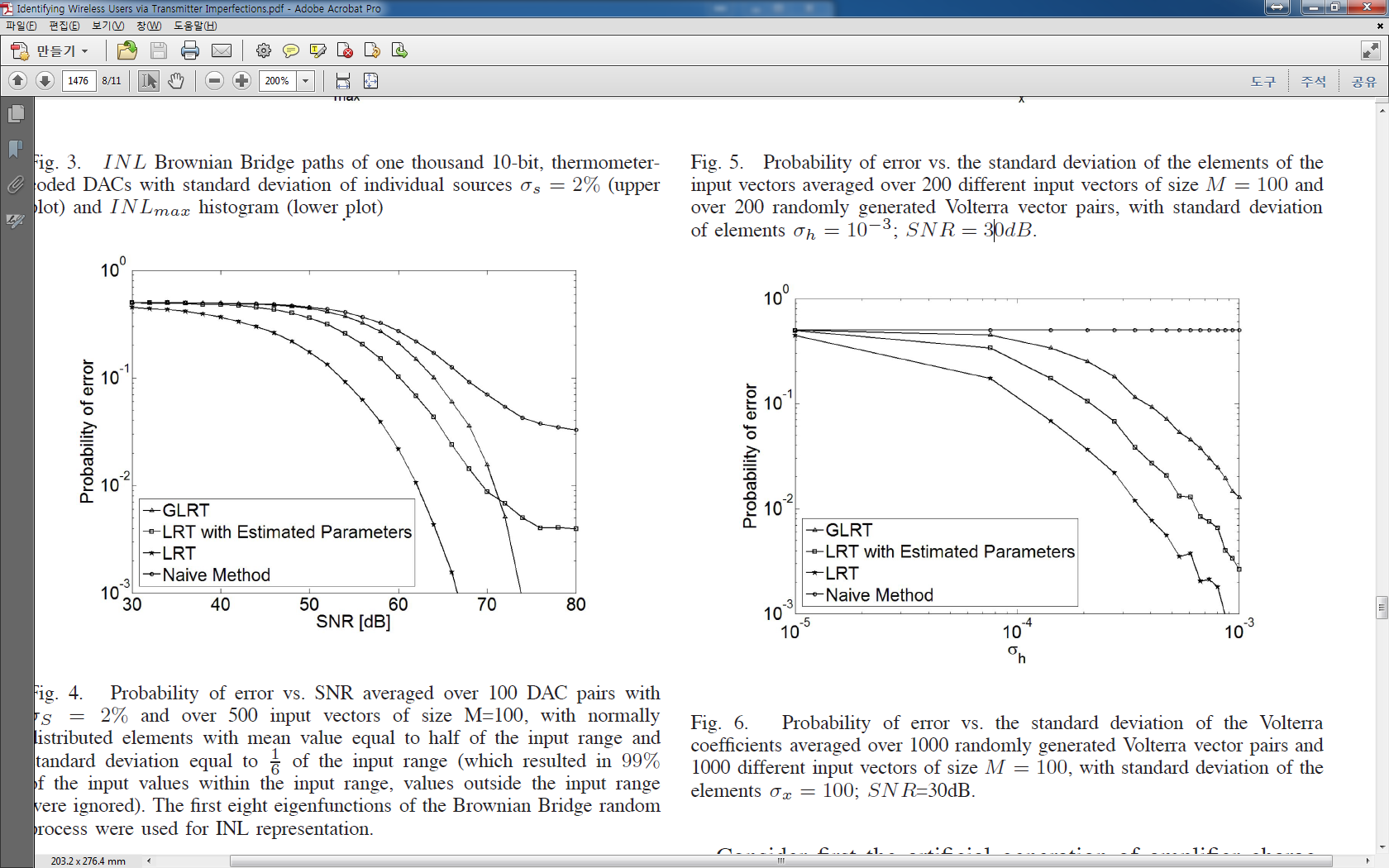
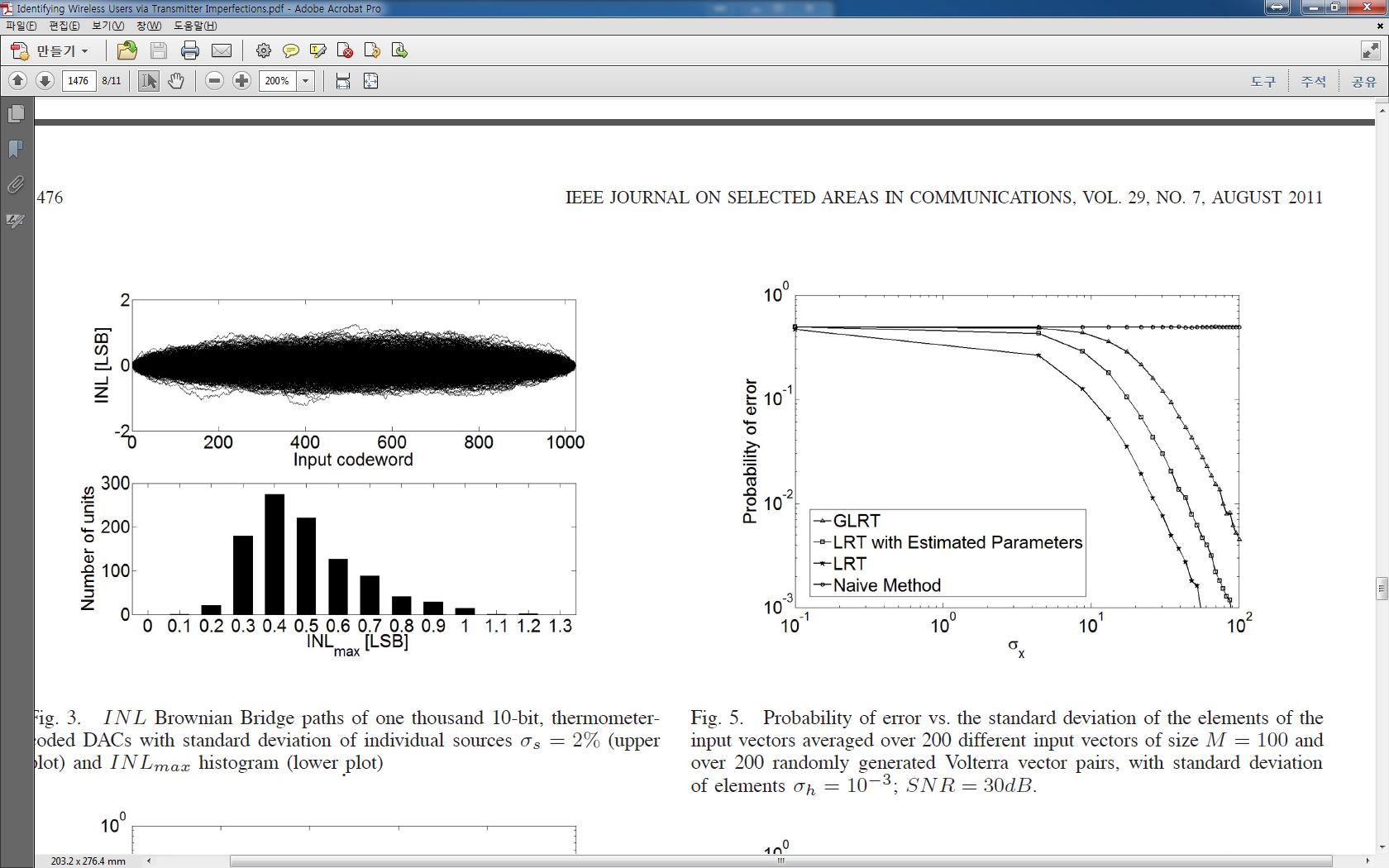


Figure 5 and 6 show the behavior of as a function of increasing input power and increasing difference of the Volterra representations of considered units.

As expected, Figures 5 and 6 demonstrate that the performance of the methods increases when the power of input signals increases and when the differences among amplifiers get larger.

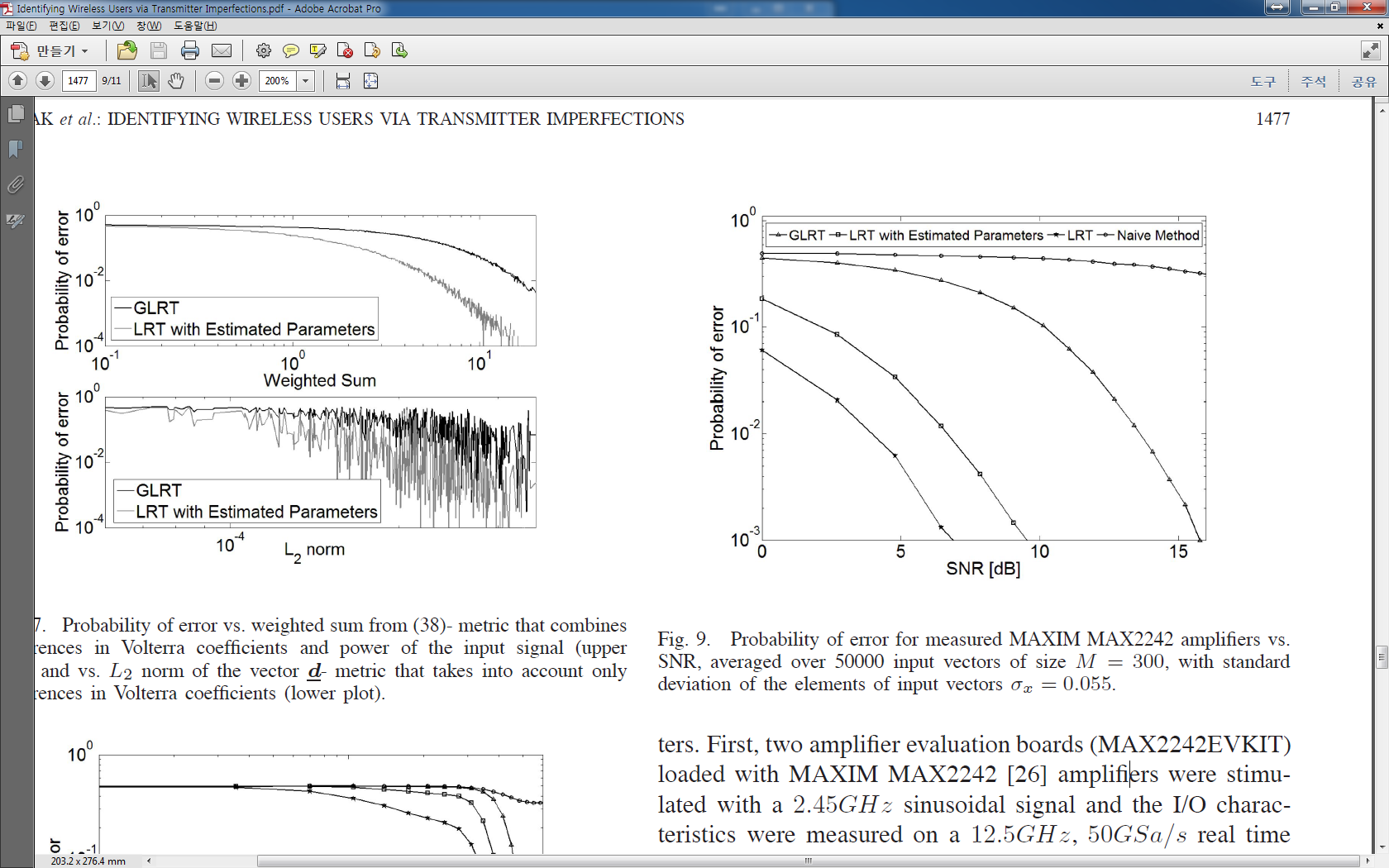


Figure 9 shows how the behaved for a fixed  as a function of *SNR.*

Note that the identifiability of the parts was very good, even at the low SNR of 15*dB* and for the very short input sequence of only 300 physical-layer symbols.

For SNR= 35*dB*, for both of the methods, we found no identification errors for any pair in 50000 trials.

## Evaluation of the Results

The model based approaches similar to ours have not yet been investigated. Thus, it is to hard conduct a comparison of our results with results of the previous work. : statistical simulation result v.s. empirical result

In [13], [13] reports an average success rate of 94-100% with 14 802.11 txs. In [18], [18] reports error rates at 0.34% with 138 802.11 txs. In this papers, with 35dB, even for very short input sequences, no error were observed during 50000 simulation trials with 8 802.11b txs.

One advantage of our model-based approach is that the results are easy to replicate for comparison to the performance of methods developed by others in the future.

# Conclusions and Future Work

In this paper, a new approach based on minute imperfections of different components of the transmitter hardware has been proposed for breaking user anonymity in wireless communication systems. The general models used to model the transmitter components allow for the determination of the probability of error of the decisions, which makes the proposed methods especially interesting for establishing probable cause and for use in court. Simulations have shown that the nonlinear variations of digital-to-analog converters can only be exploited when the signal-to-noise ratio is relatively high. However, in the case of power amplifiers, measurements from commercial chips indicate that amplifiers can be easily identified at typical power levels even at low SNRs and with very short observed sequences.

# Discussion