
Internet address:
http://dx.doi.org/10.4304/jcp.7.4.827-837

Please notice:
This is the journal's pdf version.
Journal of Computers is an open access journal.

© Reprinted with permission from
© Academy Publisher
Layout Dependent Phenomena
A New Side-channel Power Model
Geir Olav Dyrkolbotn
Norwegian Information Security Laboratory, NISlab,
Gjøvik University College, Norway,
Email: geirolav.dyrkolbotn@gmail.com
Knut Wold and Einar Snekkenes
Norwegian Information Security Laboratory, NISlab,
Gjøvik University College, Norway,
Email: {knut.wold,einar.snekkenes}@hig.no

Abstract—The energy dissipation associated with switching in CMOS logic gates can be used to classify the microprocessor’s activity. In VLSI design, layout dependent phenomena, such as capacitive crosstalk, become a major contributor to the power consumption and delays of on-chip busses as transistor technology get smaller. These effects may be known to the security community but have received little attention.

In a recent paper we presented a new power model, taking into consideration capacitive crosstalk. We have shown that capacitive crosstalk has a significant effect on gate energy dissipation. Our results confirm that the dissipated energy from CMOS switching gates depends not only on the hamming distance (HD), but also on the direction of switching activity on nearby data lines. We show that for an 8 bit data bus, crosstalk may improve detection performance from 2.5 bits (HD based detector) to a theoretical 5.7 bits and simulated 5.0 bits (crosstalk based detector) of information per sample.

In this paper we elaborate on the theory and simulations of layout dependent phenomena and how they must be considered when analyzing security implications of power and electromagnetic side-channels. We have also added a small case study, i.e. the electromagnetic side-channel of a smart card, that supports our simulations/theoretical results.

Index Terms—Crosstalk, Power model, Switching CMOS, Side-channels, Classification, Entropy

I. INTRODUCTION

When a microprocessor executes its program, power consumption (or resulting electromagnetic emanation) can be used to reveal the contents of program and/or data memory of the microprocessor. The correlation between power consumption and microprocessor activity has found many uses: to recover cryptographic keys [2], [3], [10]–[12], to reveal hidden hardware faults (trosjans) on integrated circuits [1], to control the emanation through subversive software in the Wireless Covert Channel Attack [7] and to reverse engineer the code executed by microprocessors [14].

In side-channel attacks, a common power model used to simulate the power consumption is the Hamming Distance (HD) model, as it is simple and generic [12]. The model assumes the power consumption to be proportional to the number of transitions taking place. If this assumption is appropriate, signals transmitted on a parallel bus (e.g. intermediate values of the cryptographic algorithm) with the same HD should have equal power consumption and therefore be indistinguishable. This is not always the case, e.g. if Bayes classifier is used, as suggested by the template attack [3]. It has also been demonstrated in [8] that signals with the same number of transitions can be classified using a modified template attack.

The phenomena behind this may be known in the security community, but has received little attention. One paper by Z. Chen, S. Haider and P. Schaumont [4], investigates the effect of the coupling capacitance on masking schemes without a detailed examination of the phenomena. In their book "Power Analysis Attacks", S. Mangard, E. Oswald and T. Popp [12] mention power simulation at analog level as "the most precise way to simulate the power consumption of digital circuits...". Parasitic elements, such as capacitances between the wires and unwanted capacitances in the transistors are mentioned, however, it is also stated that it is very common to make simplifications by lumping together extrinsic and intrinsic capacitances into a single capacitance to ground. This will, in fact, make the model incapable of explaining the results we are addressing in this paper.

Parasitic couplings, and the coupling capacitance in particular, are however, a great concern within sub-micron VLSI design [5], [13], [15]. CMOS technology is currently being pushed into deep sub-micron range. As the number of transistors increase, the need for on-chip wiring increases as well and must be scaled accordingly. Parasitic couplings between interconnects, such as on-chip busses, must be taken seriously as they influence both the power consumption and maximum obtainable speed [5]. F. Moll, M. Roca and E. Isern [13] did a detailed analysis of the energy dissipation from two metal lines running close together. The lines were driven by CMOS inverters and transitions in one or two wires were studied. The effect of coupling capacitance between the two lines on the power consumption was shown analytically and simulated.
in HSPICE. The main result was that if two bus lines have transitions in the same or opposite direction at the same time, the total energy is either lower or higher than if the two transitions are treated independently. This is due to the coupling capacitance. C. Duan, V.H.C. Calle and S.P. Khatri [5] focus on crosstalk avoidance codes that aim to reduce the effect of the coupling capacitances by avoiding specific data transition patterns. Their model considers coupling capacitance, $C_C$, between three adjacent lines. They show that 3 bit transition patterns can be divided into 5 crosstalk classes based on the influence of the coupling capacitances, $C_C$. The energy consumption therefore depends on which crosstalk class the transition pattern belong to, as seen in Table I reprinted from [5].

### Table I.
**Classes of Crosstalk from [5]. $C_{eff}$ is the Efficient Capacitance, $C_L$ the Load Capacitance and $\lambda = C_C/C_L$**

<table>
<thead>
<tr>
<th>Class</th>
<th>$C_{eff}$</th>
<th>Transition Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>0C</td>
<td>$C_L$</td>
<td>000 $\rightarrow$ 111</td>
</tr>
<tr>
<td>1C</td>
<td>$C_L(1+\lambda)$</td>
<td>011 $\rightarrow$ 000</td>
</tr>
<tr>
<td>2C</td>
<td>$C_L(1+2\lambda)$</td>
<td>010 $\rightarrow$ 000</td>
</tr>
<tr>
<td>3C</td>
<td>$C_L(1+3\lambda)$</td>
<td>010 $\rightarrow$ 100</td>
</tr>
<tr>
<td>4C</td>
<td>$C_L(1+4\lambda)$</td>
<td>010 $\rightarrow$ 101</td>
</tr>
</tbody>
</table>

The focus within VLSI design, such as [5] and [13] is on power consumption and delays caused by the coupling capacitance. They do not consider security implications, such as the ability to use the variation in energy consumption to classify transition patterns. However, correlations between data and energy consumption are exactly what side-channel attacks, such as DPA and Template attack, rely upon.

In this paper we elaborate on the hypothesis put forward in [9] that layout dependent phenomena, such as capacitive coupling between wires, can explain why it sometimes is possible to distinguish transition patterns with the same HD. We extend the theory and simulations of how the new power model, which takes into account capacitive crosstalk, affect our ability to classify activity in a microprocessor.

We look at the total dissipated energy from a parallel data bus driven by CMOS inverters. Our model is a generalization of [13], with inverters consisting of two MOSFET transistors, a load capacitance $C_L$, connected to each inverter output and a coupling capacitance $C_C$, connected between each bus line. Our model is generalized to $n$ lines and simulations in PSPICE are done with eight bus lines.

The purpose of our simulation is to show that when the dissipated energy depends on the direction of change of nearby data lines, and not only the number of transitions taking place, the number of possible energy levels dissipating from the bus will increase, thus allowing classification of a larger number of transition patterns. Our hypothesis is that this can be used to explain why some signal with the same HD can be distinguished. Our model can easily take into consideration other layout dependent phenomena, potentially offering an explanation to classification of an even larger set of transition patterns. We will use entropy as our classifier performance indicator and show that a detector capable of detecting energy levels due to crosstalk can extract more information than a detector based on HD only.

Finally, in order to probe the practicality of our theory and simulations, we have included a small case study, in which the objective is to see if analysis of electromagnetic side-channel information also supports the division into crosstalk energy levels.

This paper is organized as follows: Section II presents the hypothesis of layout dependent phenomena. Section III presents our model and necessary theory to calculate the energy dissipation. Section IV is an analytic analysis of security implications. Section V presents simulation results. Section VI presents a case study and future work. Finally, a conclusion is drawn in Section VII.

### II. Layout Dependent Phenomena

In a physical implementation of any circuit (e.g. CMOS based microprocessor) a number of phenomena will influence the energy dissipation and the resulting radiated electromagnetic field. These phenomena include inductance and capacitance of conductors, inductance and capacitance between conductors, wireless transmission characteristics (i.e. antenna properties) of conductors and other circuit elements and complex combinations of these phenomena. These phenomena apply to any transistors and wires in a circuit, but we choose to look at a portion of wires running parallel, as we expect them to be relatively good antennas and therefore a good source for side-channel information. This is illustrated in the model of a parallel bus, driven by CMOS inverters, seen in Fig. 1.

#### A. Inductance and Capacitance of Conductors

Any conductor, $W_j$, carrying an electric current will have an associated distributed resistance $R_j$, inductance $L_j$, conductance $G_j$ and capacitance $C_j$, expressed as a characteristic impedance, $Z_0$. The characteristic impedance is often modeled as an infinite series of lumped components. The inductance $L_j$ and capacitance $C_j$ will both block high frequency signals and act as a low pass filter. Small variations in the length and width of conductors result in small variations in the inductance. Small variations in the area and distance to ground plane result in small variations in the capacitance. There will therefore be small variations in how signals on different conductors (e.g. bus lines) are filtered.

#### B. Inductance and Capacitance between Conductors

Crosstalk can be defined as the coupling of energy between two conductors. Inductive coupling is caused by mutual inductance, $L_{ij,j+1}$, (i.e. magnetic field) and capacitive coupling is caused by mutual capacitance, $C_{ij,j+1}$, (i.e. electric field) between wire $j$ and $j+1$. These couplings occur along the entire length of the conductor, but are also modeled as lumped components (Fig. 1). The
interaction of magnetic and electric fields will effectively change the characteristic impedance, $Z_0$, associated with the conductor. This interaction is layout dependent (e.g. distance and length of wires) and will effect both delays and energy dissipation. An important property of crosstalk is its dependency on the activity on the wires. F. Moll, M. Roca and E. Isern [13] state that, "coupling capacitance is very different from the capacitance to ground because it depends on the switching activity... ". If two lines are low and rise at the same time, the mutual capacitance coupling, $C_{j,j+1}$, does not have to be charged. However, if one line remains low and the other rises, $C_{j,j+1}$ must be charged, resulting in increased rise time and power consumption.

D. Complex Combinations of Factors

Finally, complex combinations of layout dependent phenomena may be the key to identify minute differences in microprocessor activity, e.g. the radiation efficiency of bus lines combined with data and layout dependencies of the line characteristics due to crosstalk suggest that the emanation detected will have data and layout dependent variations in power consumption and delay. In the following, we will assume that the coupling capacitance is the dominating factor, and show how this can explain why some signals with the same HD can be distinguished. This will show the potential effect of layout dependent phenomena on classifying microprocessor activity. Our work can easily be extended by including other layout dependent phenomena if a more precise model is needed.

III. THEORETICAL CONSIDERATIONS

By limiting the model to only coupling and load capacitances, the model in Fig. 1 can be simplified as seen in Fig. 2. This is a generalization of the model for two lines used in [13] and includes a model of the CMOS inverter.

In order to run simulations in PSpICE, we need an expression for the total energy dissipation, $E_T$. The energy dissipation for wire $j$ in the $p$ and $n$ type transistor can be expressed as:

$$E_{pj} = \int (V_{DD} - V_j)i_{pj}dt$$

$$E_{nj} = \int V_ji_{nj}dt$$

The overall energy dissipation for an $n$ wire bus is then
In order to compare the simulated energy dissipation ($E_T$) with analytic values ($\hat{E}_T$), different expressions than (4) and (5) are needed.

It is only when the individual line has a transition, that it is subject to capacitive crosstalk. Quantifying this crosstalk influence has to take into consideration voltage changes on the line itself and one (edges) or two adjacent lines. Let $\delta_j \in \{0, \pm 1\}$ be the normalized voltage change on line $j$, then the voltage change between two lines $j$ and $k$ is $\delta_{j,k} = \delta_j - \delta_k$. The crosstalk influence $\alpha_j$ on line $j$ can then be defined as:

$$\alpha_j = \begin{cases} 0 & \text{no transition line } j \\ \frac{\delta_{j,j-1} + \delta_{j,j+1}}{\left| \delta_{j,j-1} \right| + \left| \delta_{j,j+1} \right|} & \text{otherwise} \end{cases}$$

It can be shown that $\alpha_j = \{0, 1, 2\}$ for lines with only one adjacent line (edges), and $\alpha_j = \{0, 1, 2, 3, 4\}$ for lines with two adjacent lines. Let the total crosstalk influence for an $n$ line bus be called a crosstalk index $\alpha$, defined as the sum of the crosstalk influence of each line:

$$\alpha = \sum_{j=1}^{n} \alpha_j$$

If the contributions from the load ($C_L$) and coupling capacitance ($C_C$) are dominant to the dissipated energy, then $E_T$ can be expressed by the following power model [9]:

$$E_T = \frac{1}{2} C_L V_{DD}^2 (k + \alpha \lambda) = E_0 (k + \alpha \lambda)$$

where $E_0 = \frac{1}{2} C_L V_{DD}^2$, $V_{DD}$ is the power supply voltage, $k$ is the number of transitions on the data bus, $\lambda = C_C/C_L$ and $\alpha$ is the crosstalk index of (7) indicating the coupling capacitance induced crosstalk, similar to the crosstalk classes in [5].

In the next section we will use (8) to analyze which transition patterns can be distinguished.
IV. SECURITY IMPLICATIONS

The relationship between energy dissipation, number of transitions, crosstalk index, load capacitance and coupling capacitance in (8) can be used to analyze delays and energy dissipation of sub-micron VLSI design [5], [13], [15]. However, we are interested in the security implications of layout dependent phenomena, and in this paper the coupling capacitance in particular. How will a power model (8) that includes coupling capacitance affect our ability to predict the energy dissipation of activity in a microprocessor, such as data transfer on a parallel bus?

Let \( T \) be the set of possible transitions on an \( n \) bit parallel bus. Since "no transition" can be both \( 0 \to 0 \) and \( 1 \to 1 \) there are \( |T| = 4^n \) possible transition patterns for an \( n \)-bit bus. Assuming that each transition pattern’s energy dissipation is unique, a model should ideally predict a total of \( |T| = 4^n \) energy levels. This may not be possible if physical properties are such that multiple transition patterns indeed use the same amount of energy.

Classification by energy dissipation can only distinguish transition patterns by the distinct energy levels explained by the model. A model that assumes energy dissipation proportional to the number of transition, can therefore only distinguish transition pattern into subsets \( T_k \), \( k = \{0, \cdots, n\} \) being subsets of \( T \) that has \( k \) transitions. The number of transition patterns in each subset is given by: \( |T_k| = 2^n \binom{n}{k} \). The total number of possible transitions on an 8 wire bus \((|T| = 65536)\) can be divided into 9 subsets, \( T_0, T_1, \cdots, T_8 \) based on the number of transitions, \( k \). The energy dissipation, \( E_T \) (using (8) with \( \alpha = 0 \)), associated with each subset \( |T_k| \) can be seen in Table II. A model that assumes energy dissipation proportional to the number of transition, can only classify transition pattern by the energy level of these 9 subsets. For example, in Table II there are 14336 transition patterns with energy level 3\( E_0 \) that are indistinguishable by the number of transitions alone.

Using the new power model (8), taking into consideration the coupling capacitor, each subset \( T_k \) can be split into a number of new energy levels. This gives a number of smaller subsets \( T^\alpha_k \), \( |T^\alpha_k| > |T^\alpha_k| \) and \( \sum_{q \in q_k} |T^\alpha_k| = |T_k| \), where \( \alpha \) is the crosstalk index of (7) and \( q_k \) is the set of possible values of \( \alpha \) for \( k \) transitions.

Computing \( |T^\alpha_k| \) for a fixed number of bus lines \( n \) can be done by constructing a table of \((2^k)^2\) elements corresponding to all possible transition patterns. For each of these, first compute the crosstalk index \( \alpha (7) \), then the energy dissipation \( E_T (8) \), \( |T^\alpha_k| \) can then be computed by counting the table entries for each tuple \([k, \alpha]\). Notice that for a finite \( n \), there are restrictions on the sets \( q_k \) of possible values of \( \alpha \). As the number of transitions increase, all energy levels are not possible. This applies to 6,7 and 8 transitions for an 8 bit bus.

The results for an 8 bit bus can be seen in Table II. The results show that taking into consideration the coupling capacitance increases the number of energy levels from 9 in the HD model to 93 in the crosstalk model, e.g. the 14336 transition patterns with 3 transitions previously indistinguishable can now be split into 10 energy levels. The largest increase in energy levels is found for 6 transitions with 21 new energy levels. Note that energy level \( \alpha = 20 \) does not exist.

Also notice that given an ideal classifier, there is no confusion between subsets of the same \( k \) as they all have unique energy levels. There may, however, be confusion between subsets of different \( k \). The extent of this confusion is architecture dependent, expressed by \( \lambda \), e.g. subset \( T^\beta_k \) has the same energy level as \( T^\beta_j \) if \( \lambda = 1/4 \), in case they should be treated as one subset. It is easy to show that confusion between transition \( A \) (energy \( E_{TA} \), \( k_A \) transitions and crosstalk index \( \alpha_A \)) and \( B \) (energy \( E_{TB} \), \( k_B \) transitions and crosstalk index \( \alpha_B \)) happens when:

\[
\lambda_{AB} = \frac{k_B - k_A}{\alpha_A - \alpha_B} \tag{9}
\]

\( \lambda_{AB} \) values that are close to the real \( \lambda = C_C/C_L \) indicate subsets that will be difficult to distinguish.

Finally, we have only shown how to split the subset \( T_k \) into smaller subsets \( T^\alpha_k \) by considering the effect of the coupling capacitance (i.e. \( \alpha \)). This idea can easily be generalized, such that \( T_k \) is split into subsets \( T_{k+1}^{\beta_k} \), where \( |T_k| > |T_{k+1}^{\beta_k}| \), and \( \beta \) is the influence of other layout dependent phenomena. Examples of phenomena for future work include: variations in coupling and load capacitance, coupling capacitance between line \( j \) and \( j+2 \), inductance, effect of bends in circuit paths and multi-layer capacitance (3-dimensional). We believe that the key to identify minute differences in microprocessor activity is to combine several layout dependent phenomena, \( \beta_1, \cdots, \beta_m \), such that:

\[
|T_k| > |T_{k+1}^{\beta_k}| > |T_{k+2}^{\beta_k}| > \cdots > |T_{k+\sum_{i=1}^m \beta_i}^{\beta_k}| \tag{10}
\]

A. Classification Performance

Table II shows that, taking into consideration the coupling capacitance, we are able to increase the number of subsets (or energy levels) \( T_k \) to \( T^\alpha_k \). For the purpose of comparing alternative detectors we will assume uniform random transition. Thus for an 8 bit bus we would like the detector to extract 16 bits of information, i.e. high or low (2 bits of information) for each of the 8 wires. We will use entropy as our classifier performance indicator. The entropy (i.e. bits of information) for a detector, when there are \( r \) energy levels, can be calculated using:

\[
H(x) = -\sum_{i=1}^r p(x_i) \log p(x_i) \tag{11}
\]

In the following, we have assumed an 8 bit bus width, thus there are \( 4^8 = 65536 \) possible transitions. Call the detector that can extract 16 bits of information a level detector. If we assume that one only has bus activity when initial and final state are different, and that \( 0 \to 1 \) and \( 1 \to 0 \) can be distinguished, an observation will give us the following entropy: \(-1/(2\log1/2 + 1/4\log1/4 + 1/4\log1/4) = 3/2 \) bits as we cannot distinguish \( 0 \to 0 \)
from 1 → 1, but 0 → 0, 0 → 1, 1 → 0 can be distinguished. Thus, each observation will give us 3/2 bits per line. The theoretical optimum for an 8 bit bus with a ‘transition detector’ would be 8 · 3/2 bits = 12 bits, assuming all observable transitions are distinguishable. In other words, by observing transitions rather than levels, we loose 4 bits (1/2 bit per line) compared to the setting where we would observe the states.

Using the results of Table II, we can now calculate the entropy extracted by a detector that can distinguish HD only (T_b) and a detector that can distinguish energy levels due to crosstalk (T_{crosstalk}).

The entropy extracted by a HD detector is found using (11) with 9 energy levels (r = 9) and p(x_i) = |T_b−1|/65536 (|T_{crosstalk}−1|) from column 6 and 12 Table II) giving an entropy of 5.7 bits.

The difference between the ideal value of a level detector and the entropy extracted by other detectors, represent the amount of guessing needed for classifying an observation. By considering the coupling capacitance and not only HD, we extract more information out of each observation, therefore reducing the amount of “guessing” needed for classification. In the next section we present simulations validating the effect of the coupling capacitance.

V. Simulations

The simulations are performed in PSPICE with $C_L = 400 fF$, $C_C = 250 fF$, $Vdd = 3V$ and a rise- and fall-time of 200ps of the input voltages (same as [13]). The inverter drivers are equal and balanced. Equation (4) is used in PSPICE to find the simulated energy dissipation $E_T$. 

© 2012 ACADEMY PUBLISHER
A. Model Validation

Simulations were initially carried out and compared with the results of [5], [13] as a model validation. The results are shown in Table III and IV. Transition pattern refers to transitions in the output voltage $V_j$ (Fig. 2) and also shows the number of bus lines used. Column 2 is the number of transitions $k$ followed by the crosstalk index $\alpha$. Theoretical energy, $E_T$, is calculated from (8) and simulated energy, $\hat{E}_T$ is from PSPICE simulations.

The simulations for two lines are consistent with [13]. For two wires, as seen in Table III, it is clear that the energy dissipation for two simultaneous transitions is either lower or higher than if treated as two single transitions, depending on the direction of the transitions, as expected. This means that introducing the coupling capacitance it is possible to explain a difference in the energy dissipation for transition patterns $00 \rightarrow 11$ from $01 \leftrightarrow 10$. Without this difference in energy dissipation the two transition patterns should not be distinguishable.

Simulations of three lines confirm the difference in energy dissipation of the 5 crosstalk classes (Table I) introduced in [5]. Notice that only the transition pattern with the same number of transitions (first and last, second and fourth) can be used to evaluate the effect of the coupling capacitance.

The small differences between analytic and simulated energy dissipation can be explained by simplifications in deriving (8) (e.g. omitting leakage currents, such as short-circuit and sub-threshold currents). Having validated our model, all the following simulations are done on an 8 bit bus.

B. Results and Discussion

Simulation results for 8 lines are shown in Table V. The table is not exhaustive, but includes results for all possible subsets $T_k$.

The simulated energy levels $\hat{E}_T$ are similar to the analytic values $E_T$. The results confirm that energy consumption is proportional to the number of transitions and the crosstalk index, $\alpha$. The crosstalk index depends on switching activity on adjacent lines and position, edge (one adjacent wire) or middle (two adjacent wires). As seen in Table V, the results also confirm that there is no confusion between energy levels for subsets of an equal number of transitions. However, there may be some confusion between some of the 93 subgroups, e.g. the energy dissipation of subset $T_2^2$ and $T_3^4$ are almost equal. This is expected as $\lambda_{AB} = 0.5$ is close to $\lambda = 0.63$ used in this experiment. Other examples can be found and this reduces the number of subsets depending on how accurate our detector is. A theoretical crosstalk detector capable of separating all 93 energy levels can extract 5.7 bits of information. It is therefore expected that a practical crosstalk detector will extract less information, due to some subset having almost equal energy levels. Which of the simulated energy levels that should be considered indistinguishable will depend on the accuracy of the detector and the number of observations available. A random loss of 20% of the subsets will still, on average, have an entropy of 5.0. Even with this loss due to similar energy levels, the information gain is still 2.5 bits compared to the HD detector. The performance of the detectors is summarized in Table VI.

VI. Case Study and Future Work

In order to probe the practicality of our theory and simulations, we collected a small set of experimental data. The objective was to see if analysis of electromagnetic side-channel information also supports the division into crosstalk energy levels of Table II and V.

When classifying two transition patterns by their energy dissipation, we expect a lower probability of error ($P_e$) when the difference in energy level is large and higher $P_e$ as the difference in energy levels decreases. When the energy dissipation of two transition patterns are equal, we don’t expect to be able to do any better than flipping a coin. Transition patterns with an unequal number of transitions have a relatively large difference in energy dissipation and are therefore fairly easy to distinguish, as shown in [8].

Consider two transition patterns $A$ (crosstalk index $\alpha^A$) and $B$ (crosstalk index $\alpha^B$) of an equal number of transitions. Let $\alpha$-distance, $\Delta \alpha = |\alpha^A - \alpha^B|$, be the difference in crosstalk index between transition patterns $A$ and $B$. According to our model (8), patterns with $\Delta \alpha = 0$ dissipate the same amount of energy and
are therefore assumed to be indistinguishable with an expected classification error, \( P_e = 0.5 \) (guessing), e.g. 00000000 \( \Rightarrow \) 00111111 and 00000000 \( \Rightarrow \) 11111100 both belonging to \( T_5 \) (Table II). Patterns with \( \Delta \alpha > 0 \) are assumed to be distinguishable with \( P_e < 0.5 \) and \( P_e \) is expected to decrease as \( \Delta \alpha \) increases, e.g. it is expected to be easier (lower \( P_e \)) to classify 00000000 \( \Rightarrow \) 00111111 from 00010100 \( \Rightarrow \) 00101010 (\( \Delta \alpha = 17 \), than 00000000 \( \Rightarrow \) 00111111 from 00000000 \( \Rightarrow \) 01011011 (\( \Delta \alpha = 4 \)) (Table V), simply because \( \Delta \alpha = 17 \) indicate a larger difference in energy levels than \( \Delta \alpha = 4 \).

An experiment was designed to validate the expected relationship between difference in energy levels, \( \Delta \alpha \), and classification error, \( P_e \). The experiment consisted of three steps: (i) Measure the electromagnetic emanation from a set of transition patterns. (ii) Evaluate the performance (\( P_e \)) of a classifier trained by the recorded data. (iii) Look at average \( P_e \) as a function of \( \Delta \alpha \). Do we see the expected relationship?

An 8 bit internal data bus on a smart card (i.e. PIC 16F84A microprocessor) was chosen as the source for the electromagnetic radiation. All 18 possible crosstalk indexes for transition patterns with 5 transitions were studied (\( T_{5+k}^\alpha \), \( \alpha = 1, \ldots, 18 \) in Table II). A total of 1000 traces (observations) of the electromagnetic radiation, for each of the 18 transition patterns, were collected. A 10 G/s oscilloscope with a broadband E near-field probe was used. The probe was positioned as close to the microprocessor as possible, without any decapsulation, see Fig. 3.

The challenge of this experiment is to manipulate the microprocessor, such that the appropriate transition patterns are generated on the internal data bus. It is also essential that the power consumption (i.e. electromagnetic radiation) is correlated with this bus activity and not dominated by noise (e.g. other irrelevant microprocessor activities).

Our approach was to combine detailed knowledge of
the execution sequence of the microprocessor with careful assembly programming. The code was written off-line, using vendors development kits, and loaded to the smart card with a standard smart card terminal. To execute the code, a customized smart card terminal was used to provide power and clock signals only. This was to limit noise from external circuitry. The microprocessor automatically executed from the beginning of the program and no I/O communication was required, with the exception of an initial trigger signal.

The objective of the code is to create a transition between value \(a\) and value \(b\) on the microprocessors internal 8 bit data bus and minimizing irrelevant activity. PIC 16F84A has a 2-stage pipeline architecture. Each instruction is executed during four clock cycles (Q1-Q4). The transition between value \(a\) and \(b\) should ideally take place in consecutive clock cycles or "near" consecutive such that no data bus activity takes place between handling the two values. In addition, parallel activity, due to pipelining, must be avoided or kept constant for all transitions. This is possible to achieve through careful choice of instructions as shown in [8]. The code used in this experiment is seen in Table VII.

Code lines 1 – 4 toggles the smart cards I/O and provides a trigger-point for the oscilloscope. The following 10 NOP’s create a buffer from electromagnetic disturbances caused by the relatively strong I/O toggle and the rest of the program. Code lines 15 – 18 are used to

<table>
<thead>
<tr>
<th>Test Code</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main program</strong></td>
</tr>
<tr>
<td><strong>Start</strong></td>
</tr>
<tr>
<td>; Trigger Turn I/O ON and OFF</td>
</tr>
<tr>
<td>1 movlw 80h ; Turn I/O ON</td>
</tr>
<tr>
<td>2 movwf PORTB ; by moving the value 80h onto port B</td>
</tr>
<tr>
<td>3 movlw 00h ; Turn I/O OFF</td>
</tr>
<tr>
<td>4 movwf PORTB ; by moving the value 00h onto port B</td>
</tr>
<tr>
<td>; 10 NOP’s to create buffer from I/O disturbances</td>
</tr>
<tr>
<td>5 nop</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td>14 nop</td>
</tr>
<tr>
<td>; Transition: a:0000 0000 - b: 0001 1111</td>
</tr>
<tr>
<td>15 movlw 00h ; a into W register</td>
</tr>
<tr>
<td>16 movwf DATA1 ; mov a from W to DATA1 register</td>
</tr>
<tr>
<td>17 movlw 1Fh ; (b-a) into W register</td>
</tr>
<tr>
<td>18 addwf DATA1,1 ; Q2 read a, Q4 write b=(a+(b-a))</td>
</tr>
<tr>
<td>; Transition: a:0000 0000 - b: 1000 1111</td>
</tr>
<tr>
<td>19 movlw 00h ; a into W register</td>
</tr>
<tr>
<td>20 movwf DATA1 ; mov a from W to DATA1 register</td>
</tr>
<tr>
<td>21 movlw 8Fh ; (b-a) into W register</td>
</tr>
<tr>
<td>22 addwf DATA1,1 ; Q2 read a, Q4 write b=(a+(b-a))</td>
</tr>
<tr>
<td>; Continue for all 18 transition patterns</td>
</tr>
<tr>
<td>23 goto Start</td>
</tr>
</tbody>
</table>

**Table VII.** Code used to generate 18 different transition patterns on the internal data bus of microprocessor PIC 16F84A

© 2012 ACADEMY PUBLISHER
create a transition from bit pattern 00000000 to 00111111 ($T_2$). Code lines 15 – 17 are initialization, making sure value $a$ is available in DATA1 register and value $(b-a)$ is found in the working register. Transition between value $a$ and $b$ is then made possible by the ADDWF instruction of line 18. In clock cycle 2 (Q2) the value $a$ is read over the data bus, in clock cycle 3 (Q3), $a$ is added to $(b-a)$ found in the working register. The result, $b$, is written back over the databus in clock cycle 4 (Q4), creating the desired transition without unwanted data bus activity. The process can now be repeated for all other values of $a$ and $b$, as shown in code lines 19-22. Finally code line 23 repeats the program indefinitely.

Classification between all pairs of transition patterns was done according to the Modified Template Attack [8]. This includes feature selection, training and evaluating the performance of a quadratic Bayes classifier (for details refer to [8]). The probability of error, $P_e$, was found from the confusion matrix [6]. Since the classification accuracy depends on how the observations are split, the average of 100 random permutations of 200 training observations and 800 test observations was used. Finally, the average classification error as a function of $\alpha$ distance ($\Delta\alpha$) was calculated and plotted in Fig. 4.

The results (Fig. 4) show that transition patterns belonging to equal energy levels ($\Delta\alpha = 0$) have $P_e = 0.5$. This is equal to guessing as expected. When the difference in energy level increase (larger $\Delta\alpha$) the results suggest that the average classification error decrease. This supports our simulation/theoretical results. We hypothesize that the discrepancy between our simulation/theoretical results for alpha distance 4 is a consequence of statistical uncertainty/noise in the experimental data. Currently, we cannot offer any explanation for why classification error seems to increase for high values of the alpha distance.

The classification results are a result of measuring electromagnetic emanation from a real system. The assumptions made for analytic ($E_T$) and simulated ($E_T$) energy dissipation are therefore not necessarily valid. A real system will be subject to all layout dependent phenomena of section II, and not limited to coupling capacitance and the assumptions that load and coupling capacitances are equal. To evaluate these assumptions and how the power model can be revised if the assumption do not hold is subject of future work. Future work also include calculating confidence intervals for the results in Fig. 4. This is not trivial, due to non-uniform distribution of energy levels (see Table II). When noise causes energy levels to overlap, the results looks to be a function of which energy levels merge, and must therefore be modeled carefully.

Finally, to validate the impact of the new power model on security, we encourage the power model to be applied in a DPA attack and compared to the performance of other power models (e.g. HD/HW). Such comparison could also be done with extended versions of our power model, e.g. adding other layout dependent phenomena or removing some of the simplifying assumptions made (equal coupling and load capacitance).

VII. CONCLUSION

It is known that one can distinguish bus activity generated from signal transitions having different HD. In this paper we elaborate on the theory and simulations on the hypothesis from [9] that layout dependent phenomena, such as inductance and capacitance in and between conductors and radiation properties of circuit elements, can explain why it sometimes is possible to distinguish transition patterns with the same HD. Our simulations show that capacitive crosstalk has a significant effect on gate energy dissipation, and confirm that the dissipated energy from CMOS switching gates depend not only on the HD, but also on the direction of switching activity on nearby data lines. For an 8 bit bus, this increases the number of possible energy levels from 9 (HD) to 93 (crosstalk), and therefore allows us to explain why signals with the same HD sometimes can be distinguished. Where as an HD based detector can provide about 2.5 bits of information per sample, a crosstalk based detector will yield about 5.7 bits (theoretical) or 5.0 bits (simulated) of information per sample - in all cases for an 8 bit bus.

In this paper we have also shown that experimental data, i.e. the electromagnetic side-channel of a smart card, suggest that the average classification error gets reduced as $\alpha$ distance (i.e. difference in energy level due to capacitive crosstalk) increases. This supports our simulations/theoretical results that layout specific phenomena (e.g. capacitance) must be considered when analyzing security implications of electromagnetic side-channels.

REFERENCES


© 2012 ACADEMY PUBLISHER
Figure 4. Average classification error as a function of difference in energy level (expressed as $\alpha$ distance, $\Delta\alpha$)


