Power Consumption and Energy Estimation in Smartphones

Ahmed Alsheikhy¹, Reda Ammar², Raafat Elfouly³, Mosleh Alharthi⁴

¹Department of Electrical Engineering, Northern Border University, Arar, Saudi Arabia
²Department of Computer Science and Engineering, University of Connecticut, Storrs, USA
³Department of Computer Engineering, Cairo University, Cairo, Egypt
⁴Department of Electrical Engineering, Taif University, Taif, Saudi Arabia

Abstract—A developer needs to evaluate software performance metrics such as power consumption at an early stage of design phase to make a device or a software efficient especially in real-time embedded systems. Constructing performance models and evaluation techniques of a given system requires a significant effort. This paper presents a framework to bridge between a Functional Modeling Approach such as FSM, UML etc. and an Analytical (Mathematical) Modeling Approach such as Hierarchical Performance Model (HPM) as a technique to find the expected average power consumption for different layers of abstractions. A Hierarchical Generic FSM “HGFSM” is developed to be used in order to estimate the expected average power. A case study is presented to illustrate the concepts of how the framework is used to estimate the average power and energy produced.

Keywords—Finite State Machine (FSM), Hierarchical Performance Model (HPM), Power consumption, Real-time embedded systems, Smartphones.

I. INTRODUCTION

In today’s world, using embedded systems is rising up extremely rapidly [1]. Many of those systems run on batteries and power consumption is considered to be an important criteria throughout the design process of an implementation [1,2,3,4]. Very often, a designer has to rely on a simulation tool to take a decision about which design is the best among others. Sometimes that tool can be particularly time consuming or insufficient. Due to heavy demands on embedded systems, it is essential to estimate performance metrics such as the delay and power consumption. The focus of performance analysis methods for real-time embedded systems is more on the analysis of timing aspects and power consumption [1,2]. In particular, a designer intends to determine which design produces less power while a system meets its real-time requirements [2,3]. A performance modeling scheme is required to evaluate the power consumption caused by communication and computation by distributed system architectures and existing software on different platforms [1,3,4,5]; and also to identify where bottlenecks occur. Engineers rely on performance modeling to predict the expected power before moving to the final stage of implementation. However, in the absence of a performance evaluation scheme, they must design and implement a system to predict the performance defects or bottleneck. Waiting to spot the performance defects or bottleneck to occur until the final stage of implementation and integration between different components results in increased project costs, reduced productivity and delays in schedule [1,2,3,6]; applying performance modeling and evaluation from the first stage of design in any system exhibits better results than those using a “fix-it-later” approach [1,6]. Currently, three approaches exist that evaluate system performance and analysis which are: 1- Simulation Based Method, 2- Analytical Based Method and 3- Direct Measurement [1]. The Hierarchical Performance Model (HPM) combines and collaborates direct measurements with the analytical approach with the help of software performance engineering [2,3,4,5], queueing networks [6], hardware and software co-design to predict the average power consumption.

Our contribution in this paper is done by developing the hierarchical generic FSM, converting it to a Markovian model and also integrating it with its affiliated hierarchical performance model (HPM) in order to estimate system performance metrics such as power consumption and time delay. Only power consumption is considered in this paper. The developed framework was applied only on one device due to the availability and presented within this paper.

In the reminder of this paper, we present related work on power consumption estimation schemes in Section 2, followed by a detailed discussion of the hierarchical generic finite state machine (HGFSM) and its affiliated hierarchical performance model (HPM) to estimate the power consumption. Section 4 includes a case study that applies the mapping scheme inside the framework and evaluate its power and energy and also to determine locations of bottlenecks. Section 5 concludes the paper.
II. RELATED WORK

Functional Modeling techniques and Analytical Modeling ones are used to estimate the system performance metrics such as power estimation at an early stage if possible. Queueing schemes have been used since the 1970s to model performance metrics of any software systems [6]. An FSM is used to evaluate system level performance [9,10,11]; however, that FSM was not applicable to any system since it was designed for a specific system. So designing a hierarchical generic FSM to be used in evaluating performance for any embedded system is developed in section 3. Many techniques were developed to estimate the power consumption at gate-level, circuit-level and register-transistor-level. However, those approaches are impractical to evaluate power consumption due to the lack of availability of circuit and gate levels information of a system under investigation. In [4], power analysis was done based on Y-chart scheme by Amit Nandi. He integrated power and performance analysis into the system level and claimed that the analysis became an integral part of the design process as it helped to find a proper architecture for a target application. Arafat, Ammar and Fergany in [8] applied the HPM method to evaluate the software power consumption based on measurements of the consumed power by each instruction. They claimed that their approach can be used to estimate power consumption of a software application based on physical measurements and computation modeling. In [13], rapid performance and power consumption evaluations were done at the system level only and that did not include the task level, module level and operations level. Kumar, Ben Attallah, Niar, Senn and Dekeyser in [14] presented a fast and accurate hybrid power estimation methodology for embedded systems at the system level only and did not provide any information about more levels. Many methods were developed to estimate the power consumption at the system level only as in [15].

Numerous parameters are required for each layer when using HPM to estimate the average power consumption in different levels; those parameters propagate from a bottom level to a higher one [1,2,3]. Nevertheless, accessing this information within a level or communication of information between these layers result in a complex manner [3]. The HPM is used to manage and distribute performance information between different layers of the framework. This paper considers the problem of mapping a functional modeling approach such as FSM to the analytical (mathematical) modeling approach such as (HPM) for performance analysis. A general overview of the developed framework using the functional modeling approach “HGFSM” which refers to Hierarchical Generic Finite State Machine and the analytical modeling approach (HPM) to estimate the power consumption is depicted in figure 1.

![Image](image1.png)

Fig.1: developed framework

Hierarchical Generic FSM is designed and then converted into the Markovian Model “MM”. Each state in the Markovian model is decomposed into another sub-markovian states if possible. Furthermore, Hierarchical Performance Model “HPM” is applied to each state to derive the expected average power consumption equation(s) using bottom-up methodology which represent(s) the objective function(s).

III. HIERARCHICAL GENERIC FINITE STATE MACHINE AND HIERARCHICAL PERFORMANCE MODEL

A typical FSM model is composed of 5-tuples \( \{ \Sigma, S_0, \delta, F \} \); where: \( \Sigma \) represents a set of input alphabets. \( S \) represents a set of states in the model. \( S_0 \) represents an initial state or a set of states which are sub-elements of \( S \). \( \delta \) represents a state-transition function which maps between a current state to a next state and \( F \) contains a final state or a set of states which belongs to \( S \) [9,10]. A task in any embedded system can be classified as either completed or failed. A set of states exists among those two states to form the hierarchical generic FSM “HGFSM” model as shown in fig. 2 [1,2].

![Image](image2.png)

Fig. 2: Hierarchical generic FSM
Each input alphabet is represented by a task or a set of tasks which are integrated to form a desired job. Movement from a current state $S_i$ to a next state $S_j$ is represented by a transition arrow and is done according to some existing circumstances or activities inside the system under investigation. For the hierarchical generic FSM, $S$ contains all 7 states whereas $S_0$ contains only one state which is the Initial state. $F$ contains one state which is named as completed state, referring to a successful completion, and is denoted by two circles in fig. 2. 

The states are: Initial state; each task is provided with an arrival time ($t_d$) and a deadline time ($t_d$) which refer to one of the constraints in the system. There is no transition when the system is idle which means there is no incoming task. Checking state: it performs several tasks:

- Checks if the task deadline can be met or not; if not, it forwards the task to the Failed state to restart its cycle. Otherwise, it moves to a next condition.
- Checks available resources for execution; if not, it sends the task to the Failed state. Otherwise, it performs the next operation.
- Checks if the queue in the Execution state, which is also called the Processing state as shown in fig. 2, is full or not; if not, then it forwards the task into the Execution “Processing” state. Otherwise, it forwards it into the waiting state.

The state itself is decomposed into another sub-FSM which contains 3 states: Receiving and checking state, Decision State and Recording state. More information can be found in [1]. The Execution “Processing as in fig. 2” state: represents the place where the task is executed. If the execution is done successfully, it sends the task to the Completed state. Otherwise, it sends it to the Failed state. The execution is completed successfully if the execution time ($t_e$) $<=$ deadline time ($t_d$). The state is decomposed into another two sub-FSMs to form hierarchical model as shown in fig. 2. In the Waiting state, the task waits its turn to be executed once the queue of the Execution “Processing” state is not full or the Processing Unit “P.U.” becomes available when the deadline time can be met; otherwise, the task is sent to the Failed state to restart its cycle. The Suspend state contains tasks for which their computing resources are not yet available and there is a high chance to be executed once their resources become available while the deadline time can be met. There are 16 states in the developed HGFSM which form the complete model. The HGFSM in fig. 2 can be remodeled by including the Suspend state inside the Checking state, more information about it can be found in [1].

The hierarchical generic FSM is mapped to a Markov model “MM” which represents a state diagram (a component of HPM) [1,2]. The Markovian model has 3-tuples $\{S, A, P\}$, where $S$ represents a set of states existed in the HGFSM model. $A$ denotes a vector of initial probabilities values for all states in the model while $P$ contains a matrix that represents transition probabilities among states according to some circumstances existed in the developed model. The mapping is done as follows: 1. Every state in the HGFSM is mapped to a state in the MM. 2. Each edge in the HGFSM is converted to a transition arrow $q_{ij}$ which represents the flow direction from the current state ($S_i$) to the next state ($S_j$). 3. Each transition arrow is associated with a parameter $k_{ij}$ which represents a number of tasks that go from state $S_i$ to state $S_j$. That parameter is used to calculate the probability value $P_{ij}$ which denotes the possibility of moving from the current state ($S_i$) to the next state ($S_j$) and it is calculated using the following equation, where $N_i$ represents a number of total tasks in state $S_i$:

$$P_{ij} = K_{ij} / N_i \quad (1)$$

4. Each FSM graph and state is associated with its Computation Structure Model “CSM” to show data flow in the system under consideration. CSM helps in constructing performance metrics equations. 5. If applicable, a state is decomposed into another sub-FSM and additional Markovian model is constructed to create another level in the hierarchy. A figure for the Markovian model is not shown in this paper due to the space limitation. However, the interested readers can find more information about it in [1].

Performance modeling evaluation is considered to be the abstraction of the functional and performance characteristics of the system under consideration which are combined to determine if it meets the performance requirements based on a user demands and system architectures [6,16]. The Hierarchical Performance Model (HPM) layers are illustrated in figure 3. There are 4 layers in fig. 3 which are used to derive the average power consumption and the average energy produced in the system(s) under investigation.

Fig. 3: Hierarchical Performance Model stack layers
IV. CASE STUDY: ANDROID

Android software architecture is designed and built as a stack structure as depicted in fig. 4(a) [18].

All components of the 4 layers integrate with each other to form what is known today as ANDROID. Each layer contains some blocks that integrate together to do a set of specific jobs. The 4 major layers “levels” in Android from bottom to top are: A. LINUX KERNEL, B. LIBRARIES AND ANDROID RUNTIME, C. APPLICATION FRAMEWORK and D. APPLICATIONS. More information can be found in [18,19,20,21].

Any application in Android is built based on 4 different components which are: 1- Activity, 2- Content provider, 3- Service and 4- Broadcast receiver [19,20,21]. In Android, a task can be defined as an activity or a set of activities. A typical lifecycle for any activity in Android has 7 states which are: OnCreate, OnStart, OnResume, OnPause, OnRestart, OnStop and OnDestroy. Interested readers are referred to [18,19] for more details about Android. Table 1 illustrates the mapping between the HGFSM and the Android activity lifecycle.

Table 1: Mapping HGFSM with the Android activity lifecycle

<table>
<thead>
<tr>
<th>Current State</th>
<th>Transition</th>
<th>Next State</th>
<th>Transition</th>
</tr>
</thead>
<tbody>
<tr>
<td>OnCreate</td>
<td>Initial</td>
<td>OnResume or OnStart</td>
<td>OnResume or OnStart</td>
</tr>
<tr>
<td>OnResume</td>
<td>OnStart</td>
<td>OnResume</td>
<td>OnResume or OnStart</td>
</tr>
<tr>
<td>OnPause</td>
<td>OnPause</td>
<td>OnPause</td>
<td>OnPause or OnStop</td>
</tr>
<tr>
<td>OnStop</td>
<td>OnStop</td>
<td>OnStop</td>
<td>OnStop or OnRestart</td>
</tr>
<tr>
<td>OnRestart</td>
<td>OnRestart</td>
<td>OnStop</td>
<td>OnStop or OnRestart</td>
</tr>
<tr>
<td>OnDestroy</td>
<td>OnDestroy</td>
<td>OnStop</td>
<td>OnStop or OnRestart</td>
</tr>
<tr>
<td>OnPause</td>
<td>OnPause</td>
<td>OnPause</td>
<td>OnPause or OnStop</td>
</tr>
<tr>
<td>OnStop</td>
<td>OnStop</td>
<td>OnStop</td>
<td>OnStop or OnRestart</td>
</tr>
<tr>
<td>OnRestart</td>
<td>OnRestart</td>
<td>OnStop</td>
<td>OnStop or OnRestart</td>
</tr>
<tr>
<td>OnDestroy</td>
<td>OnDestroy</td>
<td>OnStop</td>
<td>OnStop or OnRestart</td>
</tr>
</tbody>
</table>

Fig. 4(b) shows a typical overview of how a task starts and runs on an Android device.

Fig. 4(b): Task execution in Android

In fig. 4(b), solid lines indicate the control flow between the states whereas the dashed lines indicate a message that is sent among states. Once the software processes, which are displayed as the states, and the interface messages between all states are known, our next step is to determine the performance parameters associated with the graph. These parameters are: Tasks arrival rates $\lambda$, number of tasks exist in each state before processing them $N_i$, number of tasks move from the current state ($S_i$) to the new state ($S_j$) $K_{ij}$, flow probabilities $P_{ij}$, message multipliers $\beta_{ij}$ which are assumed to be unity, and lastly the computation and communication cost (service) times $E_i$. To utilize the performance parameters, at the early stage, we identify the input(s), output(s) and divide any Android system into different components if possible as shown in figure 4(c). There are
one input, one output, six components (one action, one sequence and four branches).

![Fig.4(c): System Components](image)

To determine the probabilities values, we need to know how many tasks ($N_i$) exist first in each state and then how many tasks ($k_j$) out of $N_i$ are sent from state $S_i$ to state $S_j$; all these numbers should be known in advance either by obtaining them from actual tests (experiments/simulation) or given by the designers. Several experiments were performed in order to obtain the values for different performance parameters to estimate the expected average power consumption and/or energy produced. Table 2 shows the number of tasks exist in each state while the system under investigation is running. Note that the subscript indicates the state (ID). The system within this paper will be represented by a smartphone, more specific, Galaxy Note 3 is the system to be used and tested.

**Table.2: Number of tasks in each state**

| Initial | Check | Wait | Execution | Failed | Comple |te |
|---------|-------|------|-----------|--------|--------|
| $N_1$   | $N_2$ | $N_3$ | $N_4$     | $N_5$  | $N_6$  |

The probability value $P_{ij}$ is computed using equation (1); the next step is to specify the details of the methods used to derive the power consumption equation(s). The software structure indicates the order in which the operations are executed in order to complete a desired task or computation. The software structure can be seen as the Computation Structure Method (CSM) which consists of Data Flow Graph DFG and Control Flow Graph CFG. The DFG and CFG for the Android application are not shown due to the space limitation. The interested readers are referred to [1] for more information and DFG and CFG. To derive a cost equation, we multiply each state cost with its associated flow parameter; then sum all remaining ones. The expected average power consumption is computed as follows:

$$PC_{\text{average}} = (1 \times C_{\text{initial}}) + \left( (1 + e_2) \times (C_{\text{check}} + C_{\text{test}}) \right) + \left( (1 + e_1) \times C_{\text{delete}} \right) + \left( (1 + e_3) \times (C_{\text{await}} + C_{\text{test}}) \right) + \left( (1 + e_4) \times (C_{\text{create}} + C_{\text{test}}) \right)$$

(2) PC stands for Power Consumption. To compute the energy produced from the system, just multiply the result from eq. (2) with a time ($t$) spent to perform the desired jobs. In equation (2), each parameter is associated with its flow variable(s) which is/are denoted by $e$. Each flow variable represents a value of moving through a path from a start node to an end node in the CFG. Each flow takes a value between $\{0,1,...,\infty\}$ and mainly depends on a type of distribution [1,2]. The flows also represent the data dependent aspects of the computation time [1]. They are discrete random variables and are modeled using probability distribution and statistics methods. Several probability distributions exist which are summarized as follows: Bernoulli, Binomial, Geometric, Modified Geometric and Poisson [1,2]. Given the probability distribution type of $e$, several characteristics such as Expected value $E(e)$, second moment $E(e^2)$, Variance $\text{Var}(e)$ and the coefficient of variation $C^2_e$ are easily obtained. More information about estimating the values of computations and communication aspects can be found in [1,3]. The value of $C_{\text{initial}}$ is determined from its CFG as shown in fig. 5(a). From fig. 5(a), three operations take place in every Android device which are: Assigning thread to perform a desired action, Variables initialization and finally creating the Graphical User Interface to interact with a user.

![Fig. 5(a): CFG for the Initial state](image)

From fig. 5(a), the equation for estimating the average power consumption is computed as follows:

$$C_{\text{initial}} = C_{\text{create UI}} + C_{\text{initializations}} + C_{\text{assigning thread}}$$

(3) For the Executing “Processing” state, its cost is obtained from its CFG which is now shown due to the space limitation; so

$$C_{\text{execution}} = [(1 + e_1) \times (C_{\text{handling state}} + C_{\text{test}})] + [e_4 \times C_{\text{shaded}}] + [(1 + e_3) \times C_{\text{test}}]$$

(4)
Substitute the value of $C_{\text{handling}}$ which is obtained from its CFG as depicted in fig. 5(b).

$$C_{\text{handling}} = [(I + e_3 + e_6) \cdot (C_{\text{ready}} + C_{\text{test}}) + (e_3 \cdot C_{\text{idle}}) + [(e_3+1)*C_{\text{test}}+(1*C_{\text{run}})]$$

(5)

Note that the cost value obtained for each state represents the expected power consumption which is used for the computation in a Node View in the system level [1].

The aim of this profiling is to determine the expected average power consumption and to identify which part in the smartphone produces more energy. All applications were tested several times (about 50 times) and then the average power is determined. In the architecture, all four applications were installed and then the profiling started by launching them one by one. For flows in equations (2) to (5), they occur in a single run of “if statement” and we assume equally likely for a branch to be taken so $p = q = 0.5$ since $p + q = 1$ as stated earlier.

Finding the number of visits in each state is determined using Matlab and then substituting in equation (2) will give the expected average power cost as follows;

$$E(PC) = (1*C_{\text{initial}})+(E(1+e_4)*(C_{\text{check}}+C_{\text{test}}))+(E(e_1+1)*C_{\text{decision}})+(E(e_9+e_8)*(C_{\text{wait}}+C_{\text{test}}))+(E(e_1+1)*(C_{\text{exe}}+C_{\text{test}}))$$

(8)

where $E$ refers to Expected average value, so the average energy produced is calculated using the following equation:

$$E(\text{energy}) = E(\text{Power}) \cdot t$$

(9)

Where $t$ represents the period of time taken to run the applications and it is fixed to be 45s = 45000ms. Tables 3 shows the actual values and expected ones for power consumption for all 4 applications in Note 3.

<table>
<thead>
<tr>
<th>Application Name</th>
<th>Expected power (mw)</th>
<th>Actual Power (mw)</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video</td>
<td>2.215</td>
<td>1.997</td>
<td>10.92%</td>
</tr>
<tr>
<td>Calculator</td>
<td>1.743</td>
<td>1.856</td>
<td>6.09%</td>
</tr>
<tr>
<td>Audio</td>
<td>2.189</td>
<td>1.972</td>
<td>11%</td>
</tr>
<tr>
<td>Taking Picture</td>
<td>2.206</td>
<td>2.033</td>
<td>8.51%</td>
</tr>
</tbody>
</table>
The difference between the expected and actual average power consumption lays within a small error “±11” as shown in table 3. Table 4 illustrates the actual and the expected average energy produced in Note 3 in milli joule “mj”; all values are scaled by $10^3$.

**Table 4: expected and actual energy produced in Note 3**

<table>
<thead>
<tr>
<th>Application Name</th>
<th>Expected Energy (mj)</th>
<th>Actual Energy(mj)</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video</td>
<td>99.675</td>
<td>89.865</td>
<td>10.92%</td>
</tr>
<tr>
<td>Calculator</td>
<td>78.435</td>
<td>83.52</td>
<td>6.09%</td>
</tr>
<tr>
<td>Audio</td>
<td>98.505</td>
<td>88.74</td>
<td>11%</td>
</tr>
<tr>
<td>Taking Picture</td>
<td>99.27</td>
<td>91.485</td>
<td>8.51%</td>
</tr>
</tbody>
</table>

Table 5: List of primitive operations and their average power consumption values

<table>
<thead>
<tr>
<th>List of primitive operations</th>
<th>Average value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function call</td>
<td>0.0437</td>
</tr>
<tr>
<td>Addition</td>
<td>0.119</td>
</tr>
<tr>
<td>Subtraction</td>
<td>0.147</td>
</tr>
<tr>
<td>Multiplication</td>
<td>0.492</td>
</tr>
<tr>
<td>Division</td>
<td>0.687</td>
</tr>
</tbody>
</table>

Fig. 6 shows the actual and expected energy produced in Note 3 by all four applications with scale by $10^3$.

![Energy produced in Note 3 in mj](image)

**Fig. 6: expected and actual energy in Note 3**

Samsung tried to minimize the energy produced in its devices and Note 3 yields less than expected even though that several applications were already installed and run on it. During the experiments, several applications ran on Note 3 and caused the energy to be higher than the expected. Also the initial state, which includes several activities (assigning thread, variables initializations and creating the GUI), consumes more power than the remaining states. In general, the Note 3 device is optimized in terms of the power consumption and the execution time as proved by the experiments. Table 5 illustrates the average values for several primitives operations in Samsung Galaxy Note 3 after performing different experiments. All values are in milli watt “mw” and were obtained after repeating the experiments more than 45 times.

![Average power consumption in two benchmarks](image)

**Fig. 7: Average power consumption in two benchmarks**

In each benchmark application, the first bar refers to the actual average power consumption while the second bar represents the estimated average values.

V. CONCLUSION

This paper presented the developed framework to estimate the expected average power consumption and energy in smartphones which incorporated different levels of abstraction. Two examples were given to demonstrate how the power and energy are estimated using the software structure and the architecture to represent the corresponding levels of the model. The framework can be applied on any smartphone to estimate the expected average values for the power consumption and the energy produced. The future work is to determine the average power dissipation and energy consumed with a device runs on multiple threads “up to three threads” and find the effect of rendering GPU(s) to take control on creating graphical task(s) such as drawing the UI window which is found to be the dominant factor. It takes about 75% of the average power cost in the Initial state.

www.ijaems.com
REFERENCES


