# SENSORS & TRANSDUCERS

ISSN 1726-5479

vol. 212

5 /17


International Frequency Sensor Association Publishing

IFSA

# Sensors & Transducers

**International Official Open Access Journal of the International Frequency Sensor Association (IFSA) Devoted to Research and Development of Sensors and Transducers**

Volume 212, Issue 5, May 2017

---

**Editor-in-Chief**
Prof., Dr. Sergey Y. YURISH

IFSA Publishing: Barcelona • Toronto

Published by International Frequency Sensor Association (IFSA) Publishing. Printed in the USA.

Sensors & Transducers

Volume 212, Issue 5, May 2017

www.sensorsportal.com

e-ISSN 1726-5479
ISSN 2306-8515



Sensors & Transducers journal is an open access, peer review international journal published monthly by International Frequency Sensor Association (IFSA). Available in both: print and electronic (printable pdf) formats.

Sensors & Transducers Journal

# Contents

Volume 212
Issue 5
May 2017

www.sensorsportal.com

ISSN 2306-8515
e-ISSN 1726-5479

## Research Articles





International Frequency Sensor Association (IFSA).

*Sensors & Transducers, Vol. 212, Issue 5, May 2017, pp. 1-14*

**Sensors & Transducers**

Published by IFSA Publishing, S. L., 2017
http://www.sensorsportal.com

# Game-Theoretic Optimal Power-Link Quality Topology Control in Wireless Sensor Networks

**[1] Evangelos D. SPYROU, [2] Shusen YANG and [1] Dimitrios K. MITRAKOS**

[1] School of Electrical and Computer Engineering, Aristotle University of Thessaloniki
[2] Institute of Information and System Science, School of Mathematics and Statistics, Xi'an Jiaotong University
E-mail: evang spyrou@eng.auth.gr, shusenyang@mail.xjtu.edu.cn, mitrakos@eng.auth.gr

*Received: 12 April 2017 /Accepted: 12 May 2017 /Published: 31 May 2017*

**Abstract:** One of the most significant problems in Wireless Sensor Network (WSN) deployment is the generation of topologies that maximize transmission reliability and guarantee network connectivity while also maximizing the network's lifetime. Transmission power settings have a large impact on the aforementioned factors. Increasing transmission power to provide coverage is the intuitive solution yet with it may come with lower packet reception and shorter network lifetime. However, decreasing the transmission power may result in the network being disconnected. To balance these trade-offs we propose a discrete strategy game-theoretic solution, which we call TopGame that aims to maximize the reliability between nodes while using the most appropriate level of transmission power that guarantees connectivity. In this paper, we provide the conditions for the convergence of our algorithm to a pure Nash equilibrium as well as experimental results. Here we show, using the Indriya WSN testbed, that TopGame is more energy-efficient and approaches a similar packet reception ratio with the current closest state of the art protocol ART. Finally, we provide a methodology for further optimization of our work using an indicator function to distinguish between satisfactory and poor links.

**Keywords:** Transmission power, Transmission reliability, Potential game, Pareto optimality, Packet reception ratio.

## 1. Introduction

A significant problem in Wireless Sensor Network (WSN) topology management is to guarantee connected network topologies that have a high transmission reliability. The simple approach would be to increase the radio transmission power levels of unconnected nodes. However, this is too simple and does not account for the complexities of the wireless channel. An increase in transmission power might cause an increase in interference, decreasing the number of packets received (i.e. lowering Packet Reception Ratio, or PRR). On the other hand, as we see in [43], if the distance between the transmitter-receiver and interferer-receiver is difference by approximately a factor of 2, interference does not cause packet loss. This indicates that a node may select a high transmission power level, in order to strengthen its signal, without suffering from packet loss. There is a sweet spot in PRR related to transmission power levels that can keep PRR to a high level while not using a larger transmission power level than necessary. The transmission power also affects the energy consumption of the node, directly influencing the lifetime of the WSN [4]. In order to handle this trade-off we present a discrete strategy distributed game-theoretic approach that maximizes each node's PRR while using the optimal transmission power from an optimization problem; guaranteeing connectivity. We call our approach TopGame.

Specifically, we focus on the trade-offs between energy consumption, and PRR. We use game theory, since it can appropriately describe the behavior of selfish nodes and find an optimal solution in a distributed manner. Modeling systems with selfish algorithms have been shown to provide efficient solutions that improve network performance [50]. We consider nodes to be individual players that play selfishly in order to find a best response for their objectives. In this paper we present our model as it is in [44]. We prove that this game is a potential game [32]. Potential games are games where the incentive of players to change their strategy can be expressed in a single global function, the potential function. Potential games have been used in wireless networks in a plethora of problems, including power control [21, 42], cognitive radio [36], gateway selection [41] and channel allocation [11]. In our game-theoretic formulation we prove that there is an equilibrium point. Next, we provide testbed results to show the convergence of our proposed algorithm and we compare it to the closest state of the art algorithm, Adaptive Robust Topology (ART)[1], with respect to connectivity, energy-efficiency and PRR. To our knowledge this is the first practical topology control game that has been evaluated on a real testbed system, and show the following:

- TopGame exhibits slightly lower network PRR than ART, since it exploits a Transmission Reliability metric to determine each node's final transmission power.
- Using TopGame, the network's relative energy consumption is less by 5 % than ART's, due to the fact that TopGame can use a per node transmission power setting, which remains so after the optimization process. Also, connectivity is preserved.
- TopGame's operation increases contention for accessing the wireless medium, since it keeps a steady transmission power level and it includes the bootstrapping period. This explains the slightly less PRR of our approach.
- TopGame includes mathematical proofs to support the convergence of each node's transmission power in the form of the Nash equilibrium of a potential game.
- We prove that the Price of Stability and Price of Anarchy of TopGame is 1. This shows that TopGame can find the optimal equilibrium of the game.
- We formulize participation in the maximization process of nodes with PRR less than 20 % using an indicator function and we convert the problem to its equivalent using a sigmoidal function. We employ a non-convex optimization technique to find the near-optimal solution, using the dual problem and the duality gap.
- We force a heuristic to stabilize the neighbourhood when the duality gap is positive.

The paper is structured as follows: Section 2 provides the related work, Section 3 introduces topology control in WSN, Section 4 describes game theory basics and potential games, Section 5 formally describes TopGame, Section 6 shows the experimental results obtained, Section 7 presents a discussion on further optimization and Section 8 presents the conclusions.

## 2. Related Work

The characteristics and behaviors of wireless links are now more understood. There has been work measuring the effects of varying power levels and showing the irregularity of radio ranges and the lack of link symmetry [39, 51]. The relationship between PRR and RSSI for the Chipcon CC2420 radio was established in [28]. Subsequent work then looked at the differences in behavior between indoor and outdoor networks, and fluctuations in link quality over longer durations of time [19].

Regarding of Topology Control (TC) specifically, [19] contributes a comprehensive review of this field which we summarize. Given the diversity of link behaviors influenced by their environment, experimentation for much of the early TC work was carried out using graph theory and simulation studies for tractability reasons. Yet, this work did not consider aspects like realistic radio ranges, node distributions or node capability/capacities into account, limiting their usefulness for real sensor networks [26, 27, 9, 6, 17]. For example, some have assumed that link costs are proportional to link length, but in reality a more complex relationship is evident [39, 15, 51]. The main competitors in the practical Topology control area are PCBL [39] and ART [19], which we introduce next.

PCBL was derived from link quality observations showing that links with a very high PRR remain quite stable. They then categorize links as blacklisted, middling or highly reliable. The power in the latter is minimized to their lowest stable power setting while the blacklisted are not used at all. The middling links are those that lie between the two and are set to full power. Given the expense of probing the network to establish the link categories, this protocol cannot work with dynamic routing protocols such as CTP [18]. CTP aims to find the least expensive routes through the network. To overcome such link probing, link quality metrics have been used to approximate PRR in ATPC [28]. Specifically there is a link between RSSI and PRR, and LQI and PRR over a monotonically-increasing curve. Further, linear correlations between transmission power levels and RSSI/LQI are observed at the receiver but are different for each environment monitored. Therefore, ATPC estimates the slope and uses closed feedback to adjust the model to the current situation to achieve lower bound RSSI (PRR).

[1] Note that by ART we mean the optimized ART.

Hackmann et al., showed that RSSI and LQI cannot always realistically estimate PRR in indoor environments [19], nor can instantaneous probing represent the behaviors of a link over time. They propose ART, which does not rely on estimates of link quality nor does it involve long bootstrapping phases. Being more dynamic, ART adapts link power to changes in the environment as well as contention using a gradient. Also, where applications expect acknowledgment messages, ART can piggyback these to reduce communication overhead. ART selects the appropriate transmission power based on the failures observed when the target PRR is 95 % and a contention gradient.

In [20], the authors proposed a distributed topology control and channel allocation game-theoretic algorithm. The main objective of the work is the relief of interference and the energy consumption balancing. They examined the connection between topology control and channel allocation. They designed a game-theoretic model that takes into account transmission power, energy consumption and interference suffered by a node. They have proven the existence of Nash Equilibrium and they developed an algorithm that preserves connectivity by jointly setting the transmission power and channel. Lastly, their algorithm converges to Pareto optimality.

Tan et al. [46], suggested a topology control scheme where every node tunes its transmission power adaptively, in order to use its harvested energy in an efficient manner. The authors, proposed an ordinal potential game model where high harvesting nodes cooperate with the low harvesting nodes to ensure network connectivity. They proved the existence of a Nash Equilibrium and they designed an algorithm that achieves it.

Abbasi at al. [1], investigated the issue of topology control in wireless sensor networks, in order to perform energy consumption minimization and energy balancing. Their approach accomplished their objectives by adjusting transmission power on the nodes and preserving connectivity. The authors utilized a game-theoretic scheme to address energy welfare topology control. They showed that their proposed game-theoretic solution is a potential game and it achieves a unique Nash equilibrium, which is Pareto optimal as well.

Nahir et al. [34], provided a game-theoretical solution to the topology control problem, by addressing three major issues: the price of establishing a link, path delay and path congestion proneness. They established that bad performance due to selfish play in the considered games is significant, while all but one are guaranteed to have a Nash equilibrium point. Furthermore, they showed that the price of stability is typically 1; hence, often optimal network performance can be accomplished by being able to impose an initial configuration on the nodes. Furthermore, the authors express their concern regarding the computational tractability of their solution.

Komali et al. [24], analyzed the creation of energy efficient topologies with two proposed algorithms. Specifically, their game-theoretic model specified that nodes have the incentive to preserve connectivity with a sufficient number of neighbours and that the network will not partition. They proved that their game is an exact potential game and that a subset of the resulting topologies is energy efficient. They addressed the major issue of fair power allocation by providing the argument of efficient allocation vs fair allocation.

## 3. WSN and Topology Control

Wireless sensor networks are networks of small computational devices fitted with radio transceivers for communication and sensors to capture data. Topology control can be defined by the construction of a graph that represents the nodes and links in the network that does not consist of any disjoint parts. Good topology control mechanisms can be characterized by providing an energy efficient network, offering high throughput and doing so with a low overhead. Energy-efficiency equates to the use of the minimum transmission power that guarantees connectivity, where throughput can be maximized by reducing interference and contention on the wireless medium. However, minimum transmission power does not guarantee a high reliability of transmission resulting in high throughput. This is due to a weak signal that may be significantly influenced by a small portion of interference.

For the most part, hitherto link asymmetry has been ignored, and the use of different transmission power levels when a node transmits to different neighbours may cause undesired packet loss. In addition, in a dense network, a node having a large number of neighbours may not be able to cope with transmission power changes when unicasting to different recipients in that neighbourhood while expecting to achieve a high PRR as well. As observed by Ahmed et al. [2], environmental effects and different node transmission powers are the major cause of link asymmetry in WSNs.

## 4. Game Theory and Potential Games

Game theory studies mathematical models of conflict and cooperation [49], between nodes in our work. Therefore, our meaning of the term game corresponds to any form of social interaction between two or more nodes. The rationality of a node is satisfied if it pursuits the satisfaction of its preferences through the selection of appropriate strategies. The preferences of a node need to satisfy general rationality axioms, then its behavior can be described by a utility function. Utility functions provide a quantitative description of the node's preferences and the main objective is therefore the maximization of its utility function. In this work, we focus on strategic noncooperative games, since we consider nodes to act as selfish players that want to preserve their interests. The intuition behind this is that the nodes will reach an optimal state, without having to pay a price to maximize their payoffs. The Nash equilibrium is the most important equilibrium in non-cooperative strategic form games. It is defined as the point where no node will increase its utility by unilaterally changing its strategy. It got its name from John F. Nash who proposed it [35].

In 2008, (Daskalakis et al., 2008) Daskalakis proved that finding a Nash equilibrium is PPADcomplete. Polynomial Parity Arguments on Directed graphs (PPAD) is a class of total search problems [38] for which solutions have been proven to exist, however, finding a specific solution is difficult if not intractable. This development lead researchers to concentrate a specific class of games called 'Potential Games', due to the important properties that pure Nash equilibria will always exist and best response dynamics are guaranteed to converge.

This class of games consists of the exact and ordinal potential games. In this paper we utilise exact potential games and refer the reader to [32] for details on potential games. In order to use exact potential games, it is essential to have a potential function that has the same behavior as the individual utility function, when a player unilaterally deviates.

More formally:

A game $G\langle N,A,u\rangle$, with *N* players, *A* strategy profiles and *u* the payoff function, is an exact potential game if there exists a potential function

$$V : A \to \mathbb{R} \tag{1}$$

subject to

$$\forall i \in N, \forall \sigma_{-i} \in A_{-i}, \forall \sigma_i, \sigma_i' \in A_i \tag{2}$$

where $\sigma_i$ is the strategy of player *i*, $\sigma'_i$ is the deviation of player *i*, $\sigma_{-i}$ is the set of strategies followed by all the players except player *i* and $A_{-i}$ is the set of strategy profiles of all players except *i* such as

$$V(\sigma_i,\sigma_{-i}) - V(\sigma_i',\sigma_{-i}) = u_i(\sigma_i,\sigma_{-i}) - u_i(\sigma_i',\sigma_{-i}) \tag{3}$$

## 5. TopGame

We developed the TopGame algorithm that aims to guarantee connectivity, by locating the best response of PRR and transmission power. The intuition behind this research is that TopGame will force nodes to converge to the best transmission power.

A WSN consists of a set of nodes *N* and each node $i \in N$ can switch its transmission power $p^k_i \in P$, where $k \in 3,7,11,15,19,23,27,31$ and $P$ is the set of the available transmission power levels of our example CC2420 transceiver. In this paper, we employ 4 transmission power levels, namely 11,15,19,23, in order to identify the PRR when transmission powers that operate mostly on the gray area [40] are used. Let a vector P = ($p_1$, $p_2$, ..., $p_{|N|}$) be an allocation of the transmission power level of each sensor node. The total number of possible power allocations is $4^{|N|}$. The aim of this paper is to determine a power allocation in a distributed way, which can achieve a best response rade-off between network connectivity, energy-efficiency and transmission reliability, using game theory.

### 5.1. Connectivity Definition and Measurement

In this paper we consider the small-world Model A from [16], where there are *N* nodes in the network and each one arbitrarily selects *m* nearest neighbours to connect to. Essentially, we utilise the variant of this small-world model, where node locations are being modeled by a stochastic point process. The number of neighbours consists of nearest neighbours and shortcuts. A shortcut is an edge between two nodes if either of the two nodes exist in the nearest neighbour set of the other. If a node is connected by a nearest neighbour and a shortcut, multiple edges are replaced by a single one. The presence of the shortcuts reduces the network diameter. Furthermore, we have to note that *m* is the number of neighbours a node has in terms of a spatial graph, and $(N-1)p$ is the number of neighbours it has via shortcuts. In order to ensure connectivity the quantities $m = (1+\delta)\sqrt{2\log(N)}$ and $Np = (1+\delta)\sqrt{2\log(N)}$, where $\delta > 0$, are sufficient. Hence connectivity is preserved with a smaller degree of (nearest neighbours plus shortcuts). We select a degree of 6 for each node. It is well known that the node degree can be reached by adjusting the transmission power; hence, the transmission power level that satisfies connectivity satisfies the condition that more than 6 nodes exist in the neighbourhood of each node.

### 5.2. Transmission Reliability (TR)

For a wireless link (*i*, *j*), the Packet Reception Ratio $PRR_{i,j}$ is defined as the ratio of the number of packets received by node *j* over the number of packets sent by node *i*. It can be expressed by approximation as

$$PRR_{i,j} = (1-\xi_{i,j})^l, \tag{4}$$

where l is the packet length in bits.

The Bit Error Rate (BER), which we denote as $\xi_{i,j}$, is given by the following formula [14]

$$\xi_{i,j} = \frac{1}{2}\left(1 - \sqrt{\frac{\gamma_{i,j}}{1+\gamma_{i,j}}}\right) \quad (5)$$

where $\gamma_{i,j}$ is the Signal-to-Interference-plus-Noise Ratio (SINR) of the transmission from node *i* to node *j*. $\gamma_{i,j}$ is given by

$$\gamma_{i,j} = \frac{H_{i,j}p_i}{\sum_{t\neq i,t\neq j} p_t H_{t,j} + N_0} \quad (6)$$

where $N_0$ is the white noise and $H_{i,j}$ is the channel gain of the wireless link (*i, j*) and $H_{t,j}$ is the channel gain between the receiver and an interferer. Due to the path loss, the larger the distance between nodes *t* and *j* the smaller the $H_{t,j}$. We focus on static WSNs, hence, we assume that the channel is slow fading in nature and the channel gain of every link remains constant before the convergence of the TopGame algorithm.

To measure the reliability of links around node *i*, we define a new metric called Transmission Reliability ($TR_i$) as

$$TR_i(p_i,p_{-i}) = \frac{\sum_{j\in N_i(p_i,p_{-i}),k\in N_j(p_i,p_{-i}),k\neq i} PRR_{k,j}}{\left|\bigcup_{j\in N_i(p_i,p_{-i})}(N_j(p_i,p_{-i})-\{i\})\right|} \quad (7)$$

where $p_i$ is the power level of node *i*, $p_{-i}$ means the power levels of all nodes except *i*, $N_i(p_i, p_{-i})$ is the set of nodes such that $\forall j \in N_i(p_i, p_{-i})$, $PRR_{i,j} > 0$.

For instance, Fig. 1 shows a sub-graph of a WSN for a given transmission power allocation. For each link (*i, j*) $PRR_{i,j} > 0$. In this sub-graph $N_6 = 2,15,7,11$ and $\beta TR_6 = (PRR_{1,2} + PRR_{15,2} + PRR_{12,2}+PRR_{2,15}+PRR_{9,15}+PRR_{7,15}+PRR_{15,7}+PRR_{10,7}+PRR_{11,7}+PRR_{7,11}+PRR_{14,11})/11$.

In practice, every node *i* can obtain $TR_i$ at run time by every node *j* in $N_i$ calculating the $\overline{PRR_{k,j}}$ as the average $PRR_{k,j}$, $k \in N_j - \{i\}$ and periodically broad-casting $\overline{PRR_{k,j}}$. Thereafter node *i* calculates $TR_i$.

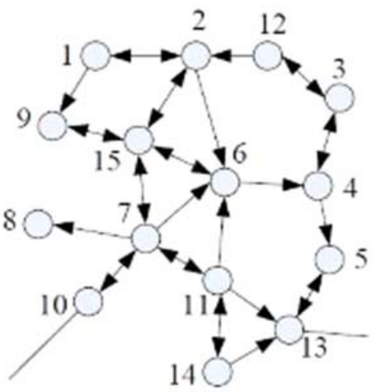



**Fig. 1.** An example to explain Transmission Reliability metric.

## 5.3. Utility and Potential Function

We define the utility function of each node *i* as,

$$u_i(p_i) = TR_i - c_i p_i \quad (8)$$

where $c_i$ is the price assigned to each strategy played by a node/player.

Our strategy domain consists of 4 strategies, which are 11,15,19,23, which correspond to the values in table 1. Notably, the 2 smallest and 2 largest transmission power levels of the CC2420 radio have been excluded. The main reason is to see TopGame operate under medium to large SINR regime. The second reason is to simplify TopGame.

**Table 1.** Transmission Power Levels and Values.

| PA LEVEL | dB | mA |
|---|---|---|
| 11 | -10 | 11.2 |
| 15 | -7 | 12.5 |
| 19 | -5 | 13.9 |
| 23 | -3 | 15.2 |

It is straightforward to see that the above utility function has a minimum under the following condition of medium to high SINR values. We do the price assignment in a similar way with [10]. The prices assigned at every node has the value 1 except when it reaches its maximizer. Each node then assigns the price given below:

$$c_i = diff(TR_i) \quad (9)$$

Hence, if we take the first derivative to obtain the minimum, it follows that there is a local minimum. Since we wish to maximize the function we simply take the negative of (8).

$$u_i(p_i) = c_i p_i - TR_i \quad (10)$$

Thereafter we wish to define the potential function and prove that the game *G* is a potential game.

**Proposition 1.** The game G is a potential game. The potential function is given by

$$V(\mathbf{p}) = \sum_i c_i p_i - \sum_i TR_i,\ p_i \in A \quad (11)$$

*Proof.* This comes as a result by taking the characterisation of the potential games in [32] where $\frac{\partial V(\mathbf{p})}{\partial p_i} = \frac{\partial u_i(\mathbf{p})}{\partial p_i}, i \in N$

$$V(p_i,p_{-i}) - V(p_i',p_{-i}) = u_i(p_i,p_{-i}) - u_i(p_i',p_{-i}) + \sum_{m\in N,m\neq i}^{N}(u_m(p_m,p_{-m}) - u_m(p_m',p_{-m}))$$

Since only one node can deviate

$$\sum_{m\in N, m\neq i}^{N} (u_m(p_m, p_{-m}) - u_m(p'_m, p_{-m})) = 0$$

Hence we conclude that Γ is an exact potential game. This proof comes as a result of the fact that given a strategy of a node/player $m$, $p_m \in N$ and an alternative strategy $p'_m \in N$ and taking the assumption that the strategies of all the other nodes remain the same, we have

$$u_i(p_i, p_{-i}) - u_i(p_i, p'_{-i}) = | V_i(p_i, p_{-i}) - V_i(p_i, p'_{-i}), \quad (12)$$

where $p_{-i}$ is the transmission power strategy of all the nodes excluding that of the node $i$. Hence, the game is a potential game.

**Remark 3.1:** The potential function is significant since its maximization, when a specific policy is played, results in this policy being an equilibrium of the designed game. In this work, the strategy set is discrete; hence, in the case that the potential function satisfies particular types of concavity, such as the Larger Midpoint Property (LMP) [48], the converse is true as well. If a policy is an equilibrium, it maximizes the potential function. Thus, we may consider the TopGame as the following optimization problem.

$$\hat{p}_i = \arg\max V_i(i) \quad (13)$$

As presented in [3], we consider two n-dimensional vectors δ(1), δ(2). *Definition 1:* [29] A vector δ(2) majorises δ(1), which we denote as δ(1) ≺ δ(2), if δ(2) is more 'irregular' in the following fashion:

$$\begin{cases} \sum_{i=1}^{k} \delta_{[i]}(1) \leq \sum_{i=1}^{k} \delta_{[i]}(2), k = 1, 2, ..., n-1 \\ \sum_{i=1}^{k} \delta_{[i]}(1) = \sum_{i=1}^{k} \delta_{[i]}(2) \end{cases}, \quad (14)$$

where $\delta_{[i]}(m)$ is the permutation of $\delta_i(m)$ satisfying the condition $\delta_{[1]}(m) \geq \delta_{[2]}(m) \geq ... \geq \delta_{[n]}(m), m = 1,2$

Equation (8) suggests that the largest element of δ(2) is larger than the largest element of δ(1). Consequently, the smallest element of δ(2) is smaller than the smallest element of δ(1). Thereafter we proceed in Schur convexity properties of majorisation.

*Definition 2:* A function $f : R^n \to R$ is Schur concave if δ(1)≺δ(2) suggests $f(\delta(1)) \geq f(\delta(2))$. $f$ is Schur convex if the inequality suggests that $f(\delta(1)) \leq f(\delta(2))$.

Definition 1 dictates that there is strong majorisation; however, at least one of the inequalities of (8) is strict. Furthermore, Proposition *C*.2 of [29] dictates that a function $f : R^n \to R$ that is symmetric and convex (concave), is also Schur-convex (concave). Hence, we need to show that our potential function is Schur-concave, in order to proceed with the majorisation properties.

**Lemma 1.** Function V is concave in N.

**Proof.** It is obvious that the function is concave, since if we take the second derivative test the first term will be set to 0 and the second term is a concave term (raised to power) for medium to large SINR values. Note that for very high SINR values the second derivative of (10) become positive and the function becomes convex as we can deduct from [30].

**Proposition 2.** If the function u(p) is concave then the function V(p) is Schur concave.

**Proof.** The proof is given by using the following corollary from [29].

**Corollary 5.0.1.** Let $\phi(x) = \sum_{i=1}^{n} g(x)$ where g is concave (convex). Then φ is Schur-concave (convex)

**Theorem 5.1.** The Game G reaches the global optimum via the potential function V(p) maximization.

**Proof.** Recall that the potential $V(p)$ Schur concave and it satisfies the LMP. It follows that if $p^*$ is a Nash equilibrium strategy, then it maximizes the potential and is the global maximum. Assume that there is another strategy profile $p^{0*}$ that maximizes the potential and is the global maximum. This means by $p^*$ majorises $p^{0*}$. Since $V(p)$ is Schur concave it follows by definition that $V(p^{0*}) \geq V(p^*)$. Since, $p*$ maximizes the potential, this is only possible when $V(p^{0*}) = V(p^*)$. Hence, $p^*$ is the global optimum.

This also comes as a result of the fact that we have shown that there is a critical point in the function $V(p)$. It follows from [23] - Theorem 2.22 - that the critical point $p^*$ is the global optimum.

Notably, Schur concavity of $V$ not only allows us to capture the optimal policies, but it allows the comparison of the performance of two non-optimal strategies, whenever one of the policies majorises the other.

**Theorem 5.2.** The price of stability of the game is 1.

**Proof.** It follows from the previous theorem that shows that the game reaches the global optimum.

Thereafter, we will proceed with the derivation of the Price of Anarchy (PoA) [37], in order to further check the optimality of the game. Firstly, though, we start with the following result.

**Definition 5.1.** (Pareto efficient) [33] A strategy profile ($p^{OPT}_i$, $p^{OPT}_{-i}$), is considered to be strongly Pareto efficient if and only if there exists no other strategy profile ($p_i$, $p_{-i}$) such that $u_i(p_i, p_{-i}) \geq u_i(p^{OPT}_i, p^{OPT}_{-i}), \forall i \in N$ and $u_i(p_i, p_{-i}) > u_i(p^{OPT}_i, p^{OPT}_{-i})$ for at least one node $m$. On the other hand, a strategy profile ($p^{OPT}_i$, $p^{OPT}_{-i}$) is weakly Pareto efficient if and only if there exists no strategy profile ($p_i$, $p_{-i}$) such that $u_i(p_i, p_{-i}) > u_i(p^{OPT}_i, p^{OPT}_{-i}), \forall i \in N$. We use the term Pareto efficient for both weak and strong cases.

**Definition 5.2.** A pure strategy NE is a Pareto efficient pure strategy NE if it is Pareto efficient.

**Theorem 5.3.** A maximizer of V, which coincides with the optimal solution of (11), is a Pareto efficient pure strategy NE.

**Proof.** We have shown previously that the game $G$ reaches the maximum which is a pure strategy NE. Hence ($p^{OPT}{}_i$ , $p^{OPT}{}_{-i}$ ) constitutes an optimal solution of (11). There is no other strategy that maximizes the potential. That is that there is no strategy profile ($p_1$,...$p_i$)∈ $P_{i i \in N}$, such that

$$u_i(p_1,...p_i) = V(p_1,...p_i) > u_i(p_m^{OPT}, p_{-i}^{OPT}) \\ = V(p_i^{OPT}, p_{-i}^{OPT}), \forall i \in N \quad (15)$$

Thus, considering Definition 5.2, ($p^{OPT}{}_i$, $p^{OPT}{}_{-i}$ ) is Pareto efficient. Moreover, let us assume the ∀$i$ ∈ $N$, $p_i$ is an alternative strategy of node/player $i$, where $p_i$ 6= $p^{OPT}{}_i$ . Then, we obtain

$$u_i(p_i, p_{-i}^{OPT}) \geq u_i(p_i^{OPT}, p_{-i}^{OPT}) \quad (16)$$

We see that there is no node that can unilaterally change its transmission power/ strategy, in order to increase its utility. Furthermore, the strategy profile ($p^{OPT}{}_1$ ,..., $p^{OPT}{}_i$ ) is also a pure strategy NE. To summarize, ($p^{OPT}{}_1$ ,..., $p^{OPT}{}_i$ ) is a Pareto efficient pure strategy NE.

Since, the game $G$ may have more than one pure strategy NEs, we will check the optimality of the NE to show the relationship between the local optimal NE and the Pareto efficient NE. Even though we have shown that the Game $G$ goes to the global optimum, we will strengthen this proof even further, by evaluating the ratio between the highest utility and the worst case NE, namely the PoA.

**Theorem 5.4.** PoA = 1, i.e. a pure strategy profile of G is Pareto efficient.

**Proof.** We assume that $p^{OPT}{}_i$ is a Pareto efficient NE. Also, assume that $p^*{}_i$ is an arbitrary pure strategy NE $\mathbf{p}_i^* = (p_i^*, p_{-i}^*)$. Then for any arbitrary node/player $i$, we have

$$u_i(\mathbf{p}^*) = V(\mathbf{p}^*) = c_i p_i^* - TR_i^* \quad (17)$$

Note that $u_i$(p*) ≥ $u_i$($p^{OPT}{}_i$ ,$p^*_{-i}$) according to the definition of a game. Therefore, we have

$$u_i(\mathbf{p}^*) = V(\mathbf{p}^*) = c_i p_i^* - TR_i^* \\ \geq u_i(\mathbf{p}^*) = V(\mathbf{p}^{OPT}) = c_i p_i^{OPT} - TR_i^{OPT} \quad (18)$$

Furthermore, since we have assumed that $p^{OPT}$ is a Pareto-optimal pure strategy NE, ∀$i$ ∈ $N$

$$V(p^{OPT}) \geq V(p^*) \quad (19)$$

Combining (18) and (19) we have $V(p^{OPT})$ ≥ $V(p^*)$,∀$i$ ∈ $N$. Hence, $PoA$ = 1.

## 5.4. Algorithm Design

The TopGame algorithm is a cross-layer approach that encapsulates information taken from the routing, MAC and physical layers. In particular, the PRR and neighbour information are obtained from the routing layer, the transmission power used is acquired from the radio and the MAC is responsible for triggering the game to determine the topology in the case of nodes failing or newly added to the network.

Initially, all nodes start communicating at their maximum transmission power $p_{max}$ = 23. Node $i$ collects the neighbour information, such as current transmission power levels used by its neighbours and their respective PRR. This occurs simultaneously, since the number of neighbours is determined via periodic beacons being broadcast and the PRR obtained by unicasting to random neighbours using a gossip-based protocol [12]. The nodes are also synchronized using beacons with a firefly-based [8] algorithm.

Node $i$ iterates through its 4 available transmission power levels, it computes $TR$ for each power level and it finally maximizes its utility function $u_i$. Note that for practical reasons the pricing of each node's utility function is set to 1. The global optimum is accomplished as we can see in Theorem 5.1. Pseudocode of TopGame is presented in Algorithm 1.

In the case of the addition or a failure of node, nodes that detect a change in their neighbour table initiate TopGame from the start, since their $TR$ will be affected by the topological change. This is due to the fact that TopGame is a repeated game only on topological alterations.

**Algorithm 1** TopGame at node $i$

**Require:** $A_i = \{p_1^i, p_2^i, \ldots, p_{max}^i\}$
**Require:** $degree = 6$, $p_i = p_{max}$
1: **for** $i = 1$ to $N_{i,p}$ **do**
2: $get\ p_{-i}, PRR_{i,k} \in N_i$
3: $N_i \leftarrow N_i^{p_i}$
4: $compute\ TR_i$
5: $u_i(p_i, p_{-i}) = c_i * p_i - TR_i$
6: **end for**
7: $\hat{p}_i = \arg\max u_i(m)$

### 5.4.1. Message Overhead

The message overhead per transmission power consists of the sum of the broadcast messages for synchronization and the unicast messages transmitted to each of the neighbours of every node. That is,

$$O_{TopGame} = f_{sync} * N + N * f_{link} * m * M \quad , \quad (20)$$

where $N$ is the number of nodes, $m$ is the window of the unicast transmission to obtain *PRR* and $M$ is the number of neighbours of every node for a given transmission power.

Since $f_{sync} = \Theta(1)$, $f_{link} = \Theta(N)$ and $M = \Theta(1)$, provided some constants $c_1$, $c_2$, $c_3$ we have: $f_{sync} \leq c_1$, $f_{link} \leq c_2 * N$, $M \leq c_3$. Therefore, 20 can be expressed as follows:

$$O_{TopGame} \leq f_{sync} * N + N * f_{link} * m * M$$
$$\Rightarrow O_{TopGame} \leq c_1 * N + c_2 * N * m * c_3$$
$$\Rightarrow O_{TopGame} \leq N * (c_1 + m * c_2 * c_3)$$
$$\Rightarrow O_{TopGame} = O(N)$$

## 6. Experimental Evaluation and Results

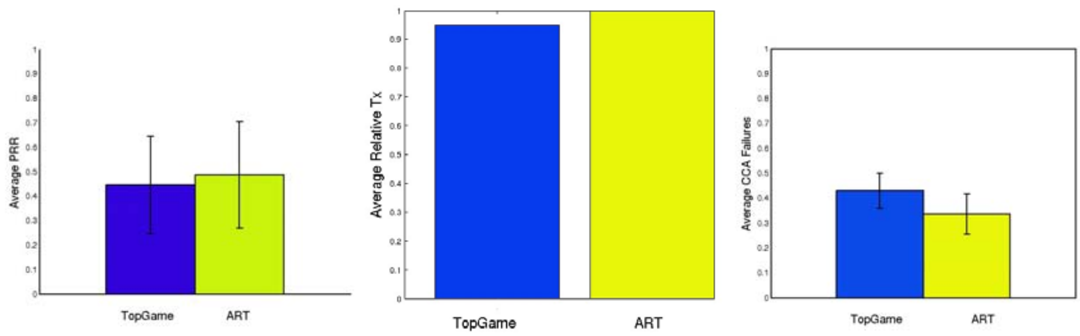
In order to evaluate TopGame, in comparison to ART, we performed 120 minute experiments using 50 nodes selected at random on Indriya. The data rate of the nodes was 250 *kbps* and each node transmits 4 packets per second. In addition each node calculates its PRR over a window of 8 packets. Each node finishes iterating through all the transmission power levels and the utility function of each power level has been obtained in order to proceed with the maximization.

Our aim is to show that converging to lower transmission powers, may provide similar reception performance. In this section we will provide results that show that TopGame approaches ART [19] in terms of PRR and is slightly better in energy-efficiency while ensuring that the network is connected.

### 6.1. Performance and Energy Consumption

Initially, we obtained the average PRR and relative energy consumption, in order to evaluate both algorithms globally. The network average PRR is provided in Fig. 2 (a). Specifically, we observe that TopGame exhibits an average network PRR of 44.7 %, while ART 48.1 %. Moreover, the standard deviation of ART is higher than TopGame's by 3 %.

The difference in the PRR between the two schemes is not quite significant; however, we have shown that a game theoretic algorithm, with a more systematic approach, exhibits similar performance with a state-of-the-art practical algorithm such as ART. Further, the formation of less links does indicate that TopGame uses its utility function that finds the sweet spot per node. On the other hand, ART fluctuates between its two packet failure thresholds; thus, forming more links. As we have seen in a previous section, contention is related to the number of neighbours ($K$) of each node. Table 2 presents the average degree of the network and the number of links that are formed with TopGame and ART.




(a) Average PRR (b) Average Relative Energy Consumption (c) Average CCA Failures Ratio

**Fig. 2.** TopGame and ART average Relative Energy, Mean PRR and CCA failures Mean of 50 nodes.

**Table 2.** Average K and formed links.

| | Average K | Number of Links |
|---|---|---|
| ART | 7 | 360 |
| TopGame | 7 | 338 |

In order to examine whether the difference in the PRR average of TopGame and ART is a result of channel collision we performed 2 hour experiments measuring the Clear Channel Assessment (CCA) failures. Fig. 2 (c) presents the CCA failures ratio of TopGame and ART. Briefly, a CCA operation occurs when the MAC layer receives a packet to transmit, then it instructs the physical layer to check channel availability (CCA) in two consecutive slots. If the channel is found to be available in both slots, the node proceeds with its transmission. Otherwise, the node attempts CCA again after a random back-off, which it repeats a certain number of times and it calls a failure of access to the upper layer. Hence, with TopGame exhibiting nearly 10 % more CCA failures than ART, it is natural to assume that the difference in average PRR comes from a higher interference and collisions of TopGame of the bootstrapping period, since it initially forms a larger number of links that are included in the graph. Specifically, TopGame exhibits 43 % failures, while ART's percentage is 33.6 %.

In our experiments we were unable to directly monitor the energy consumed by the listening and transmitting periods of each node. Thus, we decided to use unicast communications as an indicator to calculate the relative average energy. Making the assumption that all nodes spend the same amount of energy in listening, to get a rough idea of relative energy consumption, we added the number of unicast messages transmitted by ART and TopGame with their respective transmission powers and multiplied them with the corresponding *mA* radio energy consumption. The relative energy consumption of the two algorithms can be seen in Fig. 2 (b). TopGame consumes 5 % less energy than ART, including the bootstrapping period.

This is due to the fact that the nodes do not fluctuate on a per packet basis and they are not targeting a very high PRR value as dictated in the ART thresholds; hence, TopGame is slightly more energy efficient. We present an example of two links from both TopGame and ART, in order to show the difference in the switching between transmission powers and the convergence of TopGame. From the Figs. 3 and 4, it is clear that ART switches its transmission power according to packet drops; hence, the Tx fluctuation in the figure. On the other hand, TopGame collects TR for each available transmission power and converges to the transmission power maximizing the utility function. Also, we are not aware of the energy cost of the continuous Tx switch. We assume it is negligible. Note that TopGame is repeated only when a neighbourhood change is detected.

Recall that ART's intention is to reach the target PRR of 95 %, yet we observe that its reception quality is significantly lower. TopGame also does not attain this lofty figure. We believe that is the case for our scheme because of the bimodal distribution of 802.15.4 link qualities [45].

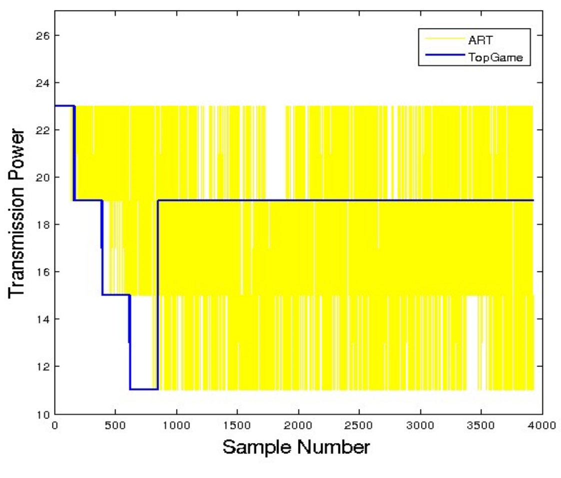



**Fig. 3.** ART and TopGame Node 13 Tx levels.

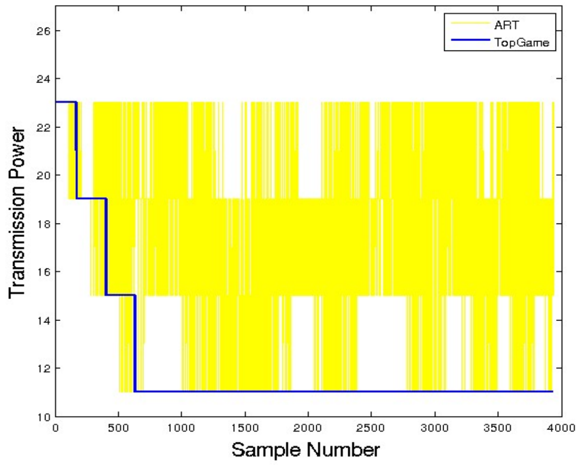



**Fig. 4.** ART and TopGame Node 41 Tx levels.

By looking at the Cumulative Density Functions (CDF) of the two algorithms in Fig. 5, we observe that TopGame has a slightly higher probability of forming poorer quality links of PRR lower than 20 %. ART has a lower probability of forming medium to high quality links. Furthermore, TopGame exhibits a slightly higher probability of establishing links with PRR over 80 %. It would be strong to claim that TopGame is better than ART; however, approaching the numbers of ART is significant, since it relies in concrete theoretical basis.

Finally, we compared the RAM and ROM overhead of TopGame with ART. Table 3 shows that TopGame consumes 2348 more ROM bytes than ART, while it produces an overhead in RAM of 342 bytes.

**Table 3.** RAM and ROM (bytes).

| | RAM | ROM |
|---|---|---|
| TopGame | 3426 | 25228 |
| ART | 3084 | 22880 |

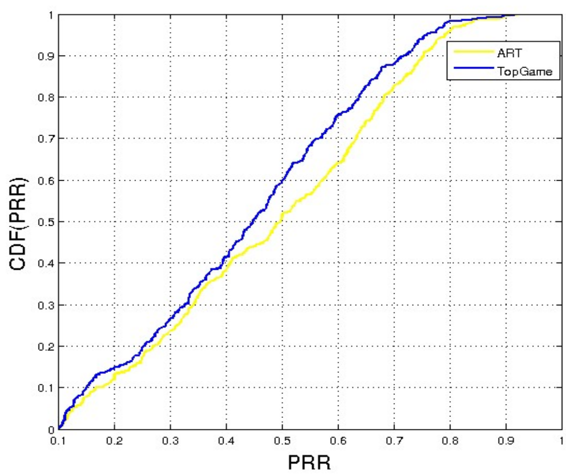



**Fig. 5.** CDF for ART and TopGame.

### 6.2. Connectivity

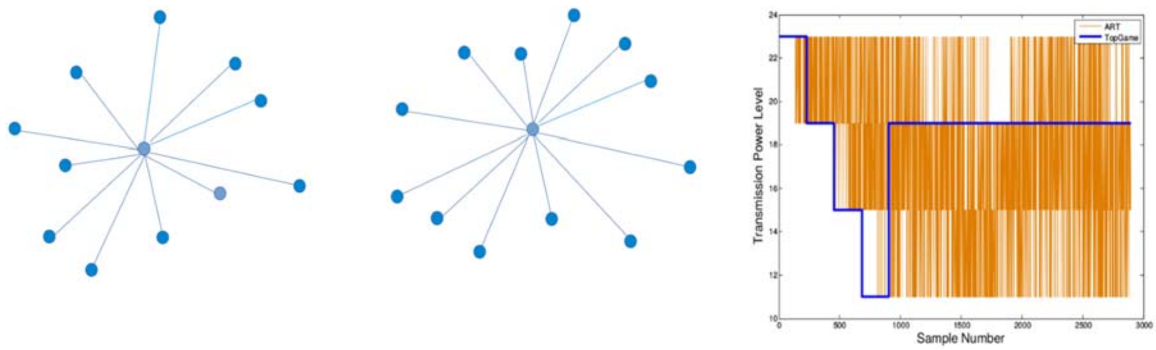
After we obtained the results we evaluated connectivity offline using method that determines whether the resulting graph is connected. This was to evaluate the connectivity model we discussed in a previous section. We have shown that the average degree of each node is greater than 6 nodes. We derived the available links from TopGame and ART's data sets and we created their respective adjacency matrices. Thereafter, we used the matrices to find a zero eigenvalue. In the case that the corresponding eigenvector has 0s, then a sum of non-zero number of rows/columns of the adjacency is 0 [22]. Hence, the degrees of these nodes are 0 and the graph is disconnected. Both TopGame and ART resulted in fully connected graphs.

### 6.3. Self-configuration and Self-healing

In the case of a node addition or failure the two algorithms are expected to behave differently. ART will adjust its transmission power levels depending on the number of failure threshold and the target PRR.

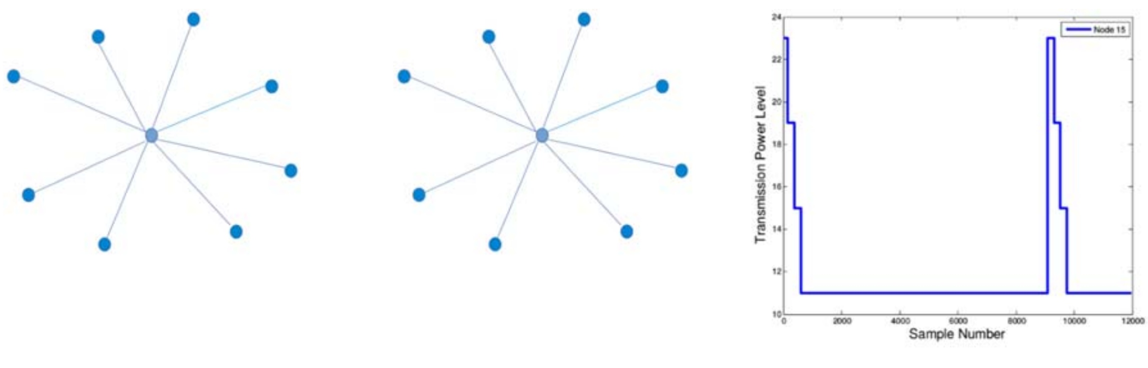
On the other hand, TopGame includes a mechanism, which allows it to become aware of the change in the neighbourhood. Specifically, in the case that TopGame detects a topological change it iterates through the 4 transmission powers again, starting from the highest one. This will trigger the other nodes' initialization period as well, since the transmission ranges of the nodes affected by the change will be informed about the topological change. We simulated a node addition on Indriya by starting the radio of node 18, calling the Tinyos AMControl.start() and AMControl.stop() interfaces respectively after a period of 30 minutes that was measured by a timer. Figs. 6 (a) and 6 (b) show the node degrees of TopGame and ART respectively. TopGame converges with 11 neighbours, while ART has 12. Furthermore, Fig. 6 (c) presents the tx levels of the two algorithms. TopGame after iterating through the 4 tx levels converges to tx level 19. ART, on the other hand fluctuates through the entire set of the transmission power levels, showing the failures it suffers from on a per link basis.

Figs. 7 (a) and (b) show node 15 degree of both algorithms before and after node 18 starts participating in the network. Node 15 detects the topological change in its neighbourhood and start the TopGame algorithm iterating through the available set of transmission power levels as can be seen in Fig. 7 (c). The node degree before and after the topological change remain the same.




(a) TopGame node 18 degree (b) ART node 18 degree (c) Tx levels of node 18

**Fig. 6.** Node 18 degree and Tx levels after it joins the network.




(a) Node 15 degree before node 18 start (b) Node 15 degree after node 18 start (c) Tx levels of node 15

**Fig. 7.** Node 15 degree and Tx levels after 30 min start.

## 7. Further Optimization Discussion

The result presented above include links between nodes that exhibit a very poor PRR, due to their location and/or conditions of the experiments. We characterize these links as bad links and we set the threshold of this characterisation to a PRR of 20%. We denote this threshold as $PRR_{thr}$. Thereafter, we define an indicator function to indicate whether a link participates in the utility maximization process with a poor or satisfactory/good PRR as

$$I(PRR_i) = \begin{cases} 1, & PRR_i \geq PRR_{thr} \\ 0, & \text{otherwise} \end{cases} \tag{21}$$

We are facing the issue of links to participate in the neighbourhood establishment of the utility that significantly affects the maximization process; therefore, the transmission power that it will converge to. Thus, we should have a level of cooperation between links of the neighbourhood, in order to accomplish some level of global optimality. Thus, we formulate our communication participation problem as an optimization problem:

$$\min \sum_{link(i,j),i=1}^{N} I(PRR_i(i,j)) \tag{22}$$

where $N$ is the set of nodes and $link(i, j)$ is the established link between a node $i$ and a node $j$.

The problem above is a non-smooth optimization problem due to the indicator function. However, the indicator function can be 'smoothened' by using the sigmoid function. The sigmoid function is a continuous function and is given below:

$$sig(x) = \frac{1}{1+e^{-a(x-b)}} \tag{23}$$

We can see that when $x$ is larger than the threshold $b$, $sig(x)$ rises to 1. On the other hand, when $x$ is less than the threshold it drops to 0. The parameter $a$ is the factor the influences the behaviour of the sigmoid function near the threshold. To continue, we replace the indicator function with the sigmoid function and we transform the optimization problem as

$$\min f(PRR) = \sum_{conv(i,j),i=1}^{N} sig(PRR_i(i,j)) \quad \text{s.t } PRR_i^{min} \leq PRR_i \leq PRR_i^{max} \tag{24}$$

Our problem is relaxed and can be solved by nonconvex optimization methods [13], [7]. We know that the duality gap in non-convex problems is positive; hence, this type of problems cannot be solved in a distributed manner. However, as identified in [47], a subset of non-convex optimization problems can be identified for which the duality gap is zero. Thus, the distributed gradient-based algorithm is suitable to converge to the optimal solution.

To this end, we define the Lagrangian function as

$$L(PRR,\lambda)= f_i(PRR)-2PRR_i\lambda+PRR^{min}_i\lambda-PRR^{max}_i\lambda \tag{25}$$

The gradient of the Lagrangian function with respect to the variables $PRR_i$ and $\lambda$ are

$$\frac{\partial L(PRR_i,\lambda)}{\partial PRR_i} = f_i'(PRR) - 2\lambda \tag{26}$$

$$\frac{\partial L(PRR_i,\lambda)}{\partial \lambda} = -2PRR_i + PRR_i^{min} - PRR_i^{max} \tag{27}$$

Thus, we have the iterative algorithm equation defined as

$$PRR_i(t+1) = PRR(t) + \beta_{PRR}(t)\frac{\partial(PRR,\lambda)}{\partial PRR_i} \tag{28}$$

$$\lambda_i(t+1) = \lambda(t) + \beta_\lambda(t)\frac{\partial(PRR,\lambda)}{\partial \lambda_i}, \tag{29}$$

where $\beta(t)$ are the small positive constants.

At this point, we have to clarify that we might not be able to locate and reach the optimal solution in this problem. This is due to the fact that the duality gap may be positive. However, by using sigmoidal functions the optimal $PRR_i$ may lie in the concave region of $f(PRR_i)$; hence according to [47] the duality gap will be zero and our algorithm will converge to the optimal primal solution, following the properties regarding the step sizes of $\beta_\lambda$ and $\beta_{PRR}$ of the Gradient Method [5].

However, as we see in other works [25], sigmoidal functions may result in an oscillation of the algorithm between zero and positive values of $PRR_i$, under specific conditions. In cases like these, we employ a heuristic to produce a stable communication between participant nodes. A heuristic is to remove the participant nodes with low PRR, since they are destabilizing the communication party because of their inability to engage.

In the Fig. 8, we can see the average PRR of TopGame as opposed to ART's when the bad PRR links have been minimized. We see that the PRR of TopGame slightly increases. This coincides with the CDF in Fig. 5, where we concluded that we establish bad links with small probability.

## 8. Conclusions

Compared with the state of the art protocol ART, we showed that TopGame is a general solution providing efficient robust Topology Control minimizing the costs of communications while ensuring connectivity.

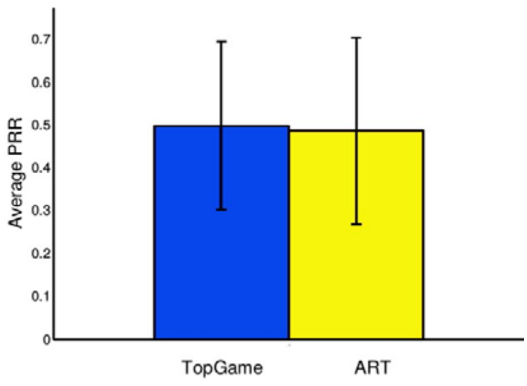



**Fig. 8.** PRR of TopGame and ART after TopGame excludes bad links.

First, we evaluated ART and TopGame on Indriya using 50 nodes to determine the average PRR and the average relative TX power. We evaluated connectivity based on a method that checks the eigenvector of each algorithms adjacency matrix (links) and determined that the resulting graph is connected. The experiments on Indriya showed that TopGame reduces power consumption compared with ART without significantly degrading link quality. Macro-benchmarks comparing TopGame to ART protocol indicated that TopGame provides guaranteed connectivity and exhibited slightly lower PRR than its competitor and 5 % improvement on energy consumption. The corresponding Probability Density Functions showed that TopGame has a slightly higher probability of having links of low quality (< 20 %) than ART. Moreover, ART has a lower probability of creating links of > 20 – 80 % PRR. Finally, TopGame has a better probability of creating high PRR links (> 80%) This is a promising factor of the comparison between the two algorithms in terms of performance, since the average network PRRs were not significantly different. In terms of energy consumption, we presented results that show that TopGame converges to lower transmission power levels than ART making it more energy-efficient. The main differences between ART and TopGame are that ART establishes per-link power levels while TopGame establishes power settings for a given neighbourhood of nodes, and thus can be seen as non link-based. A node running ART will have to switch between transmission powers to transmit packet to its neighbours. This has an impact on the transmission power selection in larger networks, since the target PRR (95 %) is not reached and nodes select high transmission powers. TopGame's power is set to cover the neighbourhood and therefore has no such switching overhead. ART obtains data and makes decisions by indirectly considering link asymmetry in that; hence ART selected higher transmission power levels. In fact, their non optimized version not using the contention gradient verify these phenomena and even the improved version also shows a decrease in PRR. However, link asymmetry is taken into account in TopGame; where bi-directional information helps to ensure both the connectivity of all nodes and that we will converge at a Nash Equilibrium. Finally, in terms of implementation, ART is closely coupled to CTP whereas, though TopGame is slightly more expensive in terms of speed and footprint, it is agnostic to WSN Operating System or stack implementations and is therefore more generally applicable. We aim to interface our approach with CTP or other state of the art routing protocol such as the Backpressure Collection Protocol (BCP) [31].

## References

[1]. Mohammadjavad Abbasi and Norsheila Fisal, Noncooperative game-based energy welfare topology control for wireless sensor networks, *IEEE Sensors Journal*, 15, 4, 2015, pp. 2344–2355.

[2]. N. Ahmed, P. Misra, S. Jha, and D. Ostry, Characterization of link asymmetry in wireless sensor networks, in *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems (ACM SenSys)*, 2009, pp. 373–374.

[3]. Eitan Altman, Anurag Kumar, and Yezekael Hayel, A potential game approach for uplink resource allocation in a multichannel wireless access network, in *Proceedings of the 4th International ICST Conference on Performance Evaluation Methodologies and Tools*, 2009, p. 72.

[4]. C. Antonopoulos, A. Prayati, T. Stoyanova, C. Koulamas, and G. Papadopoulos, Experimental evaluation of a WSN platform power consumption, in in *Proceedings of the IEEE International Symposium on Parallel & Distributed Processing (IPDPS'09)*, 2009, pp. 1–8.

[5]. Dimitri P. Bertsekas, Nonlinear programming, *Athena Scientific*, Belmont, 1999.

[6]. D.M. Blough, M. Leoncini, G. Resta, and P. Santi, Topology control with better radio models: Implications for energy and multi-hop interference. *Performance Evaluation*, 64, 5, pp. 379–398, 2007.

[7]. Paul T Boggs and Jon W Tolle, Sequential quadratic programming, *Acta Numerica*, 4, 1995, pp. 1–51.

[8]. M. Breza and J.A. McCann, Lessons in implementing bio-inspired algorithms on wireless sensor networks, in *Proceedings of the NASA/ESA Conference on Adaptive Hardware and Systems (AHS'08)*, 2008, pp. 271–276.

[9]. M. Burkhart, P. Von Rickenbach, R. Wattenhofer, and A. Zollinger, Does topology control reduce interference?, in *Proceedings of the 5th ACM International Symposium on Mobile Ad Hoc Networking and Computing (ACM MOBIHOC)*, 2004, pp. 9–19.

[10]. Utku Ozan Candogan, Ishai Menache, Asuman Ozdaglar, Pablo Parrilo, et al., Near-optimal power control in wireless networks: A potential game approach, in *Proceedings of the 29th IEEE Conference on Information Communications INFOCOM, 2010,* pp. 1–9.

[11]. Jiming Chen, Qing Yu, Peng Cheng, Youxian Sun, Yanfei Fan, and Xuemin Shen, Game theoretical approach for channel allocation in wireless sensor and actuator networks, *IEEE Transactions on Automatic Control*, 56, 10, 2011, pp. 2332–2344.

[12]. S. M. Dammer and H. Hinrichsen, Epidemic spreading with immunization and mutations, *Physical Review E*, 68, 1, 2003, pp. 016114.

[13]. Roger Fletcher and Michael J. D. Powell, A rapidly convergent descent method for minimization, *The Computer Journal*, 6, 2, 1963, pp. 163–168.
[14]. Yong Fu, Mo Sha, Gregory Hackmann, and Chenyang Lu, Practical control of transmission power for wireless sensor networks, in *Proceedings of the 20th IEEE International Conference on Network Protocols (ICNP'12),* 2012, pp. 1–10.
[15]. D. Ganesan, B. Krishnamachari, A. Woo, D. Culler, D. Estrin, and S. Wicker, Complex behavior at scale: An experimental study of low-power wireless sensor networks, Technical report, *Citeseer*, 2002.
[16]. A. Ganesh and F. Xue, On the connectivity and diameter of small-world networks, *Advances in Applied Probability*, 39, 4, 2007, pp. 853–863.
[17]. Y. Gao, J. C. Hou, and H. Nguyen, Topology control for maintaining network connectivity and maximizing network capacity under the physical model, in *Proceedings of the 27th Conference on Computer Communications (INFOCOM'08)*, 2008, pp. 1013–1021.
[18]. O. Gnawali, R. Fonseca, K. Jamieson, D. Moss, and P. Levis, Collection tree protocol, in *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems (SenSys' 09)*, 2009, pp. 1–14.
[19]. G. Hackmann, O. Chipara, and C. Lu, Robust topology control for indoor wireless sensor networks, in *Proceedings of the 6th ACM Conference on Embedded Networked Sensor Systems (SenSys' 08)*, 2008, pp. 57–70.
[20]. Xiao-Chen Hao, Mei-Qi Wang, Shuang Hou, QianQian Gong, and Bin Liu, Distributed topology control and channel allocation algorithm for energy efficiency in wireless sensor network: From a game perspective, *Wireless Personal Communications*, 80, 4, 2015, pp. 1557– 1577.
[21]. Tiina Heikkinen, A potential game approach to distributed power control and scheduling, *Computer Networks*, 50, 13, 2006, pp. 2295–2311.
[22]. R. A. Horn and C. R. Johnson, Matrix Analysis, *Cambridge University Press*, 2005.
[23]. Eduard Jorswieck and Holger Boche, Majorization and matrix-monotone functions in wireless communications, *Foundations and Trends in Communications and Information Theory*, 3, 6, 2006, pp. 553–701.
[24]. R. S. Komali, A. B. MacKenzie, and R. P. Gilles, Effect of selfish node behavior on efficient topology design, *IEEE Tran. Mobi. Comput.*, 2008, pp. 1057–1070.
[25]. J.-W. Lee, Ravi R. Mazumdar, and Ness B. Shroff, Nonconvex optimization and rate control for multi-class services in the internet, *IEEE/ACM Transactions on Networking*, 13, 4, 2005, pp. 827–840.
[26]. L. Li, J. Y. Halpern, P. Bahl, Y. M. Wang, and R. Wattenhofer, A cone-based distributed topologycontrol algorithm for wireless multi-hop networks, *IEEE/ACM Trans. Netw.*, 13, 1, 2005, pp. 147–159.
[27]. N. Li, J.C. Hou, and L. Sha, Design and analysis of an MST-based topology control algorithm, *IEEE Trans. Wireless Commun.*, 4, 3, 2005, pp. 1195–1206.
[28]. S. Lin, J. Zhang, G. Zhou, L. Gu, J. A. Stankovic, and T. He, ATPC: adaptive transmission power control for wireless sensor networks, in *Proceedings of the 4th ACM Conference on Embedded Networked Sensor Systems (SenSys'09*), 2006, pp. 223– 236.
[29]. Albert W. Marshall, Ingram Olkin, and Barry Arnold, Inequalities: Theory of Majorization and its Applications, *Springer Science & Business Media*, 2010.
[30]. Farhad Meshkati, H. Vincent Poor, Stuart C. Schwartz, and Radu V. Balan, Energy-efficient power and rate control with qos constraints: a game-theoretic approach, in *Proceedings of the International Conference on Wireless Communications and Mobile Computing*, 2006, pp. 1435–1440.
[31]. Scott Moeller, Avinash Sridharan, Bhaskar Krishnamachari, and Omprakash Gnawali. Routing without routes: The backpressure collection protocol, in *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks*, 2010, pp. 279–290.
[32]. D. Monderer and L. S. Shapley, Potential games, *Games and Economic Behavior*, 14, 1996, pp. 124–143.
[33]. Roger B Myerson, Game theory: analysis of conflict, *Harvard University*, 1991.
[34]. A. Nahir, A. Orda, and A. Freund, Topology design and control: A game-theoretic perspective, in *Proceedings of the IEEE INFOCOM Conference*, 2008, pp. 1620–1628.
[35]. J.F. Nash Jr., The bargaining problem, *Econometrica: Journal of the Econometric Society*, 1950, pp. 155–162.
[36]. James O. Neel, Jeffrey H. Reed, Robert P. Gilles, et al., Convergence of cognitive radio networks, in *Proceedings of the Wireless Communications and Networking Conference (WCNC)*, Vol. 4, 2004, pp. 2250–2255.
[37]. Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V. Vazirani, Algorithmic Game Theory, Vol. 1, *Cambridge University Press,* Cambridge, 2007.
[38]. Christos H Papadimitriou, On the complexity of the parity argument and other inefficient proofs of existence, *Journal of Computer and System Sciences*, 48, 3, 1994, pp. 498–532.
[39]. D. Son, B. Krishnamachari, and J. Heidemann, Experimental study of the effects of transmission power control and blacklisting in wireless sensor networks, in *Proceedings of the 1st Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON '04),* 2004, pp. 289–298.
[40]. Dongjin Son, Bhaskar Krishnamachari, and John Heidemann, Experimental study of the effects of transmission power control and blacklisting in wireless sensor networks, in *Proceedings of the 1st Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON '04),* 2004, pp. 289–298.
[41]. Yang Song, Starsky H. Y. Wong, and Kang-Won Lee, A game theoretical approach to gateway selections in multi-domain wireless networks, *Gateways*, 1, 2011, S1.
[42]. Evangelos D. Spyrou and Dimitrios K. Mitrakos, Approximating NASH equilibrium uniqueness of power control in practical WSNs, *International Journal of Computer Networks & Communications (IJCNC)* Vol. 7, No. 6, November 2015.
[43]. Evangelos D. Spyrou and Dimitrios K. Mitrakos, On the homogeneous transmission power under the sinr model, in *Proceedings of the 4th International Conference on Telecommunications and Remote Sensin, (ICTRS'15)*, 2015.
[44]. Evangelos D. Spyrou, Shusen Yang, and Dimitrios K. Mitrakos, Discrete strategy game-theoretic topology control in wireless sensor networks, in *Proceedings of*

*the 6th International Conference on Sensor Networks (SENSORNETS '17)*, 2017.

[45]. K. Srinivasan, M. Kazandjieva, S. Agarwal, and P. Levis, The beta-factor: Improving bimodal wireless networks, in *Proceedings of the 6th ACM Conference on Embedded Networked Sensor Systems (ACM SenSys)*, 2007.

[46]. Qian Tan, Wei An, Yanni Han, Yanwei Liu, Song Ci, Fang-Ming Shao, and Hui Tang, Energy harvesting aware topology control with power adaptation in wireless sensor networks, *Ad Hoc Networks*, 27, 2015, pp. 44–56.

[47]. George Tychogiorgos, Athanasios Gkelias, and Kin K. Leung, A new distributed optimization framework for hybrid ad-hoc networks, in *Proceedings of the IEEE GLOBECOM Workshops (GC Wkshps)*, 2011, pp. 293–297.

[48]. Takashi Ui, Discrete concavity for potential games, *International Game Theory Review*, 10, 01, 2008, pp. 137–143.

[49]. J. Von Neumann, O. Morgenstern, A. Rubinstein, and H.W. Kuhn, Theory of Games and Economic Behavior, *Princeton Univ. Pr.*, 2007.

[50]. M. K. H. Yeung and Y. K. Kwok, A game theoretic approach to power aware wireless data access, *IEEE Transactions on Mobile Computing*, 2006, pp. 1057–1073.

[51]. J. Zhao and R. Govindan, Understanding packet delivery performance in dense wireless sensor networks, in *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems ACM SenSys*, 2003, pp. 1–13.

*Sensors & Transducers, Vol. 212, Issue 5, May 2017, pp. 15-22*

**Sensors & Transducers**

Published by IFSA Publishing, S. L., 2017
http://www.sensorsportal.com

# Ultra-Low Power Sensor System for Disaster Event Detection in Metro Tunnel Systems

**Jonah VINCKE, Scott KEMPF, Niklas SCHNELLE, Clemens HORCH, and Frank SCHÄFER**

Fraunhofer Institute for High-Speed Dynamics, Ernst-Mach-Institut, EMI,
Eckerstrasse 4, Freiburg, Germany
E-mail: jonah.vincke@emi.fraunhofer.de, scott.kempf@emi.fraunhofer.de,
niklas.schnelle@emi.fraunhofer.de, clemens.horch@emi.fraunhofer.de,
frank.schaefer@emi.fraunhofer.de

*Received: 10 April 2017 /Accepted: 29 May 2017 /Published: 31 May 2017*

**Abstract:** In this extended paper (see [1]), the concept for an ultra-low power wireless sensor network (WSN) for underground tunnel systems is presented highlighting the chosen sensors. Its objectives are the detection of emergency events either from natural disasters, such as flooding or fire, or from terrorist attacks using explosives. Earlier works have demonstrated that the power consumption for the communication can be reduced such that the data acquisition (i.e. sensor sub-system) becomes the most significant energy consumer. By using ultra-low power components for the smoke detector, a hydrostatic pressure sensor for water ingress detection and a passive acoustic emission sensor for explosion detection, all considered threats are covered while the energy consumption can be kept very low in relation to the data acquisition. In addition to [1] the sensor system is integrated into a sensor board. The total average power consumption for operating the sensor sub-system is measured to be 35.9 µW for lower and 7.8 µW for upper nodes.

**Keywords:** Ultra-low power, Wireless sensor network, Energy harvesting, Tunnel system, Natural disaster management.

## 1. Introduction

More than half of the planet's population now lives in urban areas. This creates the need for various mass rapid transport systems including metro systems. New vulnerabilities for society that arise due to disaster events, such as terrorist attacks, flooding or fire are increased by a higher population density and current political processes. To address these challenges – in particular for underground metro systems – as part of the bi-national research project SenSE4Metro (Sensor-based Security and Emergency management system for underground Metro systems during disaster events) [2], a concept for a wireless sensor system is introduced that can detect the most significant threats and which provides rescue forces with the relevant and necessary information in the case of an emergency. The particular operation site leads to the requirement for the WSN that each node must be energy autarkic, which in turn necessitates the application of ultra-low power (ULP) components. The focus in this work lies on the needed sensors to achieve these goals and to fulfill the project requirements.

First, the state-of-the-art of wireless sensor networks for different kind of tunnels is presented. After that, an overview of the proposed wireless sensor network is given, taking the special linear topology for a tunnel system into account. Finally, a concept for the needed sensors that can cover all addressed threats is presented that focuses on the power consumption and takes the applicability for a metro tunnel system into account. This includes the integration of the sensor subsystem into the sensor node and the test of the system. At the end, the presented results are discussed and a conclusion is taken.

## 2. State-of-the-Art

The use of wireless sensor networks for tunnel systems has been investigated in several works. Of the systems described, most are designed primarily for road tunnels [3-5] or mine tunnels [6-7]. Of the latter, some focus on the radio transmission in tunnels (e.g. D. Wu and H. Jiang) and others on special protocols designed to increase robustness against underground collapses [8]. Of the former, Ceriotti, |*et al.* present a WSN consisting of 40 nodes to monitor the light conditions of a 260 m long tunnel. Mottola, *et al.* compared the data of a traffic tunnel (here, railroad tunnels are proposed as analogous with the assessed road tunnel) with a WSN for a vineyard to make suggestions for the communication in road tunnels.

For underground rail tunnels, only a few works exist. Wischke, *et al.* [9] discuss the generation side of the energy autarkic nodes by proposing a vibration energy harvesting solution for maintaining the wireless sensor nodes in rail tunnels. Bennett, *et al.* [10] used MICAz boards as wireless sensor nodes to monitor cracks in a 170 m long part of the Prague Metro and in a 115 m long section of the London Metro. Sivaram Cheekiralla [11] used a WSN to monitor the deformation of a train tunnel during construction using 18 nodes using only a star topology.

Raza, *et al.* introduced a combination of an ULP wake-up receiver with model-based sensing to reduce the power consumption of the rail tunnel WSN by Ceriotti, *et al.* [3]. They simulated an increased lifetime of the nodes by a factor of over 2000 due to their method, which demonstrated the influence of a data model to reduce necessary data transmission and the use of a wake-up receiver.

The optimization of wireless communication in tunnel systems has been widely discussed. Especially Raza, *et al.* [12] showed that the power consumption of the data acquisition is of high interest as they reduced the power consumption of the communication drastically using a wake up receiver. Therefore, the focus of this paper lies on the conception of the sensor system that is based on an initial layout and requirement specification for the WSN.

## 3. Wireless Sensor Network

For the purposes of defining the requirements of the WSN, an assessment of past terrorist attacks on underground, tunnel and rail infrastructure was performed [13]. The assessment results highlighted several distinct differences between attacks on underground systems as compared to above-ground networks, with respect to tactics and effectiveness (in terms of casualties). Ultimately, event scenarios were defined based on explosive and arson attacks targeting the trains and tunnels themselves. While the assessment also highlighted the threat of biological/chemical attacks, due to previous studies [14], these were explicitly omitted from the scope of the project.

Adding recent historical flooding in underground networks [Prague 2002, New York 2012] and accidents [Valencia 2006, Moscow 2014] the following events should be detected and the corresponding data acquired by the WSN:

- Train passage (positioning and movement);
- Fire (temperature and smoke presence);
- Explosion (impact peak pressure and specific impulse);
- Flooding (water presence and depth).

The combined data will be applied to determine danger levels and traversability of tunnel segments and to coordinate paths of access (rescue forces) and escape (passengers).

In order to acquire the necessary data above, sensors are positioned both at ground and ceiling level and all along the tunnel segment. It was decided that the implementation would be performed as a parallel linear topology (as shown in Fig. 1) with cable wired master nodes at each metro station that work as a gateway to the control center. The topology provides added robustness via path redundancy, as both nodes can be used to forward messages. Using two parallel strings of nodes the upper nodes can be used to measure smoke and impact pressure using a wind energy harvester as their power source. The lower nodes on the other hand measure water ingress and ambient temperature and they are powered using a piezoelectric vibration energy harvester attached to a rail.

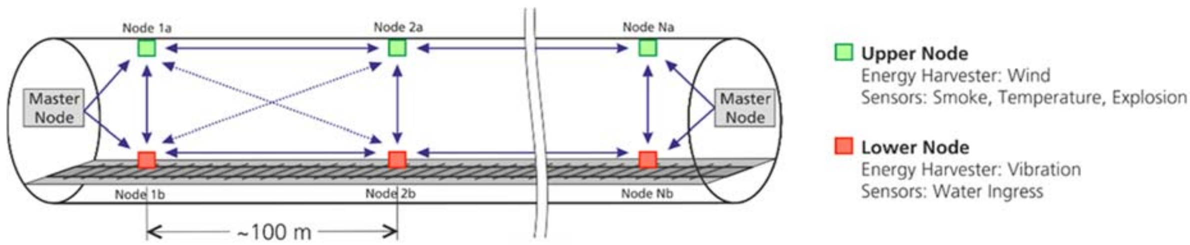



**Fig. 1.** Applied WSN topology in underground tunnel.

In standard (non-emergency) operation, situational status messages are transmitted during train passage events, which means the energy required for data acquisition and transmission is provided directly via energy harvesting processes. For emergency event detection, an energy storage system is provided to ensure a constant power supply.

Its relatively long path lengths however require special tuning and adaptation of routing algorithms. While classic tree based routing protocols, such as Contiki Collect or CTP [15], can in principle be applied directly, nodes at the end of long paths would have to forward all the messages generated by preceding nodes in the path thus creating significant load and using disproportionally more energy. To counter this, modifications are applied to the Contiki Collect protocol. One modification allows data from different nodes to be combined or filtered so as to only forward messages for significant changes instead of simply forwarding all generated data towards the nearest gateway. This also allows nodes to detect when upstream nodes detect an event allowing the network to increase its measurement and forwarding frequencies on demand. Additionally changes are implemented enabling longer paths and passive non-forwarding nodes, which can dynamically switch to a forwarding role when an ongoing situation is detected. This however remains an area of active research and development.

Modifications to the Contiki Collect protocol as well as the uncommonly long path lengths have been evaluated in simulations using the Cooja simulation framework where dozens of nodes can be tested without time intensive reflashing. Additionally tests have been performed on real hardware using SensorTag boards which use the same chip in a convenient package. Development in this area remains ongoing as prototype boards and tunnel testing opportunities become available.

For all nodes, the CC2650 from Texas Instruments is chosen as the MCU due to its very low power consumption and integrated RF module.

## 4. Sensors

To achieve low power consumption, a holistic concept has been developed. This includes the application of modern ultra-low power sensors, enabled only when necessary, as well as the re-application of the same sensors for various disaster events if possible. The precision of the sensors is of less importance in contrast to the power consumption. A robust detection of a dangerous event is sufficient.

### 4.1. Water Ingress

There are several methods for the detection of water ingress and determination of the resulting water level. These vary from mechanical solutions using floats to change a resistance, a capacitance or to close a contact to pure capacitance or resistance measurements as well as hydrostatic, ultrasonic and radar methods. Many can be realized with an ultra-low power consumption but vary with respect to their robustness, dependence on the medium and the tunnel's shape.

Mechanical solutions have the disadvantage that their dimensions need to be in the same range as the measurable water level and that their shapes are limited. On the other hand they are independent from the media and can be realized as ultra-low power systems.

Optical or ultrasound distance sensors are not dependent on such limitations nor on the media or the shape of the tunnel. But they lack on the measurable distance and power consumption. As an example the infrared distance sensor GP2Y0A710 from Sharp needs above 1 mW for one measurement every 5 seconds while only covering a distance of up to 5 m. Ultrasonic sensors that can measure distances of up to 8 m or more have commonly a power consumption of over 1 Watt during operation and need several hundreds of milliseconds until the first measurement is possible. As an example the UC30-2 from SICK needs up to 1.2 W for approximately 450 ms until a measurement can take place. Sensors for smaller distances such as the LV-MaxSonar-EZ have a power consumption of about 10mW for half a second for a measurable distance of 6.45 m.

Of the other solutions, measuring the hydrostatic pressure seems most promising for achieving very low power consumption. Here the pressure caused by the water ingress at the bottom of the tunnel has to be measured as well as that above the water level. The disadvantage of this principle is that calculating the water depth according to the induced pressure difference is dependent on the mediums density by design. Also, the system's robustness in a harsh environment such as an underground tunnel has to be investigated. Since passing trains induce pressure disturbance, the measurement has to be adjusted during train passage events. As an example simulations [16] have shown that a train with a cross-sectional area of 8 m² in a 5 m high tunnel with a speed of 200 km/h creates a pressure difference (in time) of up to 1.36 kPa in the tunnel. If this would be measured as the spatial pressure difference it would be equivalent to a water level of 138.7 mm. How the pressure is disturbed exactly over the cross section and over the time in a real tunnel has to be investigated in further works.

Because of the independence on the tunnel's shape, the water ingress detection will be based on measuring the pressure induced by the water. Since in most cases the media will be ground water the density of the media will be similar in most cases and therefore the dependence of the system on the media can be neglected. The sensors will be located at the wall. The pressure at the bottom of the tunnel is measured using a tube mounted to the wall that is connected to one of the sensors and goes to the ground as shown in Fig. 2. Using two MS5806 pressure sensors, water levels of up to 9 m with a precision of 0.13 cm can be measured while consuming less than 3 µW for each sensor when measuring once per second in theory. A temperature sensor is also included that can be used for the other sensors in addition, reducing the overall power consumption.

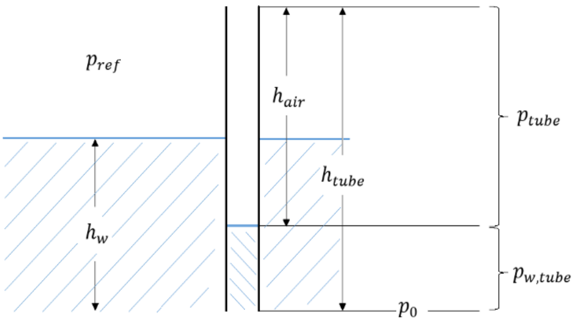



**Fig. 2.** Schematic of the water level measurement system using a tube to measure the pressure at the ground of the tunnel.

The water ingress depth $h_w$ can be determined using the pressure $p_0$ at the base of the tube, the reference pressure $p_{ref}$ above the water and assuming that the media is water:

$$h_w = \frac{p_0 - p_{ref}}{\rho_w \cdot g}, \tag{1}$$

where $\rho_w$ is the water's density and $g$ is the acceleration of gravity. As the air in the tube is compressed by the water ingress within the tube, the measured pressure $p_{tube}$ is lower than the real pressure at the end of the tube. In order to determine the pressure $p_0$, it is necessary to compensate the compression. This can be done using the ideal gas law:

$$p \cdot V = n \cdot R \cdot T, \tag{2}$$

and assuming a constant amount of air n and a constant temperature T. Therefore the product of the pressure and the Volume is constant. Assuming a constant cross-sectional area within the tube the height of the air column can be determined as:

$$h_{air} = h_{tube} \cdot \frac{p_{ref}}{p_{tube}} \tag{3}$$

The pressure at the tube base is a summation of the air pressure due to compression and the water ingress within the tube:

$$p_0 = p_{w,tube} + p_{tube} \tag{4}$$

Reapplying the water depth equation, the total pressure $p_0$ can be determined as:

$$p_0 = \rho_w \cdot g \cdot (h_{tube} - h_{air}) + p_{tube} \tag{5}$$

With (3):

$$p_0 = \rho_w \cdot g \cdot h_{tube}\left(1 - \frac{p_{ref}}{p_{tube}}\right) + p_{tube} \tag{6}$$

Finally, inserting back into (1):

$$h_w = h_{tube}\left(1 - \frac{p_{ref}}{p_{tube}}\right) + \frac{p_{tube} - p_{ref}}{\rho_w \cdot g} \tag{7}$$

This results in an air compression compensated formula to determine the water depth with respect to the measured pressure in the tube and above the water level.

### 4.2. Fire

To detect fires, the temperature and the amount of smoke are measured. Heat is measured using the integrated sensor of the pressure sensor. For the smoke detection three classical systems exists. While measuring the concentration of carbon monoxide either consumes too much power or is limited in life time, ionic sensors can reach a power consumption as low as 25 µW [17] but consist of a radioisotope. Common photoelectric smoke detectors consume approximately 90 µW. This power can be reduced down to less than 6 µW by using ultra-low power microcontroller units (MCUs) and operational amplifiers and by reducing the sampling ratio down to one sample each 8 s [18]. Based on this and because of the German regularities regarding ionic materials, the smoke detectors are realized based on the photoelectric effect.

As shown in Fig. 3, an infrared LED emits light in a smoke chamber that has to be reflected by particles such that the reflection can be measured by a photodiode. To increase the robustness of the sensor against disturbances such as ambient light, the sensor measures the output of the photodiode when the IRLED is turned off additionally. As in [18], a measurement is done every 8 seconds. If smoke is detected, the interval is reduced down to 4 seconds and after another detection to 2 seconds. After three detections an alarm is triggered. This is done to reduce false alarms, while still keeping the response time lower than 22 seconds in the worst case.


**Fig. 3.** Schematic of a photoelectric smoke detector.

### 4.3. Explosion

A typical blast wave as shown in Fig. 4 consists of a leading shock wave and positive pressure phase followed by a negative pressure phase. To estimate its effect on the tunnel's structure and the potential harm on train passengers the most important information is its peak pressure and its impulse of the positive pressure phase.

To measure the blast wave, an acoustic emission sensor, the VS150-M, has been chosen which has been tested successfully in previous projects performed by researchers at the EMI [19-20]. It generates an electrical signal from the deformation of the sensor using a piezoelectric element and therefore needs no supply power. Nevertheless, to measure its signal, an electric subsystem is needed because it is damped if the ADC of the CC2650 is used directly. As in previous work, the most power saving system [21] is to measure the exceedance of several thresholds instead of amplifying the signal. For this purpose three MOSFETs are used that will consume up to 9 µW during normal operation. As opposed to the water ingress and fire detection sensors, the signal of the VS150-M can be used to wake up the MCU in case of an explosion. This allows measuring the start time of the event, the exceedance of different thresholds and the duration of the blast. As it is not important for the system to be sustainable during a disaster event, as it will have enough stored power, the power consumption after the wake up can be neglected.

Given the time how long the thresholds are exceeded, the peak pressure and impulse of the blast wave can be approximated. Using several nodes this is enough to estimate the severity of the blast as well as the energy released.

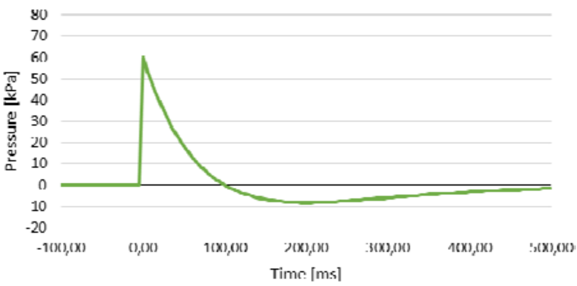



**Fig. 4.** Example of the Friedlander waveform of a blast wave.

Due to the physical sensor node layout and the limitations of ultra-low power WSNs with regard to time resolution and synchronization, an exact determination of explosion position is not feasible. Using empirically determined thresholds based on the results of in-house explosion experiments at the EMI [19], [22-23] the sensor system can however determine the remaining structural capacity of tunnel walls based on the peak pressure and specific impulse observed at the node. The expected variable impact distance of 0-50 meters (based on 100 m node spacing) can be taken into account when establishing these thresholds.

## 5. Sensor Nodes

The sensor nodes consists of 3 boards, one for the communication, one for the sensors and one for the energy supply. The communication board consists of the TI CC2650 EM board and an adapter board for the node's stack (as shown in Fig. 5). The microcontroller has an integrated radio module and is used to control all other boards. The EM board integrates a PCB antenna and has the option to mount an external antenna using a SMA connector. It has a sensitivity of -97 dB and a maximum transmission power of +5 dBm. On outdoor range tests distances of up to 200 m where possible without packet losses.

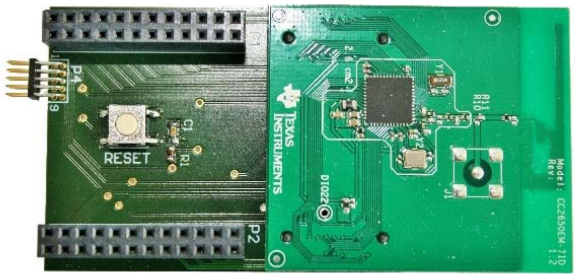



**Fig. 5.** CC2650 EM Board and Stack Adapter.

The sensor board (see Fig. 6) integrates all three sensor types. The explosion detection circuit is shown on the left side. It consists of the three MOSFETs in combination with an adjustable voltage divider for the threshold exceedance detection and an operational amplifier for the direct signal measurement.

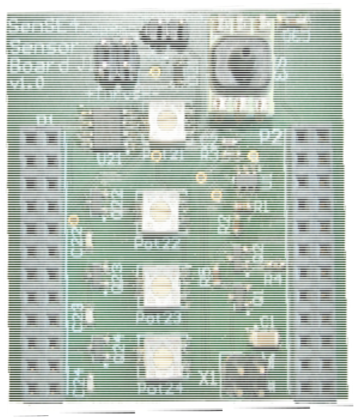


**Fig. 6.** Sensor Board.

The smoke detection circuit is shown on the right. It consists of a constant current source and a current to voltage conversion circuit. The IR LED and photo diode as well as the ultrasound sensor are located externally. The pressure sensor is shown at the upper right corner and is connected to the I2C interface of the microcontroller.

The energy harvesting board is under development.

## 6. Experiments

To measure the different sensor subsystems power consumption, the Agilent B2901A has been used as the power source and measure unit. In all cases the power consumption of the whole system has been measured. The power consumption of the single parts is then derived as the difference between the power consumption of the system when the subsystem under test is not powered by the microcontroller and when it is powered and used for data acquisition. This is done according to parasitic effects that occurs for example when the pressure sensor if powered directly but connected via I²C to the sensor board. It is also necessary to measure the subsystems integrated in the sensor board.

For the water ingress detection system a single pressure sensor increases the power consumption by 14.79 µW when no measurement is done and 40.08 µW if the pressure and temperature is measured once per second. Therefore the power consumption of the data acquisition is 25.29 µW which could be reduced to 3.16 µW if the pressure is measured only once every 8 seconds. This would result in a total power consumption of 17.95 µW for the water ingress detection sensor system.

The power consumption of the smoke sensor varies between 4.06 µW for the breadboard design as shown in [1] and 7.75 µW for the sensor board. In both cases one measurement has been taken every 8 seconds. This time interval is chosen as an acceptable response time.

The power consumption of the explosion detection system is measured to be 0.05 µW during normal operation. As the power consumption of the explosion detection system after a detection is not of importance it has not been measured.

To validate the sensor system different functional tests has been done. In Fig. 7 a depth measurement in a water pipe filled with tap water is shown. The hydrostatic pressure at the bottom is measured using a hose connected to a MS5806 pressure sensor located above the pipe.

The water depth is calculated according to Equation (7) using the measured pressure and a previous measured reference pressure. The compression of the air is compensated for the 645 mm long hose. Here the mean error between the nominal and measured depth increases from 1.2 mm for a real depth of 0 mm up to -14.6 mm for a real depth of 500 mm. This corresponds to a relative error of up to 4.32 % as shown in Fig. 8.

To validate the smoke detection system its response to the application of a test spray is shown in Fig. 9. Here a measurement is done every 0.25 seconds in contrast to normal operation. The response of the smoke detector when the IRLED is turned on is shown in red. This can be compared with the response when the IRLED is turned off as shown in orange. The induced difference between both measurements is increased from an average difference of 203.77 mV before the test spray is applied to 758 mV for the peak difference after the application.

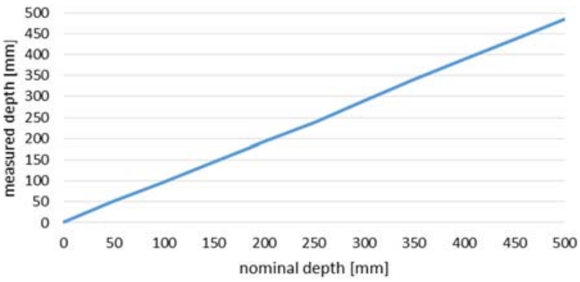



**Fig. 7.** Calculated depth from pressure readings.

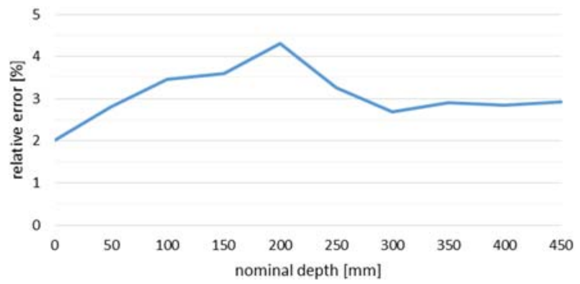



**Fig. 8.** Relative difference between the nominal and measured depth with respect to the nominal depth.

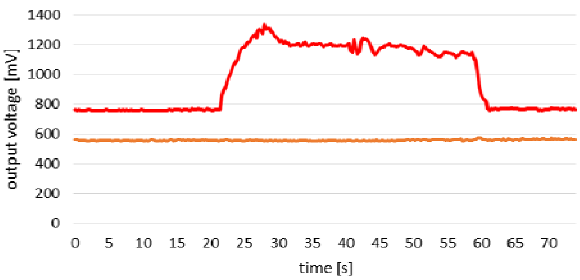



**Fig. 9.** Response of the smoke sensor when using a test spray as a function of voltage [mV] over time [s]. Red: IR LED turned on, Orange: IR LED turned off.

The explosion detection system has been validated in previous work. The characterization of the actual system and its adaptation to a metro tunnel is ongoing.

## 7. Discussion

The power consumption of most subsystems complies with the previous assumptions. Only the pressure sensor increases the power consumption higher as predicted. The power needed for taking one measurement per second of 25.29 µW can be also caused by the activation of the I²C interface of the microcontroller. In contrast to that even the needed

14.79 μW when no measurement is done is about five times higher than the power consumption given by the sensors datasheet. A final explanation cannot be found. Nevertheless the output of the senor is very good. Its response is very linear and the accuracy of the derived water depth is sufficient for the rescue forces to get a situational awareness as its relative error is always lower than 4.32 %. The functional test is not sufficient to draw a conclusion, what the reason is for the more or less constant relative error. To exclude a gain error of the sensor itself a measurement of specific pressures would be necessary. More likely the several assumption for the compression compensated formula to calculate the depth will be the reason. Nevertheless, this error is acceptable to create a situational awareness.

The power consumption of the smoke sensor varies between 4.06 μW for the breadboard design and 7.75 μW for the integrated system. This difference can be caused by losses induced by the printed circuit board track and by slightly different components used for the PCB design. All in all the power consumption lies in the range expected with respect to [18]. The response of the smoke detector to the application of a test spray caused an increase of the output difference of a factor of up to 3.7. This verified the general functionality of the sensor. A more precise application of a specified amount of smoke particles per volume would be necessary to characterize the smoke detector.

The power consumption of 0.05 μW for the explosion detection circuit is less than the power consumption that can be measured reasonably with the given test setup. Therefore it can't be seen as a correct measurement. But it shows that the power consumption will be less than the predicted one according to the maximum ratings of the datasheet of the MOSFETs. The functional test and calibration of the integrated explosion detection system remains an area of active research and development.

## 8. Conclusions

A concept for a wireless sensor network for monitoring underground metro systems, specifically focusing on the requirements and design of the sensors themselves, has been presented. As energy autonomy is desired, low energy consumption of the components, while maintaining the minimum sensing integrity and resolution, is of highest priority.

Using a highly integrated MEMS pressure sensor, ULP components for a smoke detector with a reduced sampling rate and an acoustic sensor for the explosion detection, the expected average power consumption for the sensor system in normal operation can be reduced to 7.8 μW for upper nodes and 35.9 μW for lower nodes. In this case, smoke and water depth are measured every eight seconds. The accuracy of the depth measurement is capable for giving the rescue forces a situational awareness. The smoke detector is very sensitive as its output is increased by a factor of more than three. Therefore, the combination of both nodes provides the capability of detecting explosion, fire and water ingress, while only consuming very low power.

In further work, a hose ending for the pressure sensor has to be developed that ensures robustness of the system in such a harsh environment like a metro. In addition, the pressure disturbances in the tunnel systems induced by passing trains has to be analyzed.

Using the presented sensor system, a larger security management and emergency response system will be developed. All interested parties, such as metro network operators and rescue forces, will be informed in real-time of critical developments. This includes the degradation of tunnel structural integrity and impairment of traversability of tunnel segments in addition to the direct sensor readings. This will help to minimize the secondary damage (e.g. in terms of human life) of disaster events in metro tunnel systems.

## Acknowledgements

This work is part of the research project "Sensor-based Security and Emergency management system for underground Metro system during disaster events" (SenSE4Metro) and is funded by BMBF (German Federal Ministry of Education and Research) through the joint program "International cooperation in civil security research: Cooperation between Germany and India" under Grant No. 13N13039.

## References

[1]. J. Vincke, S. Kempf, N. Schnelle, C. Horch, F. Schäfer, A Concept for an Ultra-low Power Sensor Network - Detecting and Monitoring Disaster Events in Underground Metro Systems, in *Proceedings of the 6th International Conference on Sensor Networks*, 2017, pp. 150-155.

[2]. SenSE4Metro. [Online]. Available: http://www.sense4metro.org/ [Accessed: 04-04-17].

[3]. M. Ceriotti, *et al.*, Is there light at the ends of the tunnel? Wireless sensor networks for adaptive lighting in road tunnels, in *Proceedings of the 10th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, 2011, pp. 187-198.

[4]. L. Mottola, G. P. Picco, M. Ceriotti, Ş. Gună, A. L. Murphy, Not All Wireless Sensor Networks Are Created Equal: A Comparative Study on Tunnels, *ACM Transactions on Sensor Networks*, Vol. 7, No. 2, Sep. 2010, pp. 15:1–15:33.

[5]. Z. Sun, I. F. Akyildiz, Channel Modeling of Wireless Networks in Tunnels, in *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*, New Orleans, LA, USA, 2008, pp. 1-5.

[6]. D. Wu, L. Bao, R. Li, A holistic approach to wireless sensor network routing in underground tunnel environments, *Computer Communications Archive*, Vol. 33, No. 13, Aug. 2010, pp. 1566-1573.

[7]. H. Jiang, L. Chen, J. Wu, S. Chen, H. Leung, A Reliable and High-Bandwidth Multihop Wireless Sensor Network for Mine Tunnel Monitoring, *IEEE Sensors Journal*, Vol. 9, No. 11, Nov. 2009, pp. 1511-1517.

[8]. H. Li, Z. Maohua, S. Congling, S. Jiehong, C. Haicheng, X. Qiaoxiang, Experimental Research on Investigation of Metro Passenger Evacuation Behaviors in Case of Emergency, in *Pedestrian and Evacuation Dynamics*, *Springer,* US, 2011, pp. 173-184.

[9]. M. Wischke, M. Masur, M. Kröner, P. Woias, Vibration harvesting in traffic tunnels to power wireless sensor nodes, *Smart Materials and Structures*, Vol. 20, No. 8, 2011, p. 085014.

[10]. P. J. Bennett, Y. Kobayashi, K. Soga, P. Wright, Wireless sensor network for monitoring transport tunnels, in *Proceedings of the Institution of Civil Engineers - Geotechnical Engineering*, Vol. 163, No. 3, Jun. 2010, pp. 147-156.

[11]. S. Cheekiralla, Wireless sensor network-based tunnel monitoring, in *Proceedings of the REALWSN Workshop*, 2005.

[12]. U. Raza, A. Bogliolo, V. Freschi, E. Lattanzi, A. L. Murphy, A Two-Prong Approach to Energy-Efficient WSNs: Wake-Up Receivers plus Dedicated, Model-Based Sensing, *Ad Hoc Networks*, Vol. 45, 2016, pp. 1-12.

[13]. S. Kempf, F. Schäfer, G. Sitharam, *et al.*, SenSE4Metro: A bi-national multi-disciplinary project for monitoring underground metro environments in disaster events, in *Proceedings of the 11th Future Security Research Conference*, 2016, pp. 479-482.

[14]. Pflitsch Andreas, Brüne Markus, Ringeis Julia, Killing-Heinze Michael, "OrGaMIR"-Development of a safety system for reaction to an event with emission of hazardous airborne substances-like a terrorist attack or fire-based on subway climatology, in *Proceedings of the 4th International Symposium on Tunnel Safety and Security*, Frankfurt am Main, Germany, 2010, pp. 451-462.

[15]. O. Gnawali, R. Fonseca, K. Jamieson, D. Moss, P. Levis, Collection Tree Protocol, in *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, New York, NY, USA, 2009, pp. 1-14.

[16]. ThermoTun Online - Rail Tunnel Ventilation and Aerodynamics Software. [Online]. Available: http://www.thermotun-online.com/index.php [Accessed: 09-Jun-2016].

[17]. Z. Mokhtari, S. Holé, J. Lewiner, Study of an ionic smoke sensor, in *Measurement Science and Technology*, Vol. 24, No. 5, 2013, p. 055006.

[18]. M. Mitchell, Implementing a Smoke Detector with the MSP430F2012, *Texas Instruments Incorporated*, Texas, Application Report SLAA335, October 2006, pp. 1-7.

[19]. O. Millon, F. Schäfer, A. Stolz, W. Riedel, Automatisierte Informationsgewinnung und Schutz kritischer Infrastruktur im Katastrophenfall (AISIS), *Fraunhofer IRB Verlag Stuttgart*, 2013.

[20]. M. Erd, F. Schäfer, M. Kostic, L. M. Reindl, Event monitoring in emergency scenarios using energy efficient wireless sensor nodes for the disaster information management, in *International Journal of Disaster Risk Reduction*, Vol. 16, Jun. 2016, pp. 33-42.

[21]. A. Kleemann, O. Millon, C. Horch, Elasstic - D3.7 - Report of Experimental Study Results, *Fraunhofer EMI*, Freiburg, Germany, Deliverable Report D3.7, 2016.

[22]. A. Stolz, W. Riedel, C. Mayrhofer, M. Nöldgen, K. Dörendahl, Tunnel Structures subjected to Explosions, in *Proceedings of the 5th Security Research Conference,* Berlin, Vol. S.2, 2010.

[23]. F. Schäfer, A. Stolz, L. Reindl, *et al*., Energieautarkes Funksensornetzwerk zur Erhöhung der Sicherheit in Bahntunnels bei terroristischen Anschlägen mit Sprengstoffen, in *Science 2*, No. 1, 2014, pp. 19-29.


---

*Sensors & Transducers, Vol. 212, Issue 5, May 2017, pp. 23-30*

**Sensors & Transducers**

Published by IFSA Publishing, S. L., 2017

http://www.sensorsportal.com

# Performance Analysis of Synthetic Mobility Models and Mobile Ad Hoc Routing Protocols

**[1] Nisrine IBADAH, [1] Khalid MINAOUI, [1] Mohammed RZIZA and [1, 2] Mohammed OUMSIS**

[1] LRIT, Associated Unit to CNRST (URAC 29), Faculty of sciences, Mohammed V University, Rabat, Morocco

[2] Superior School of Technology, Mohammed V University, Salé, Morocco

E-mail: nisrine.ibadah@gmail.com, khalid.minaoui@fsr.ac.ma, rziza@fsr.ac.ma, oumsis@yahoo.com

*Received: 10 April 2017 /Accepted: 29 May 2017 /Published: 31 May 2017*

**Abstract:** Routing protocols heavily influenced by the node motion applied. Many performance analyses are already done with a lot of flaws. But, they do not look to all influenced constraints. Sometimes, they evaluate routing protocols without taken into consideration mobility models. They often analyze them using one routing protocol. Whereas, Simulation time employed is too short. It mainly impacts performance metrics of many mobility models. Or usually, simulation area used is small. It influences the number of packets received. In this study, we aim to summarize all these several parameters into 90 different scenarios with an average of 1350 simulated files. We will combine some well-known mobility models with the most prominent mobile Ad hoc routing protocols in order to analyze their accurate behaviors in one experimental synthesis paper. That shows results of three performance metrics combined with five mobile ad hoc routing protocols under three synthetic mobility models. All these parameters are applied to two dissimilar simulation areas, a small one with (220 m×220 m) and a large one with (1020 m×1020 m). Basing on one exhaustive analysis with all these details like this paper; leads to well understand the accurate behaviors of routing protocols and mobility models used. By displaying the ability of every routing protocol to deal with some topology changes, as well as to ensure network performances.

**Keywords:** MANET, Routing protocols, Mobility models, NS2, Boonmotion, Performance analysis.

## 1. Introduction

For almost two decades, mobile communication has become a major field of research and scientific discoveries. Mobile Ad hoc Network (MANET) has achieved a huge improvement due to its flexibility, easier maintenance, the non-existence of centralized control or fixed and static infrastructure as well as self-administration and self-configuration abilities. Therefore, MANET [1] has become an integral portion of the mobile wireless Network. This kind of wireless network can be established anytime and anywhere, with two or more mobile nodes. If they are in the same radio range, they are directly connected one another. So, they must play the roles of both routers and hosts.

Several mobility models have been proposed to overcome these situations with the aim of imitating human beings' real-life. Wireless communications display many problems related to nodes density, traffic load, autonomous energy, and mobility. Routing within this network suffers from frequent topology's updates and unconnected actives routes between mobile nodes. The main challenge of MANETs routing is to develop a dynamic routing protocol expeditiously able to find a route between mobile nodes. The choice of a mobility model (MM) can favorite some designs over others. It must be efficiently readapted to every change occurring in the network topology [2]. The performance of mobile ad hoc networks can vary significantly under different mobility models. Sometimes, they evaluate routing protocols without taken into consideration mobility models. They often analyze them using one routing protocol. Whereas, Simulation time employed is too short. It mainly impacts performance metrics of many mobility models. Or usually, simulation area used is small. It influences the number of packets received. Their optimal implementation requires a deep study of the routing protocols. Researchers find meaningful to explore mobility model decisions and metrics in modeling their wireless communication where mobile nodes move from a place to another with no fixed infrastructure.

Synthetic Mobility models [3] imitate the movement of real mobile nodes that change speed, position, and direction with time. They can be done by making prevision, mobiles move from one place to another at a given moment under varied network restrictions. They represent precisely motion characteristics of mobile nodes. They are amongst key parameters that influence performance features of the mobile network in order to judge which protocol is useful in a special scenario. Nodes' mobility needs to be analyzed to explore dependency and topology requirements.

This paper will propose an intensive performance analysis of some synthetic mobility model under a mobile ad hoc network. In order to describe mobility issue of various wireless communication scenarios that heavily impacts mobile routing protocols. The entire document is divided into three principal sections. Firstly, we present briefly related works where are used in the simulation. Secondly, we present the parameters of simulation; and also, we interpret the simulation results. And finally, we discuss the conclusion.

## 2. Related Works

Many ways are proposed to classify synthetic mobility models [4].

Firstly, the 'Entity mobility model' where every node is independent of each other. This class has been classified into the following areas: random mobility models, models with temporal dependency, models with spatial dependency and models with geographic restrictions. For random mobility models, nodes travel freely and without obstructions. Direction, speed, and destination are selected randomly and independently of prior selection. That assesses these models to be generally without a memory, e.g.: Random Waypoint Mobility Model (RWMM) [5]. Secondly, models with temporal dependency, devices are governed by motion's physical laws when its present movement depends on its movement's history, for example, Gauss-Markov mobility model [6]. Whereas, patterns with spatial dependency. On many cases, it has been observed that node's waypoint (destination) is probabilistically subordinate its current location, e.g., Probabilistic Random Walk Mobility Model [7]. However, models with geographic restriction, node's movements are not often random or have a temporal/spatial dependency. But, it can be obstructed in a bounded area, guided by paths or restricted into a building, e.g., Manhattan Grid Mobility Model (MGMM) [8]. All previous patterns are considered as 'entity synthetic MM'. Secondly the 'correlated or group based mobility model', where the device node's movement is dependent on others. In this subclass, nodes move by following a leader node in the group. That is to say, each group is governed by one leader which can be a pre-defined or a logical node, e.g., Reference Point Group Mobility Model (RPGMMM) [9].

Thirdly, the 'human or social based mobility model' where nodes are driven by socializing human behaviors, e.g., Self-Similar Least Action Walk [10].

And fourthly, vehicular mobility models emulate vehicle movement with changing speed, moving in queues along highway/street and stopping at traffic signals [11]. That follows the shortest trajectory from a given source to a destination.

However, vehicular communication becomes an important portion of the intelligent transport system.

## 3. Simulation Parameters and Results

### 3.1. Simulation Models' Description

Different scenarios have been considered in order to evaluate mobility nodes and traffic load [12-14]. Both of routing protocols and mobility models are impacted by various criterions during simulation. For instance, we can mention: the 'Traffic Generation Model' [15] which is used to investigate systematically traffic load effect. Many application traffics can be generated in such wireless communication. In our case, we use a random traffic load as Continuous Bit Rate (CBR) between mobile nodes.

And also, they are impacted by the 'Radio propagation model' [16] adopted. It predicts propagation features like received signal power of every packet, antenna features and distance of covered zone applied. At the physical layer, there has a receiving threshold for each mobile node. If, its signal power does not below to receiving threshold that leads to being dropped by the MAC layer. Mainly, there are three propagation models which are Free Space model, Two-Ray Ground reflection model, and Shadowing model. According to Free Space model. It considers propagation conditions as ideal. Where between the transmitter and receiver is direct with only one clear line-of-sight path. Although, Shadowing model is more realistic. At a specific distance, the received power is a random variable caused by multipath propagation effects. In our simulation, we use ’Two-Ray Ground reflection model’ that considers a ground reflection path in addition to the direct path.

We find also, ’Mobility Generation Model’ which is used to explore nodes mobility effect of total network performances. Movement scenario files used for each simulation are characterized by the pause time. If this latter equals 100 seconds there will be almost no movement. However, if it equals zero second that corresponds to continuous motion without stopping.

To study the effect of mobility, a set of movement scenarios correspond to different mobility strategies are generated by Boonmotion Tool [17] in our simulation. There exist many designed tools to generate mobility traces [4].

### 3.2. Configuration Parameters

This paper shows results of three performance metrics which are Packet Delivery Ratio (PDR), average end-to-end delay and throughput under different scenarios.

We combine five mobile ad hoc routing protocols which two of them are proactive, two are reactive and hybrid one. With three synthetic mobility models which are: RWMM is a random entity synthetic MM, MGMM is an entity synthetic MM with restriction geographic MM and RPGMM is group based MM. All these parameters are applied under two simulation areas; a small one with (220 m×220 m) and a large one with (1020 m×1020 m). So, our results will represent 90 different scenarios with an average of 1350 simulated files. We combine all these details in order to well understand the accurate behaviors of routing protocols and mobility models used. Simulation settings used for experiments are depicted in Table 1.

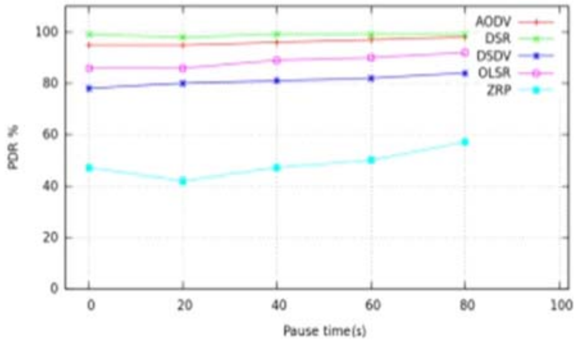
**Table 1.** Simulation parameters.

| Parameters | Values |
|---|---|
| Propagation model | TwoRawGround model |
| Bandwidth | 10 Mb/s |
| Number of nodes | 50 |
| Packet size | 512 bytes/s |
| Traffic sources | CBR |
| Pause time (s) | 0, 20, 40, 60, 80 |
| Routing protocols | DSDV, OLSR, AODV, DSR, and ZRP |
| Mobility models | RWMM, MGMM and RPGMM |
| Performance metrics | PDR, Average e-e delay and Throughput |
| Area | 220 × 220, 1020 × 1020 |
| Simulation time | 1000 s |
| Recursion | 15 times |

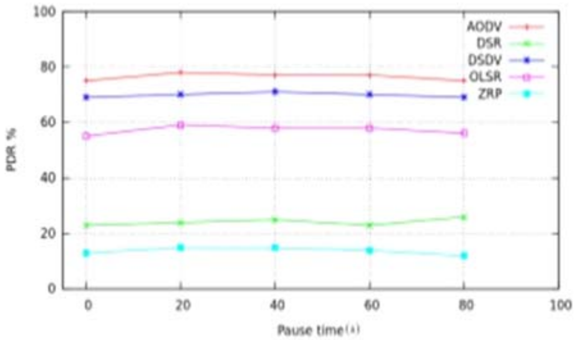
### 3.3. Configuration Parameters

To evaluate routing protocols, a wide range of performance metrics have been considered to catch characteristics of different mobility models. Our results aim to analyze their performance impacts on routing protocols over MANET [18]. So, different metrics have been used to compare and evaluate them against nodes’ mobility, as follows.

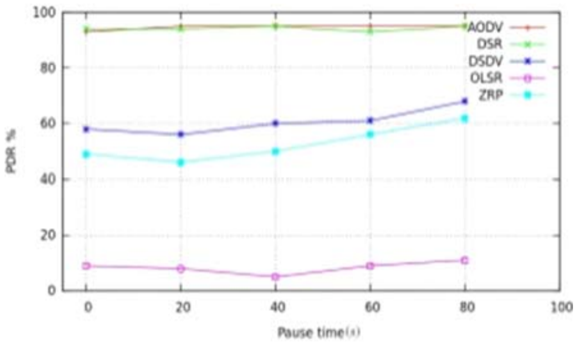
Firstly, we start with Packet Delivery Ratio (PDR) or Fraction (PDF). It represents the ratio of data packets delivered to destinations, those generated by CBR application sources. According to this metric, simulation results are shown in Fig. 1 and Fig. 2.




(a) RWMM



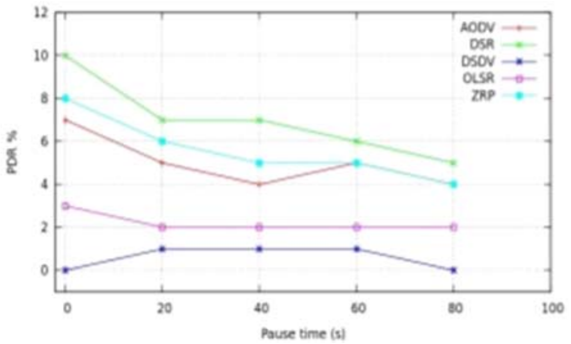

(b) MGMM



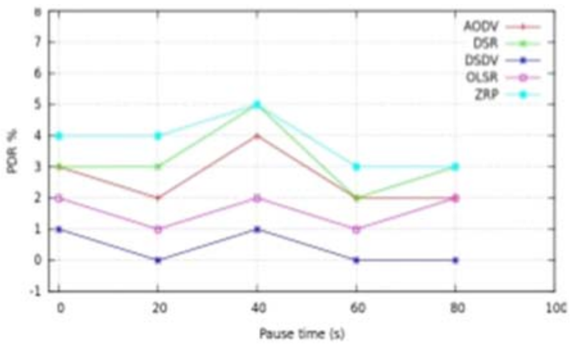

(c) RPGMM

**Fig. 1.** PDR of routing protocols under various mobility models - Small area. (a) Random Waypoint Mobility Model, (b) Manhattan Grid Mobility Model, (c) Reference Point Group Mobility Model.

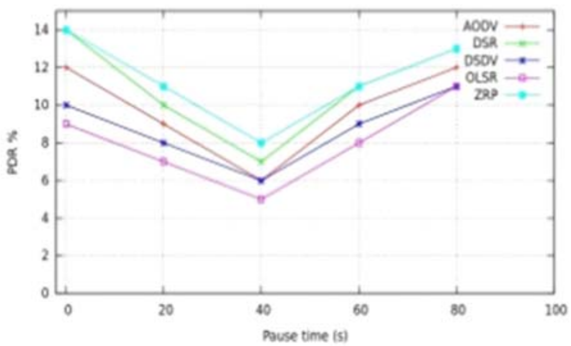
Fig. 1 is applied in the small area. From Fig. 1(a) and Fig. 1(c), the PDR of AODV and DSR present best results in both RWMM and RPGMM in which they reach approximately 100 %. Due to their reactive strategy, routes are sure which are searched on demand. But, AODV represents the best routing protocol in MGMM of Fig. 1(b). However, in RWMM and MGMM, ZRP gives the worst results in this metric, by dint of zone network used by this protocol. DSDV and OLSR in RWMM and MGMM offer acceptable outcomes, thanks to continuously update their routing table. OLSR is the worst in RPGMM.




(a) RWMM




(b) MGMM




(c) RPGMM

**Fig. 2.** PDR of routing protocols under various mobility models - Large area. (a) Random Waypoint Mobility Model, (b) Manhattan Grid Mobility Model, (c) Reference Point Group Mobility Model.

As a result of, OLSR is based on routing by cluster heads. And, RPGMM has their own groups' leader.

So, the same strategy applied for routing and mobility respectively. The coordination between clusters and leader nodes is tough in this case. In general, we notice that AODV offers best results at the PDR for all mobility used in the small area.

Fig. 2 is applied in the large area. From Fig. 2(a), (b) and (c), the DSR and ZRP offer the best PDR percentage. Due to the hidden routing table of DSR which often has an available route to the destination even in a wide field. And zone based protocol applied by ZRP which it allows to be suitable to the large area. Although, the proactive protocols OLSR and DSDV are the worst in all mobility models. Proactive protocols generally offer bad results in large simulation field. We observe that ZRP is the best in the PDR in this area. As a result of dividing spacious simulation area in a small zone which will be easier to verify transmitted packets.

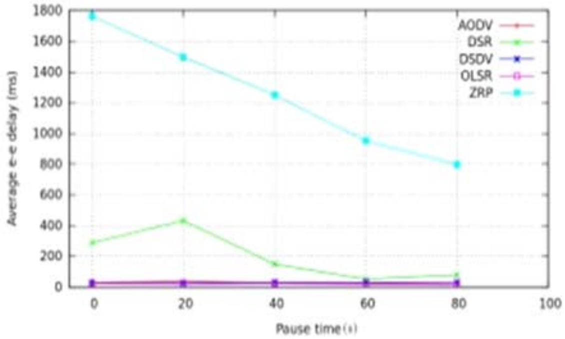
Secondly, we analyze the 'Average End-to-End Delay'. It represents total time spends between application source to destination one. The simulation results are shown in Fig. 3 and Fig. 4.

Fig. 3 is applied in a small area. From Fig. 3(a), (b) and (c), the Average end-to-end delay of DSR and ZRP are the worst in this three mobility models simulated. Due to their zone approach of ZRP and useless routes saved by DSR. However, we notice that in the small area, this metric is best with AODV, OLSR, and DSDV. Thanks to their on demand or continuous proactive strategy adopted by these routing protocols.

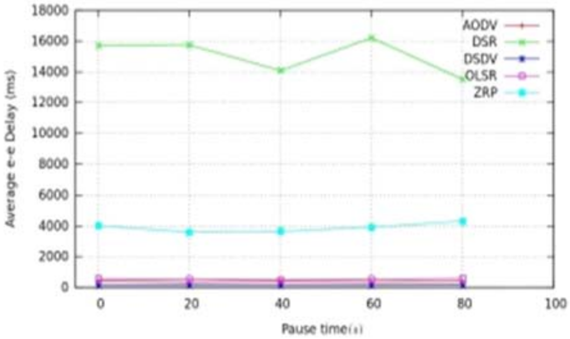
Fig. 4 is applied in a wide area. From Fig. 4(a), (b) and (c) like the small one, the average end-to-end delay of DSR and ZRP are the worst in these three mobility models simulated of Fig. 4(a), (b) and (c). Due to their zone approach of ZRP and useless routes saved by DSR. So, sometimes, they borrow prolonged routes to reach the destination. However, AODV has acceptable results. Thanks to the reactive methodology which send to one neighbor without total knowledge of a correct path to the destination. We notice that average end-to-end delay of proactive protocols OLSR and DSDV are not influenced by simulation field adopted. It offers best outcomes, thanks to their continuous proactive strategy.

Thirdly, we assess the Throughput which is the sum of data rates which are delivered to all mobile nodes, indicating bits or packets received per second. The simulation results are shown in Fig. 5 and Fig. 6.

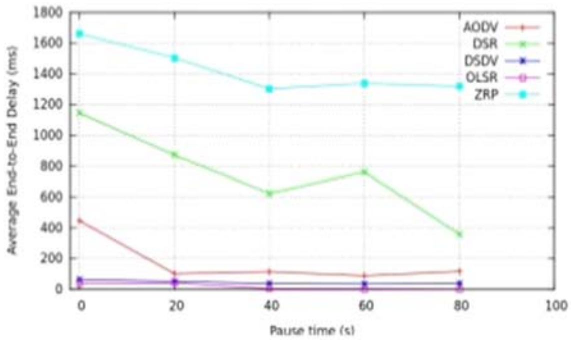
Fig. 5 is applied in a small area. From Fig. 5(a) and (c), reactive protocols AODV and DSR show best results in RWMM and RPGMM. These protocols are suitable for small areas. But from Fig. 5(b), AODV outperforms than others at MGMM due to the reliable path used. However, ZRP is the worst in RWMM and MGMM. And, it is admissible in RPGMM. Although, DSDV and OLSR offer permissible outcomes in RWMM and MGMM.




(a) RWMM

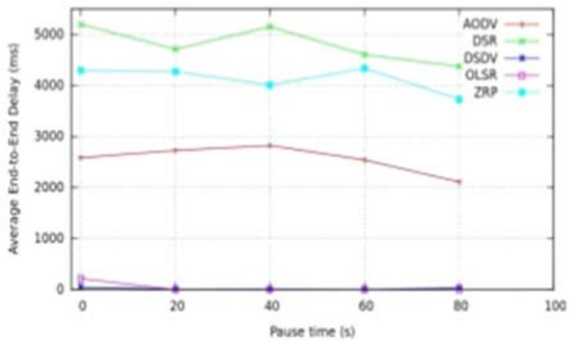



(b) MGMM

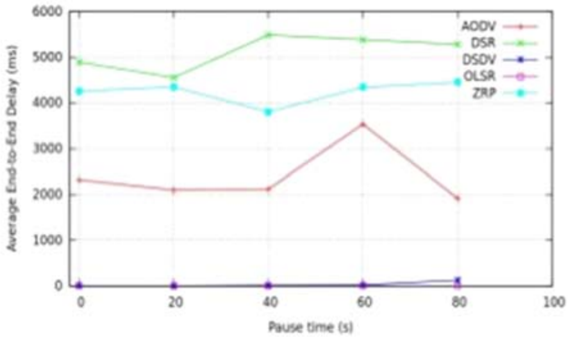



(c) RPGMM

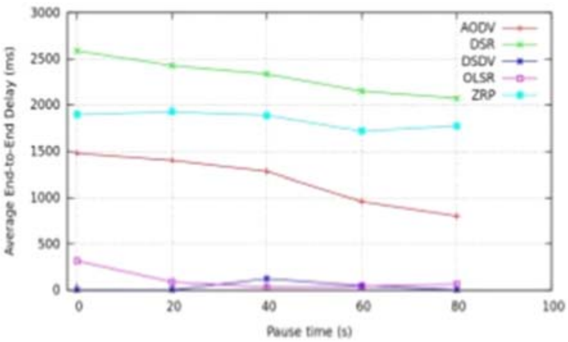
**Fig. 3.** End-to-End Delay of routing protocols under various mobility models - Small area. (a) Random Waypoint Mobility Model, (b) Manhattan Grid Mobility Model, (c) Reference Point Group Mobility Model.




(a) RWMM




(b) MGMM




(c) RPGMM

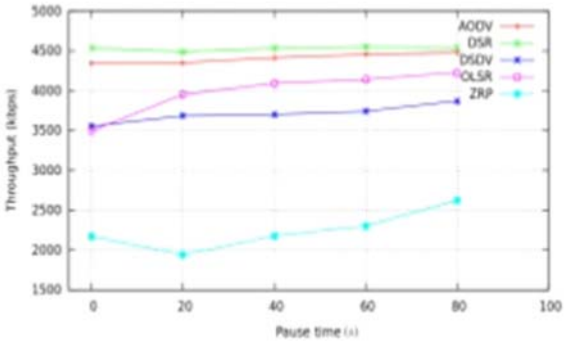
**Fig. 4.** End-to-End Delay of routing protocols under various mobility models - Large area. (a) Random Waypoint Mobility Model, (b) Manhattan Grid Mobility Model, (c) Reference Point Group Mobility Model.

But, OLSR is the worst in RPGMM due to its cluster routing process. We conclude that AODV is the most suitable for all these mobility models simulated in a small field.

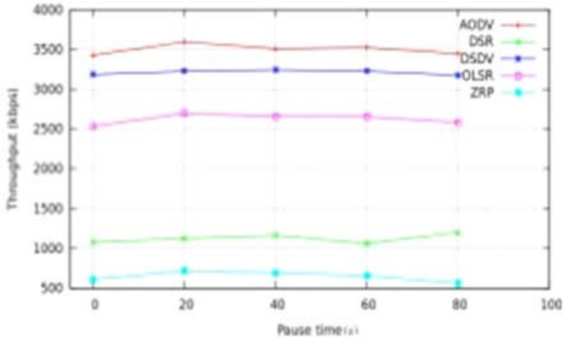
Fig. 6 is applied in a large area. From this figure, we remark that ZRP, AODV, and DSR give best results on the throughput. Furthermore, ZRP is the best according to this metric. But, OLSR and DSDV are the worst at all models experimented. We conclude that proactive protocols are bad. And, ZRP is the best one in large areas.

After simulating 1350 files of 90 different scenarios. Our results will be summarized in Table 2. When we have combined some routing protocols with synthetic mobility models. We obtain best outcomes which are displayed with green cells 1-2. And worst results with red color 4-5.

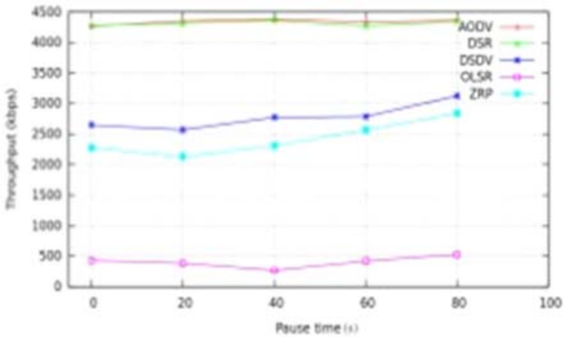
We result from that the Packet delivery ratio and Throughput in a small area. AODV achieve best outcomes as a result of on-demand concept based on route request RREQ and route reply RREP leads to possesses exactly the correct path. But, the worst one is represented by ZRP because it explores information of Intra-zone Routing Protocol (IARP) and Inter-zone Routing Protocol (IERP) which will be tedious to coordinate between them in a small one.




(a) RWMM

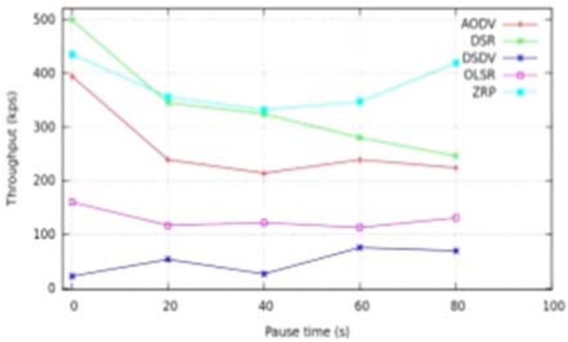



(b) MGMM

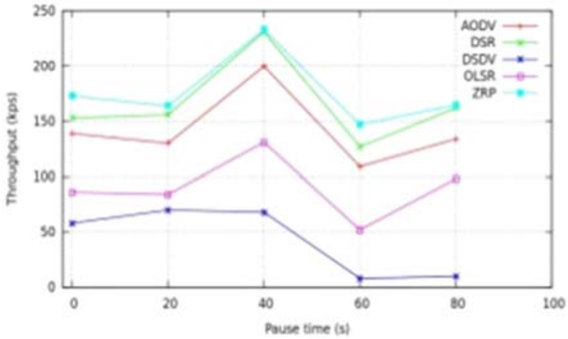



(c) RPGMM

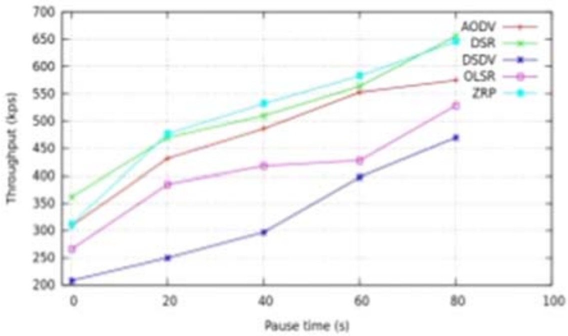
**Fig. 5.** Throughput of routing protocols under various mobility models - Small area. (a) Random Waypoint Mobility Model, (b) Manhattan Grid Mobility Model, (c) Reference Point Group Mobility Model.




(a) RWMM




(b) MGMM




(c) RPGMM

**Fig. 6.** Throughput of routing protocols under various mobility models - Large area. (a) Random Waypoint Mobility Model, (b) Manhattan Grid Mobility Model, (c) Reference Point Group Mobility Model.

For the large area, we acquire best results with DSR due to the available path to a destination node, even if in a wide area. And ZRP as a result of dividing spacious simulation area in a small zone which will be easier to verify transmitted packets.

However, for the average end-to-end delay in the two areas, we have best results with proactive protocols DSDV and OLSR, due to their researches in advance and continuous updates or routing tables. So, all the time, they possess correct paths to a destination. But, the worst are obtained with DSR as a result of hidden table without any strategy to erase it, and ZRP due to speed occupied to locate the destination in a specific zone in simulation field.

## 4. Conclusions

As a reaction to the huge research directed towards mobility models, the principal goal is to analyze performance for any kind of mobile network which is conceived to revivify the real-life scenarios better for applications. Many mobility models have been used to study the mobile ad hoc network performances and evaluate various parameters that can be suitable. This paper aimed to summarize several performance evaluation scenarios of MANET routing protocols under different mobility models.

**Table 2.** Experimental synthesis results.

| Performance metrics | Routing protocols | Mobility models | | | | | |
|---|---|---|---|---|---|---|---|
| | | Small area | | | LARGE area | | |
| | | RWMM | MGMM | RPGMM | RWMM | MGMM | RPGMM |
| PDR | AODV | 2 | 1 | 1 | 3 | 3 | 3 |
| | DSR | 1 | 4 | 2 | 1 | 2 | 2 |
| | DSDV | 4 | 2 | 3 | 5 | 5 | 4 |
| | OLSR | 3 | 3 | 5 | 4 | 4 | 5 |
| | ZRP | 5 | 5 | 4 | 2 | 1 | 1 |
| Average end-to-end delay | AODV | 3 | 3 | 3 | 3 | 3 | 3 |
| | DSR | 4 | 5 | 4 | 5 | 5 | 5 |
| | DSDV | 2 | 1 | 2 | 1 | 2 | 1 |
| | OLSR | 1 | 2 | 1 | 2 | 1 | 2 |
| | ZRP | 5 | 4 | 5 | 4 | 4 | 4 |
| Throughput | AODV | 2 | 1 | 1 | 3 | 3 | 3 |
| | DSR | 1 | 4 | 2 | 2 | 2 | 2 |
| | DSDV | 4 | 2 | 3 | 5 | 5 | 5 |
| | OLSR | 3 | 3 | 5 | 4 | 4 | 4 |
| | ZRP | 5 | 5 | 4 | 1 | 1 | 1 |

Three mobility models have been applied in order to study the impact of changed metrics as average end-to-end delay, throughput, and the packet delivery ratio. We conclude that AODV offers best results in the small area. It is usually moderate or better for all ninety divers' scenarios. It represents an adaptable routing protocol under varied mobility models for the small and large area. Due to reactive routing approach which leads it to own correct path according to packets transmitted. However, ZRP is the worst. But, it is the best in the large one. And proactive protocols are the worst in this field. Three tracks in mobility modeling are allowed which we achieve the first one in this paper. Basing on one itemized analysis with all these details; leads to well understand the accurate behaviors of routing protocols and mobility models used. Our future work will focus on modeling a human trace mobility model applied in a real world scenario. That will be interesting for mobile P2P application and suitable to a crowded area.

## Acknowledgements

This paper was supported by the project PPR n13/2016 of Mohammed V University and recherché laboratory on computer sciences and telecommunication (LRIT), Faculty of sciences, Rabat.

## References

[1]. Jonathan Loo, Jaime Lloret Mauri, Jesùs Hamilton Ortiz, Mobile ad hoc networks: current status and future trends, *CRC Press*, 2016.

[2]. Ashish Srivastava, Atul Mishra, Bikash Upadhyay, Akhilesh Kumar Yadav, Survey and overview of Mobile Ad-Hoc Network routing protocols, in *Proceedings of the IEEE International Conference on Advances in Engineering and Technology Research (ICAETR)*, 2014, pp. 1-6.

[3]. S. Umamaheswaran, Naresh Kumar, P. K. Biswas, A survey of routing protocol design patterns for mobile ad hoc networks, in *Proceedings of the IEEE 2nd International Conference on Emerging Technology Trends in Electronics, Communication and Networking (ET2ECN)*, 2014, pp. 1-6.

[4]. Suvadip Batabyal, Parama Bhaumik, Mobility models, traces and impact of mobility on opportunistic routing algorithms: A survey, *IEEE Communications Surveys & Tutorials*, Vol. 17, No. 3, 2015, pp. 1679–1707.

[5]. Guangjie Han, Jinfang Jiang, Chenyu Zhang, Trung Duong, Mohsen Guizani, George Karagiannidis, A survey on mobile anchor node assisted localization in wireless sensor networks, *IEEE Communications Surveys & Tutorials*, Vol. 18, No. 3, 2016, pp. 2220-2243.

[6]. Vinícius F. S. Mota, Felipe D. Cunha, Daniel F. Macedo, José M. S. Nogueira, Antonio A. F. Loureiro, Protocols, mobility models and tools in opportunistic networks: A survey, *Computer Communications*, Vol. 48, 2014, pp. 5-19.

[7]. Youssef Saadi, Said El Kafhali, Abdelkrim Haqiq, Bouchaib Nassereddine, Simulation analysis of routing protocols using Manhattan grid mobility model in MANET, *International Journal of Computer Applications*, 45, No. 23, 2012, pp. 24-30.

[8]. Francisco J. Martinez, Manuel F-ogue, Chai-Keong Toh, Juan-Carlos Cano, Carlos T. Calafate, Pietro Manzoni, Computer simulations of VANETS using realistic city topologies, *Wireless Personal Communications*, Vol. 69, No. 2, 2013, pp. 639-663.

[9]. Qian Dong, Waltenegus Dargie, A survey on mobility and mobility-aware mac protocols in wireless sensor networks, *IEEE Communications Surveys & Tutorials*, Vol. 15, No. 1, 2013, pp. 88-100.

[10]. Daniel Hiranandani, Katia Obraczka, J. J. Garcia-Luna-Aceves, MANET protocol simulations considered harmful: the case for benchmarking, *IEEE Wireless Communications*, Vol. 20, No. 4, 2013, pp. 82-90.

[11]. Saif Al-Sultan, Moath M. Al-Doori, Ali H. Al-Bayatti, Hussien Zedan, A comprehensive survey on vehicular

ad hoc network, *Journal of Network and Computer Applications*, Vol. 37, 2014, pp. 380-392.

[12]. Fahim Maan, Nauman Mazhar, MANET routing protocols vs mobility models: A performance evaluation, in *Proceedings of the 3rd IEEE International Conference on Ubiquitous and Future Networks (ICUFN)*, 2011, pp. 179-184.

[13]. Ola Alsaqour, Raed Alsaqour, Tariq Alahdal, Rashid Saeed, Mohammed Al-Hubaishi, A comparative study of simulation based performance evaluation of routing protocol for ad-hoc networks, in *Innovations and Advances in Computing, Informatics, Systems Sciences, Networking and Engineering, Springer*, 2015, pp. 215-221.

[14]. Arvind Kumar Shukla, C. K. Jha, Rajeev Arya, A simulation study with mobility models based on routing protocol, in *Proceedings of the 5th International Conference on Soft Computing for Problem Solving, Springer*, 2016, pp. 867-875.

[15]. Jogendra Kumar, Broadcasting traffic load performance analysis of 802.11 mac in mobile ad hoc networks (MANET) using random waypoint model (RWM), *International Journal of Information and Network Security*, Vol. 1, No. 3, 2012, pp. 223-227.

[16]. Caleb Phillips, Douglas Sicker, Dirk Grunwald, A survey of wireless path loss prediction and coverage mapping methods, *IEEE Communications Surveys & Tutorials*, Vol. 15, No. 1, 2013, pp. 255-270.

[17]. N. Aschenbruck, R. Gerhards-Padilla, E. Schwamborn, Bonn Motion. A mobility scenario generation and analysis tool, Documentation, Version July 8, 2013, Osnabruck, Germany, 2013.

[18]. Ibadah N., Minaoui K., Rziza M., Oumsis M., Experimental Synthesis of Routing Protocols and Synthetic Mobility Modeling for MANET, in *Proceedings of the 6th International Conference on Sensor Networks (SENSORNETS'17)*, 2017, pp. 168-173.

*Sensors & Transducers, Vol. 212, Issue 5, May 2017, pp. 31-38*

**Sensors & Transducers**

Published by IFSA Publishing, S. L., 2017
http://www.sensorsportal.com

# Home Sound: A GPU-based Platform for Massive Data Acquisition and Processing for Acoustic Ambient Assisted Living Applications for Behavior Monitoring

**[1] Joan NAVARRO, [2] Rosa Mi ALSINA-PAGÈS and [2, *] Marcos HERVÁS**

[1] GRITS - Grupo de Recerca en Internet Tecnologies & Storage (La Salle - Universitat Ramon Llull), Quatre Camins 30 (Barcelona), 08022, Spain

[2] GTM - Grup de recerca en Tecnologies Mèdia (La Salle - Universitat Ramon Llull), Quatre Camins 30 (Barcelona), 08022, Spain

[2] Tel.: +34-932902445, fax: +34-932902385

[*] E-mail: mhervas@salleurl.edu

*Received: 11 April 2017 /Accepted: 29 May 2017 /Published: 31 May 2017*

**Abstract:** Human life expectancy has grown over the last century, which has driven governments to increase the efforts on caring about the eldest population. Therefore, modern trends take advantage of latest advances in technology to remotely monitor those people with special needs at their home, increasing their life quality and with less impact on their social lives. This paper presents an acoustic event detection platform for assisted living that tracks patients' status by automatically identifying and analyzing the acoustic events happening in a house. Specifically, we have taken benefit of a Jetson TK1, with its NVIDIA Graphical Processing Unit, to process the acoustic data and identify a closed number of events in order to inform the care system. This is a proof of concept conducted with data of only one acoustic sensor, but we plan in the future to deploy a sensor network in several places in the house.

**Keywords:** Ambient assisted living, Sensor network, Machine hearing, Acoustic feature extraction, Machine learning, Graphics processor unit.

## 1. Introduction

Human life expectancy is increasing in the modern society [1]. Our society has to face new challenges in terms of health care because the number of patients to attend is increasing according to [2-3] the people ageing who need support [4]. Nowadays, public and private health services try to avoid long term hospitalizations and, instead, foster the elderly to remain at home for two reasons: on the one hand, it is better for their health to keep them – while not suffering from severe deterioration – in their own environment and, on the other hand, it is much cheaper for health services and care systems. However, nowadays there is still a quality gap between the service provided at medical facilities and the service provided at patients' home.

Technology is a powerful tool that can contribute to address this problem by enabling medical staff to monitor and attend patients while they are at home. Ambient Assisted Living (AAL) [5] can reduce the personnel costs in health assistance. AAL consists of monitoring the preferred living environment of the patients with intelligent devices that can track their status and improve their life quality, as well as obtain information about their behavior, which in the future can lead the doctors to conclusions that with hospital visits could not identify. Some incipient illnesses show symptoms occasionally in time, so a short medical visit is not able to identify preliminary signals. Acoustic Ambient Assisted Living, by means of acoustic event detection, and focusing on behavioral monitoring can help early diagnoses of severe diseases.

To address this hot research topic, several engineering projects have been proposed to discuss the feasibility of deploying smart robots at the home of elderly not only to cover routine tasks, but also to remind them to have their medication or interact with them through serious games [6]. One of the main challenges that these proposals open is the huge amount of data that these robots have to collect in order to provide a meaningful response for patients. Typically, these robots have limited computing capabilities and, thus, are able to process data from a reduced number of sensors.

This paper explains the proof of concept of a software and a chosen hardware platform designed to recognize a set of the predefined events from the environmental sound in a house [7]. This information can be later used to infer the in-home context and detect some situations of risk. To process data from several sources (e.g., microphones) and conduct the computations associated to audio event identification in parallel, the system implements a recognition scheme using a NVIDIA Jetson TK1 [8] Graphical Processing Unit (GPU). This platform can reach to several decisions depending on the situation and home, and the final conclusion can be activating some kind of alarm or just track the patient's behaviour for health purposes. Overall, the purpose of this work is to present an approach to the implementation of an acoustic event recognition platform based on a GPU and the obtained results when classifying a limited corpus of events.

The reminder of this paper is organized as follows. Section 2 reviews the related work on environmental sound recognition; it is specially focused on ambient assisted living environments. Section 3 elaborates on the technical details of the proposed algorithm to solve the problem, which corresponds to a basic implementation. Section 4 gives details about the selected platform and its convenient features to process audio data. Section 5 describes the algorithm used to classify the events and shows the obtained results when running on the chosen platform. Finally, Section 6 details the conclusions and future work of this project.

## 2. Related Work

There are several approaches in the literature that aim to extract features from the sound. From these features, it is possible to create a corpus of a close universe of different sounds and train a machine learning system to classify the source of the sound. Therefore, environmental sound recognition has emerged as a hot research topic today, which has led to some interesting applications [9]; from animal recognition [10] to surveillance [surveillance], including ambient assisted living use cases.

Interest in detecting in-home sounds started from the beginning of this technology in 2005. Chen, *et al.* [11] were monitoring the bathroom activity using only the sound information. Afterwards, with research not detailed in this work, robust environment sound recognition motors were designed in 2008 [12]. One of the most challenging problems to be solved in this field, is to take into account the varying acoustic background, the noise sources. In this regard, the project SonicSentinel [13] uses noise-robust model-based algorithms to evaluate the noise sources. Evolving this technology, Valero, *et al.* [14] succeeded on classifying audio scenes. Additionally, several works can be found about audio analysis in a smart home to help doctors on the early diagnose of dementia diseases for the elder [15]. Also, it is worth mentioning that conditional random fields have been used to build an event detection framework in a real-world environment of eight households [16], which led the system to be sometimes unreliable.

From the applications point of view, one of the most popular use-cases nowadays of audio event recognition is its use in the smart home [17], especially when conceiving systems to meet the needs of the elderly people. The constraints around the design of a smart home for health care [5] based on audio event classification are as follows:

1) Degree of dependency of the disabled person,
2) Quality of life to be improved by means of automatizing the processes,
3) Distress situations recognition and the activation of the preassigned protocols, including reducing the false alarm situations [18].

Even though there are several solutions in the literature [19] that consider these three constraints, the primary goal of the platform presented in this paper is to accurately address the third one. Additionally, our proposal aims to meet the needs of ambient assisted living, which are the following [20]:

1) Increasing the comfort of living at home,
2) Increasing the safety, through detecting dangerous events,
3) Supporting health care by professionals, through detecting emergencies and monitoring vital signs.

## 3. System Description

When designing and deploying an Acoustic Wireless Sensor Network, the main parameter to be considered is power consumption. Generally, the power consumption of a node in a wireless network strongly depends on its assigned duties (e.g., data acquisition, data storage, data computing, data forwarding). Therefore, system architects have to carefully select which devices conduct every task. In this regard, with the growth of the number of mobile user equipment (UE) devices and the advent of the Internet of Things, latest advances on the distributed systems field have proposed several approaches and reference models to offload the duties of each device by means of the Mobile Edge Computing (MEC) paradigm [21]. Indeed, MEC consists of deferring the power consuming activities to specialized and dedicated devices close to the wireless sensor network. Therefore, it can be best seen as a particular case of cloud computing where the storage and computation infrastructure are physically close to where data are generated, which brings appealing advantages such as data security, reduced communication delay, and energy efficiency [21]. This section enumerates the latest contributions in the MEC field, justifies the selected alternative for the Ambient Assisted Living use case proposed in this paper, and details its deployment.

So far, when designing a MEC architecture four main approaches have emerged whose details are further elaborated in [21]. These alternatives are summarized in what follows:

1. Small Cell Cloud (SCC): It consists of extending the capabilities of the UE by include a Small Cell Manager committed to forward the UE requests to a storage and computation cluster.

2. Mobile Micro Clouds (MMC): It consists of deploying a network of device managers each one committed to forward the UE requests from a small set of UEs. These device managers are interconnected between themselves and also connected to a storage and computation cluster. This approach minimized the load of the device manager, which makes it suitable for scenarios with a high number of UE devices.

3. Fast Moving Personal Cloud (FMPC): It consists of using Software Defined Networks and Network Function Virtualization to build a dynamic and adaptable set of forwarding devices to link the UE with the storage and computation cluster. Therefore, it uses a SDN enabled transport network typically residing in a cloud. This is a feasible approach for those application that change their traffic patterns frequently and are not very sensible to delay.

4. Follow Me Cloud (FMC): It consists of embedding the UE devices inside the cloud storage and computing cluster. In this way, the perception that the cloud following the roaming UE is given.

For the sake of our Ambient Assisted Living proposal in which an acoustic wireless sensor network based on microphones (i.e., UE) is deployed inside a home environment, the aforementioned four alternatives have been considered. FMPC and FMC approaches are designed for environments where UE are moving, which is not the case of AAL where the microphones are permanently installed in the same place.

MMC might be a feasible option if the home environment was big enough to deploy a high number of microphones (i.e., hundreds). Considering that there will be two microphones per room, the overall number of UE per home environment is still below the threshold imposed by MMC.

Therefore, we have selected the SCC option that is also convenient taking into account the low latency (i.e., AAL is committed to operate in real-time) and budget constraints requirements of the AAL paradigm. As a result, the proposed system diagram to monitor audio events in Ambient Assisted Living environments is shown in Fig. 1.

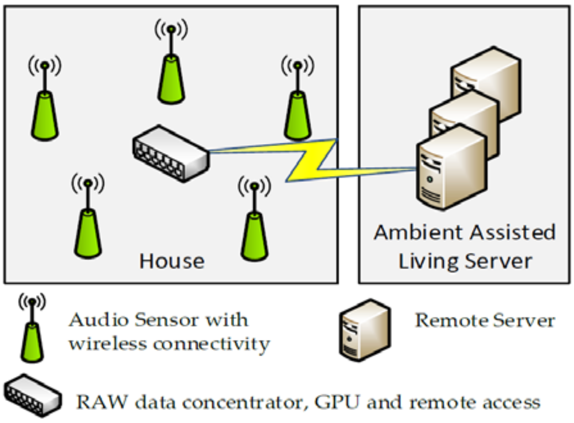



**Fig. 1.** Block diagram of the network elements of this system.

As far as the proof of concept herein presented is concerned, the system relies on a network of microphones consistently deployed around the house (see Fig. 2). The microphones are installed in such a way that they provide the maximum entropy of a given event (i.e., it is not necessary to analyze together different audio sources).

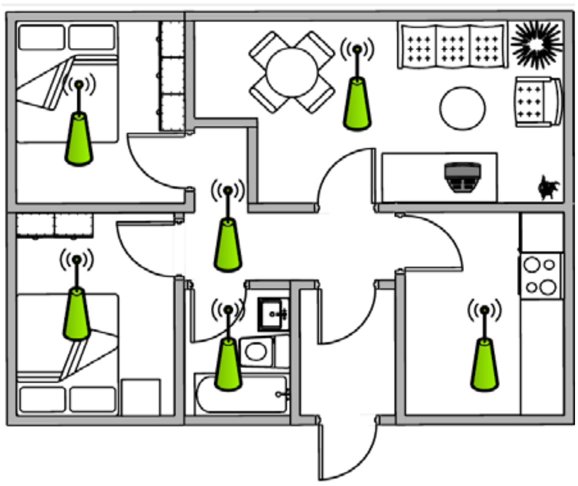


**Fig. 2.** Example of the proposed audio sensors network deployed in a house.

The microphones used in this application to sense the environmental sound should present a good trade-off between the frequency response and cost, for this reason tests are being conducted with the electret condenser microphone CMA-4544PF-W [22] of the manufacturer CUI inc. with a very low price.

In this way, each microphone transmits sounds to this device that acts as a concentrator – the core element of our proposal. As a matter of fact, this concentrator

1) Collects all the audio sounds of the house,
2) Processes them to extract their features,
3) Infers the source of the audio event,
4) Sends this information to a remote server that monitors the needs of the people living in the house.

The concentrator platform used in this work is the NVIDIA Jetson TK1 developer kit. This platform is based on the Tegra K1 SoC, which is composed of

1) NVIDIA Kepler GPU with 192 CUDA cores,
2) Quad core ARM cortex-A15 CPU.

The Tegra family is the proposal of the NVIDIA manufacturer for mobile processors in which you need GPU-accelerated performance with low power consumption.

This GPU is able to process up to 192 threads in parallel. Kepler architecture offers an improvement of performance up to 3 times more than the previous version, Fermi, [23]. This level of concurrency allows us to process audio events of several sources in real-time.

Therefore, to exploit the parallel capabilities of the concentrator, it opens a thread to process each audio source and infer the event that generated every sound.

## 4. Machine Learning

Endowing machines with the ability of hearing the acoustic environment to detect and recognize an event as humans do, is known as machine hearing. The algorithm used in this work is based on

1) Feature extraction using mel-frequency cepstral coefficients (MFCC) [24];
2) Pattern recognition using the k-Nearest Neighbors classifier (KNN) [25], see Fig. 3.


**Fig. 3.** Block diagram of a Hearing Machine algorithm.

### 4.1. Feature Extraction

Feature extraction aims to obtain a representation of audio events in which the dimensionality of this parametrization is much lower than the original samples [26]. This parametrization will be the input data of the classifier. The parametrization used in this work, MFCC [24], uses an approach based on perceptual-based frequency using the Mel scale [27] as shown in Fig. 4.

The incoming audio stream is divided into blocks of 30 ms with a sliding window. These frames are transformed into frequency domain using the DFT to measure the power of different bands of the spectrum. The power measures are conducted with a bank of 48 filters using the Mel scale (see Fig. 5).

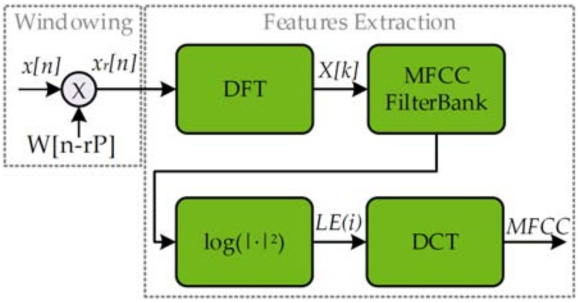



**Fig. 4.** Block diagram of the feature extraction based on the Mel coefficients used in this work.

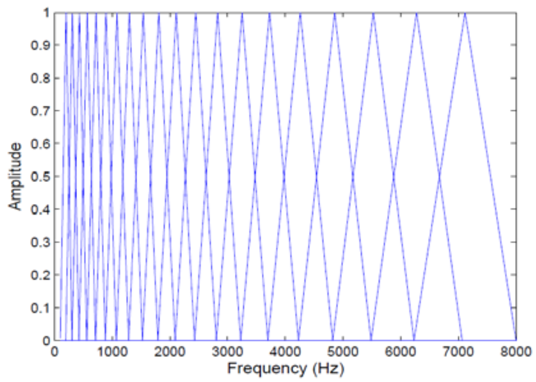



**Fig. 5.** Example of a Mel scale with a filter-bank of 20.

The MFCC coefficients are obtained from the Discrete Cosine Transform (DCT) of the logarithm of these 48 values. The higher order coefficients of the DCT are discarded to obtain a reduced dimensionality characterization of the sound event, this compression can be done because the main information is in the low frequency components of the signal's spectral envelop. The final number of MFCC coefficients is 13.

Window lengths between 10 and 50 ms are usually used to detect transient audio events [28]. A Hamming windowing is also applied to this frame of samples to improve the frequency resolution in the Discrete Fourier Transform (DFT) – as we can see comparing the differences between square and Hamming windows in Fig. 6. This sliding block has an overlap of 50 % of samples to compensate the power reduction of the data blocks due to the laterals of the Hamming window, see Fig. 6.

The Mel scale is a perceptual scale which aims to emulate the behaviour of the human hearing. As we can observe in Fig. 5, Mel scale is a bank of triangular filters.

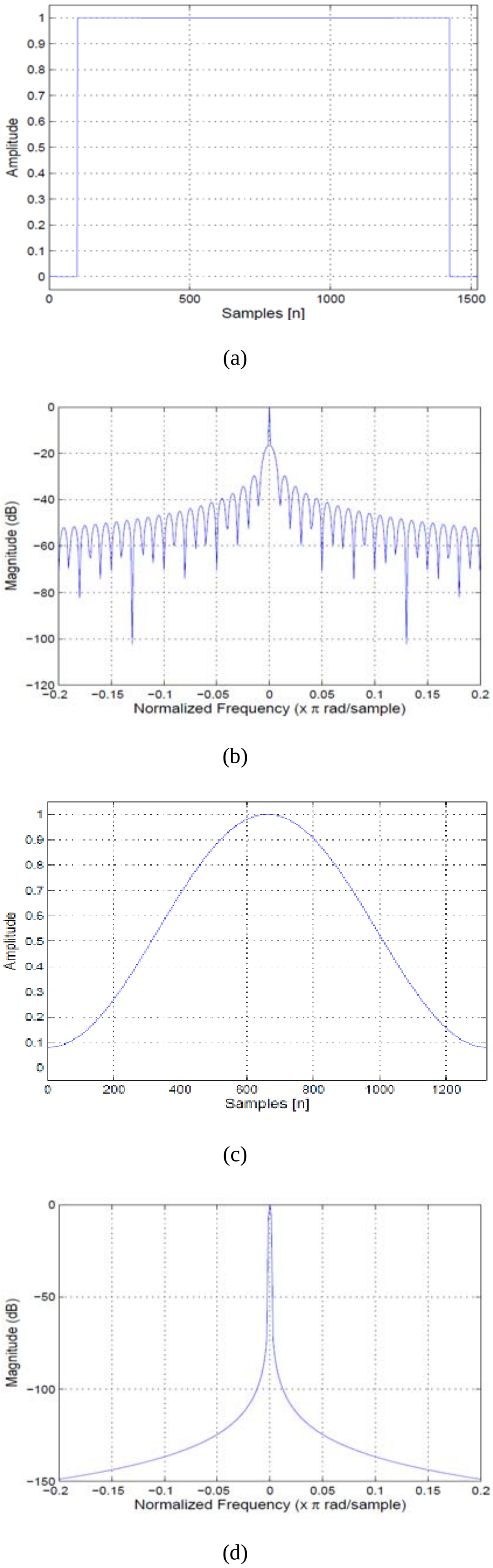



**Fig. 6.** Comparison between a squared and Hamming windows in time and frequency domain: (a) is a squared window, (b) is the spectrum of the squared window, (c) is a Hamming window and (d) is the spectrum of the Hamming window.

### 4.2. Automatic Audio Classification

Machine learning algorithms are widely used in the literature of speech technologies to automatically classify audio samples. In fact, most of the audio recognition systems settle the use of the MFCC coefficients as baseline in terms of feature extraction [26]. Then, when the signal is processed and the features are already extracted, a k-Nearest Neighbors (kNN) [25] system can be run [29].

Hence, we have followed this approach and trained a kNN classifier as follows. We have built a training data set composed by 2850 audio samples belonging to 14 in-home events lasting a total number of 20 hours. We have split every sample in several sub samples as detailed in the previous section, and for every sub sample we have computed the MFCC coefficients. This results in a vector of 13 components (each one corresponding to its associated MFCC) for every sound sub sample. As a result, a sound sample is characterized with a set of 13-component vectors.

For the sake of this paper we have implemented two classification strategies: raw-kNN and SVM.

The raw-kNN attempts to obtain a lower bound of up to what extent it is feasible to classify the training data set. In this regard, a simple 13-dimensions k-NN has been built and a grid search to come out with the best *k* parameter has been run. Actually, this classification strategy can be best seen as a worst case scenario in which the entropy provided by previous and subsequent subsamples is deliberately neglected. Thus, in order to classify a given subsample we only consider the closest *k* samples to it. As far as the computation cost is concerned, it is worth mentioning that the kNN has a linear cost (i.e., the input subsample has to be compared against all the vectors of the dataset. However, this overhead has been greatly alleviated by implementing a map-reduce inspired strategy to compare different subsets of the dataset in parallel. Specifically, we have assigned a thread to a segment of the dataset that conforms the kNN and, next, each thread shares its *k* nearest neighbors. Then, the output of a group of threads is collected by another thread in charge of selecting again the *k* nearest neighbors. This process is done recursively until the wining *k* neighbors have survived the whole recursively map and reduce process.

The second classification strategy has been designed to improve the results of the kNN classifier by considering the information of the subsamples belonging to the same sample. As the number of vectors that characterize a given sound depends on the length of the training sound, there is an inherent class imbalance in the formulation of this problem, which limits the classifier accuracy (i.e., shorter sounds of the same sound type would probably be misclassified). Therefore, to address this issue, we have built a bag of words with all the vectors belonging to the same sample using the k-means algorithm [30]. The resulting vector has a fixed length of K components. This gives an idea of how many portions of the training sound set belong to each centroid of the k-means, which at the same time removes the temporal dimension of the sound event. Next, we normalize all these resulting vectors to make the suitable for a fair comparison. With this fixed size set of normalized vectors, we finally train a Support Vector Machine (SVM) that is running on the concentrator platform and uses a one-against-all strategy to deal with this multiclass problem. That is, a SVM has been built for every class in which the class of interest has been labelled as positive and all other classes as negative. Then the output of each SVM will be considered as a confidence factor. When this confidence factor is above a heuristic threshold, the output of the SVM will be considered as positive. To come out with this heuristic threshold we have held out a 10 % of the dataset and conducted a 4-fold cross-validated grid search. Analogously, we have found the best offset parameter for each SVM using the same strategy.

Finally, when our system is in exploitation mode, the concentrator platform extracts the audio sub samples and builds the fixed size vector accordingly. Then, this vector is delivered to all the SVMs that had been previously tested to predict the event. In order to obtain a positive outcome, a single SVM has to provide an output above the aforementioned threshold. If more than one SVM provides an output above the threshold, no class is assigned to that sample.

## 5. Results

With the dataset and the techniques described in the previous section we have conducted our experimentation to detect the following events: someone falling down, slice, screaming, rain, printer, people talking, frying food, filling water, door knocking, dog bark, car horn, glass breaking, baby crying, water boiling. We have used 60 % of instances to train the classifiers and the other 40 % to test them. In order to obtain statistically significant results we have run the classification in 1000 runs, performed a 10-fold cross validation, and averaged the output.

The obtained confusion matrix for the kNN classifier is shown in Table 1 with an overall accuracy of 50.24 %. In this confusion matrix we can see how often the kNN misclassifies a given class and, thus, assigns a wrong event to an audio sample. It is shown that in general, the best results for each sample are obtained when testing the sound event against itself. Also, it depicts the skill of the classifier on distinguishing one audio event from the others. The optimal value of this confusion matrix should be an Identity Matrix with the value 100 on its diagonal.

We can see that although some events are identified with a reasonable degree of accuracy (e.g., screaming), some others (e.g., dog barking) cannot be classified properly.

**Table 1.** Confusion Matrix of the kNN classifier. Events are ordered from left to right as follows: falling down, slice, screaming, rain, printer, people talking, frying food, filling water, door knocking, dog bark, car horn, glass breaking, baby crying, water boiling.

| | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 58.54 | 0.00 | 2.44 | 0.00 | 4.88 | 2.44 | 7.32 | 4.88 | 1.22 | 2.44 | 0.00 | 10.98 | 3.66 | 1.22 |
| **1** | 0.00 | 60.26 | 0.00 | 3.85 | 0.00 | 11.54 | 3.85 | 0.00 | 6.41 | 2.56 | 7.69 | 0.00 | 1.28 | 2.56 |
| **2** | 0.00 | 0.00 | 89.33 | 0.00 | 0.00 | 0.00 | 0.00 | 3.33 | 0.00 | 4.00 | 1.33 | 0.00 | 0.67 | 1.33 |
| **3** | 0.00 | 1.54 | 0.00 | 76.92 | 0.77 | 2.31 | 3.08 | 2.31 | 0.00 | 4.62 | 6.92 | 0.00 | 0.00 | 1.54 |
| **4** | 9.52 | 1.19 | 0.00 | 2.38 | 55.95 | 4.76 | 3.57 | 0.00 | 2.38 | 7.14 | 0.00 | 8.33 | 1.19 | 3.57 |
| **5** | 2.91 | 8.74 | 1.94 | 4.85 | 3.88 | 18.45 | 1.94 | 0.00 | 6.80 | 18.45 | 7.77 | 1.94 | 3.88 | 18.45 |
| **6** | 7.89 | 6.58 | 1.32 | 3.95 | 3.95 | 3.95 | 15.79 | 15.79 | 1.32 | 15.79 | 1.32 | 2.63 | 13.16 | 6.58 |
| **7** | 1.08 | 0.00 | 3.23 | 2.15 | 0.00 | 0.00 | 5.38 | 60.22 | 0.00 | 9.68 | 10.75 | 0.00 | 4.30 | 3.23 |
| **8** | 1.41 | 4.23 | 0.00 | 0.00 | 1.41 | 9.86 | 0.00 | 0.00 | 76.06 | 1.41 | 0.00 | 5.63 | 0.00 | 0.00 |
| **9** | 2.65 | 0.88 | 4.42 | 7.96 | 1.77 | 10.62 | 6.19 | 4.42 | 3.54 | 14.16 | 25.66 | 3.54 | 3.54 | 10.62 |
| **10** | 0.00 | 5.33 | 0.00 | 10.67 | 0.00 | 8.00 | 0.00 | 13.33 | 0.00 | 32.00 | 21.33 | 0.00 | 1.33 | 8.00 |
| **11** | 8.70 | 1.74 | 0.87 | 0.00 | 6.96 | 1.74 | 1.74 | 0.00 | 3.48 | 5.22 | 0.87 | 68.70 | 0.00 | 0.00 |
| **12** | 3.08 | 0.00 | 10.77 | 0.00 | 3.08 | 4.62 | 10.77 | 9.23 | 0.00 | 6.15 | 1.54 | 1.54 | 30.77 | 18.46 |
| **13** | 1.63 | 3.25 | 1.63 | 1.63 | 0.81 | 11.38 | 1.63 | 4.88 | 0.00 | 7.32 | 4.07 | 0.00 | 4.88 | 56.91 |

For instance, on row 6 in Table 1, door knocking, people talking and frying food have similar MFCC vector patterns and, thus, the SVM features a low accuracy in these specific situations Such a poor performance can be explained because (1) a single subsample is only considered to decide to which class it belongs to (i.e., no information from previous nor subsequent subsamples is considered and (2) the MFCCs associated to some subsamples of these events are pretty similar to other subsamples from other events.

On the contrary, Table 2 shows the confusion matrix obtained when using the bag of words approach and the aforementioned SVM classifier. We can see that this classification strategy reaches an overall accuracy of 72.88 %, which is much better than the kNN. Also, we can see that with this strategy, the classification accuracy is more consistent for all the events (i.e., the worst value 56.23 %). However, despite considering information from previous and subsequent subsamples thanks to the bag of words approach, this approach still gets confused on some sound events. To address this concern, we plan to (1) complement the training vector set with other features in addition to MFCCs, and (2) use a more sophisticated classifier such as a deep net.

## 6. Conclusions

Preliminary results of our paper encourage us to keep on working on the analysis of the events happening in the house. We will work with the feature extraction improvement with other methods, as well as we will test more machine learning algorithms to increase the accuracy of the system with just one acoustic measurement.

Next steps after this proof of concept using the Jetson TK1 are the expansion of the platform, by means of using a wider sensor network, where several autonomous acoustic sensors sending data to the GPU to be processed. In this stage, an important part of the work will be focused on the optimization of the acoustic event detection algorithm to take advantage of the parallelization of the GPU unit.

## Acknowledgements

The authors would like to thank the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement (Generalitat de Catalunya) under grant refs. 2014-SGR-0590 and ref. 2014-SGR-589.

**Table 2.** Confusion Matrix of the SVM classifier. Events are ordered from left to right as follows: falling down, slice, screaming, rain, printer, people talking, frying food, filling water, door knocking, dog bark, car horn, glass breaking, baby crying, water boiling.

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 89.10 | 0.11 | 1.78 | 0.35 | 0.00 | 0.12 | 0.12 | 1.60 | 2.39 | 0.00 | 1.26 | 0.80 | 1.48 | 0.89 |
| **1** | 1.53 | 85.48 | 2.18 | 0.17 | 0.79 | 2.71 | 0.00 | 0.41 | 0.00 | 1.31 | 3.05 | 0.47 | 1.35 | 0.55 |
| **2** | 2.16 | 5.18 | 63.22 | 0.04 | 2.89 | 0.00 | 5.21 | 4.83 | 3.37 | 1.16 | 4.98 | 5.65 | 0.00 | 1.30 |
| **3** | 0.27 | 1.29 | 0.00 | 87.14 | 0.59 | 2.15 | 0.00 | 1.75 | 0.83 | 3.36 | 0.20 | 0.81 | 0.62 | 0.98 |
| **4** | 0.03 | 0.87 | 16.37 | 0.75 | 69.60 | 0.00 | 0.79 | 0.87 | 3.23 | 0.00 | 0.06 | 2.02 | 1.45 | 3.97 |
| **5** | 0.00 | 1.44 | 1.67 | 1.72 | 4.44 | 85.62 | 0.00 | 1.18 | 0.33 | 0.01 | 0.92 | 1.69 | 0.00 | 0.99 |
| **6** | 3.17 | 0.81 | 1.32 | 1.08 | 2.92 | 3.79 | 71.05 | 6.28 | 3.55 | 5.04 | 0.00 | 0.14 | 0.00 | 0.85 |
| **7** | 0.00 | 0.00 | 2.10 | 0.00 | 0.14 | 0.86 | 3.80 | 89.46 | 0.20 | 0.71 | 0.98 | 0.38 | 0.04 | 1.33 |
| **8** | 0.84 | 1.91 | 1.75 | 0.00 | 0.14 | 6.71 | 2.05 | 3.31 | 65.36 | 15.61 | 0.00 | 0.06 | 1.34 | 0.92 |
| **9** | 1.08 | 3.82 | 4.56 | 12.18 | 0.00 | 0.00 | 5.45 | 4.52 | 3.67 | 56.78 | 1.34 | 3.67 | 2.93 | 0.00 |
| **10** | 5.27 | 5.36 | 5.70 | 2.20 | 6.09 | 0.80 | 0.00 | 1.75 | 4.15 | 5.19 | 56.23 | 1.63 | 2.98 | 2.65 |
| **11** | 0.00 | 12.15 | 1.53 | 0.00 | 0.24 | 0.13 | 2.09 | 0.22 | 2.20 | 0.28 | 1.42 | 78.14 | 0.89 | 0.70 |
| **12** | 10.13 | 0.71 | 1.19 | 2.21 | 1.97 | 1.55 | 0.43 | 1.09 | 1.04 | 11.57 | 0.00 | 1.33 | 66.78 | 0.00 |
| **13** | 0.00 | 0.92 | 7.83 | 0.00 | 3.15 | 5.46 | 4.08 | 5.12 | 4.93 | 0.00 | 6.90 | 2.15 | 3.10 | 56.36 |

## References

[1]. R. Suzman, J. Beard, Global health and aging – Living longer, *National Institute on Aging*, 2015.

[2]. Karp F. (Ed.), Growing Older in America: The Health and Retirement Study, *U.S. Department of Health and Human Services*, 2007.

[3]. S. Chatterji, P. Kowal, C. Mathers, N. Naidoo, E. Verdes, J. P. Smith, R. Suzman, The health of aging populations in China and India, *Health Affairs*, Vol. 27, Issue 4, 2008, pp. 1052-1063.

[4]. G. Lafortune, G. Balestat, Trends in Severe Disability Among Elderly People, in Assessing the Evidence in 12 OECD Countries and the Future Implications. OECD Health Working Papers 26, *Organization for Economic Cooperation and Development*, Paris, 2007.

[5]. M. Vacher, F. Portet, Challenges in the processing of audio channels for ambient assisted living, in *Proceedings of the IEEE 12th International Conference on E-health Networking Applications and Services (Healthcom)*, Vol. 12, 2010, pp. 330-337.

[6]. Y. S. Morsi, A. Shukla, Optimizing Assistive Technologies for Aging Populations, *IGI Global*, 2015.

[7]. M. Hervás, R. M. Alsina-Pagès, J. Navarro, homeSound: a High Performance Platform for

Massive Data Acquisition and Processing in Ambient Assisted Living Environments, in *Proceedings of the 6th International Conference on Sensor Networks (SENSORNETS'17)*, Porto, Portugal, 2017, pp. 182-187.

[8]. JETSON TK1. Unlock the power of the GPU for embedded systems applications (http://www.nvidia.com/object/jetson-tk1-embedded-dev-kit.html).

[9]. S. Chachada, J. Kuo, Environmental sound recognition: a survey, in *APSIPA Transactions on Signal and Information Processing*, Vol. 3, 2014, pp. 14-20.

[10]. Stowell D., Wood M., Stylianou Y., Glotin H., Bird detection in audio: a survey and a challenge, in *Proceedings of the IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2016, pp. 1-6.

[11]. J. Chen, A. H. Kam, J. Zhang, N. Liu, L. Shue, Bathroom activity monitoring based on sound, in *Proceedings of the International Conference on Pervasive Computing*, 2005, pp. 47-61.

[12]. J. C. Wang, H. P. Lee, J. F. Wang, C. B. Lin, Robust environmental sound recognition for home automation, *IEEE Transactions on Automation Science and Engineering*, Vol. 5, Issue 1, 2008, pp. 25-31.

[13]. D. Hollosi, S. Goetze, J. Appell, F. Wallhoff, Acoustic Applications and Technologies for Ambient Assisted Living Scenarios, in *Proceedings of the Ambient Assisted Living Forum*, 2011, pp. 337-342.

[14]. X. Valero, F. Alías, Classification of audio scenes using narrow-band autocorrelation features, in *Proceedings of the 20th European Signal Processing Conference*, Bucharest, Romani, August 2012, pp. 2012-2019.

[15]. P. Guyot, J. Pinquier, X. Valero, F. Alías, Two-step detection of water sound events for the diagnostic and monitoring of dementia, in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, San Jose, California, USA, 2013, pp. 1-6.

[16]. D. Matern, A. Condurache, A. Mertins, Adaptive and automated ambiance surveillance and event detection for ambient assisted living, in *Proceedings of the 35th Annual International Conference of the IEEE EMBS*, Osaka, Japan, July 2013, pp. 3-7.

[17]. M. Chan, D. Estève, C. Escriba, E. Campo, A review of Smart homes – present state and future challenges, *Computer Methods and Programs in Biomedicine*, Vol. 91, Issue 1, 2008, pp. 55-81.

[18]. S. Goetze, J. Schroder, S. Gerlach, D. Hollosi, J. Appell, Acoustic Monitoring and Localization for Social Care, *Journal of Computing Science and Engineering*, Vol. 6, Issue 1, 2012, pp. 40-50.

[19]. M. Vacher, F. Portet, A. Fleury, N. Noury, Development of audio sensing technology for ambient assisted living: Applications and challenges, *Digital Advances in Medicine, E-Health, and Communication Technologies*, Vol. 2, Issue 1, 2011, pp. 35-54.

[20]. P. W. J. van Hengel, J. Anemuller, Audio event detection for in-home care, in *Proceedings of the International Conference on Acoustics (NAG/DAGA)*, Rotterdam, Netherlands, 2009, pp. 618-620.

[21]. P. Mach, Z. Becvar, Mobile Edge Computing: A Survey on Architecture and Computation Offloading, *IEEE Communications Surveys & Tutorials*, Vol. PP, Issue 99, 2017, pp. 1-1.

[22]. CMA-4544PF-W (http://www.cui.com/product/resource/pdf/cma-4544pf-w.pdf).

[23]. NVIDIA's next generation cuda compute architecture: Kepler TM GK110/210 (http://international.download.nvidia.com/pdf/kepler/NVIDIA-Kepler-GK110-GK210-Architecture-Whitepaper.pdf).

[24]. P. Melmstein, Distance measures for speech recognition, psychological and instrumental, *Pattern Recognition and Artificial Intelligence*, 1976, pp. 91-103.

[25]. T. M. Cover, P. E. Hart, Nearest neighbor pattern classification, *IEEE Transactions on Information Theory*, Vol. 13, Issue 1, 1967, pp. 21-27.

[26]. F. Alías, J. Claudi, X. Sevillano, A Review of Physical and Perceptual Feature Extraction Techniques for Speech, Music and Environmental Sounds, *Applied Sciences*, Vol. 6, Issue 5, 2016, pp. 143-187.

[27]. S. Liang, X. Fan, Audio Content Classification Method Research Based on Two-step Strategy, *International Journal of Advances in Computer Science Applications*, Vol. 5, Issue 3, 2014, pp. 57-62.

[28]. Z. Fu, G. Lu, K. M. Ting, D. Zhang, A survey of audio-based music classification and annotation, *IEEE Transactions on Multimedia*, Vol. 13, Issue 2, 2011, pp. 303-319.

[29]. M. L. Zhang, Z. H. Zhou, A k-nearest neighbor based algorithm for multi-label classification, in *Proceedings of the IEEE 1st International Conference on Granular Computing*, 2005, pp. 718-721.

[30]. Zhang Y., Jin R., Zhou Z. H., Understanding bag-of-words model: a statistical framework, *International Journal of Machine Learning and Cybernetics*, Vol. 1, No. 1-4, 2010, pp. 43-52.

*Sensors & Transducers, Vol. 212, Issue 5, May 2017, pp. 39-49*

**Sensors & Transducers**
Published by IFSA Publishing, S. L., 2017
http://www.sensorsportal.com

# *Hash Chains Sensornet*: A Key Predistribution Scheme for Distributed Sensor Networks Using Nets and Hash Chains[1]

**Deepak Kumar DALAI and * Pinaki SARKAR**
School of Mathematical Sciences, National Institute of Science Education and Research,
Bhubaneswar 752 050, India
* Tel.: +919433531020
E–mail: pinakisark@gmail.com

*Received: 24 April 2017 /Accepted: 24 May 2017 /Published: 31 May 2017*

**Abstract:** Key management is an essential functionality for a security protocol; particularly for implementations to low cost devices of a distributed sensor networks (DSN)–a prototype of Internet of Things (IoT). Constraints in resources of constituent devices of a low cost IoT (example: sensors of DSN) restricts implementations of computationally heavy public key cryptosystems. This leads to adaptation of the novel key predistribution technique in symmetric key platform to efficiently tackle the key management problem for these resource starved networks. Initial key predistribution schemes (KPS) use random graphs; while later ones exploit combinatorial approaches that assure predictable design properties. Combinatorial designs like a *(v, b, r, k)–configuration* that forms a *μ–CID* are effective schemes to design a KPS. A net in a vector space is a set of cosets of certain kind of subspaces called partial spread. A *μ(v, b, r, k)–CID* can be formed from a net. In this paper, we propose a KPS for DSN, named as Sensornet, using a net. We observe that any deterministic KPS suffer from "smart attack" and hence devise a generic method to eliminate such attacks. Resilience of a KPS can improve by clever application of a Hash Chains technique introduced by *Bechkit et al*. We improve our *Sensornet* to obtain a new *Hash Chains Sensornet (HC(Sensornet))* by the applications of these two generic methods. Effectiveness of *Sensornet* and *HC(Sensornet)* in term of crucial metrics in comparison to prominent schemes has been theoretically established.

**Keywords:** Distributed sensor networks, Key management, Combinatorial designs, Attacks, Hash function.

## 1. Introduction

Distributed (Wireless) Sensor Networks (DSN) are revolutionary information gathering systems owing to their easy deployment and flexible topology. They are decentralized with numerous low cost identical resource starved wireless devices, called sensors or nodes that deal with sensory data.

They are considered to be a nice prototype of Internet of Things (IoT) which is a sophisticated concept that aims to connect our world beyond imagination. This has boosted the study of such DSN in recent times.

Prominent scientific applications of IoT are smart homes, smart cities, smart grids, smart water networks, agriculture, health-care, etc. Of particular interest are applications of DSN to networks where

[1] This paper is thoroughly revised and substantially extended version of our "best paper" awarded conference publication at Sensornet 2017. Title of our conference version is: "Sensornet: A Key Predistribution Scheme for Distributed Sensors using Nets" [1]. Additional sections in this work in comparison to earlier version are Sections 2, 6, 7 and 8 and all their subsections.

security is a premium. For instance, security may be essential for certain sensitive scientific and military networks that are meant for (i) self-healing minefields, (ii) military surveillance, (i) force protection arenas, and so on. Primary tasks of devices of an IoT in any such application are to collect information from their surroundings, process and forward them to other devices. Depending on specific applications, they may be further required to (i) track and/or classify an object, (ii) determine parametric value(s) of a given location, etc. These sensitive tasks for such critical applications create necessity of secure message exchange among the low cost IoT devices.

### 1.1. Type of Cryptosystem: KPS

Constraints in resources of constituent tiny devices of a low cost IoT (like sensors of DSN) make us opt for lightweight symmetric key cryptosystems (SKC) over expensive public key cryptosystems (PKC) while designing security protocols for such networks. SKC require both senders and receivers to possess the same encryption/decryption key before message exchange. Standard online key exchange techniques that involve PKC are avoided due to high cost factor.

Few trivial key distribution techniques are as below. First approach is to assign a single key for the entire network. This method is vulnerable to "single point *fail*ure" (compromise of one sensor reveals this single system key). Second is to think of assigning pairwise distinct SKC keys for every pair of devices. This later strategy overloads the memory of each sensor; $\mathcal{N} - 1$ keys are required to be stored per sensor for a network of size $\mathcal{N}$. This is impractical for large networks (i.e., large value of $\mathcal{N}$ ( $\approx 10^4$, say, or greater). Treating a node (or a few) as Trusted Authority (TA) is risky. This also makes the network prone to "single point *fail*ure" because capture of sensors acting as TA leads to vulnerable systems. Thereby schemes like LEAP [2] are avoided while designing secure key management schemes for DSN.

These facts emphasize the importance of proper employment of an adequate key management scheme. This situation was wittily overcome in 2002 by Eschenauer and Gligor [3] by introducing the concept of *key predistribution* that involves applications of SKC to sensor networks. Any KPS primarily executes:

- Key distribution: Prior to deployment, keys are preloaded into sensors to form their *keyrings or key chains* from the collection of all network keys, called *key pool*. Each system key is marked with a unique identifier *(key id)*. Certain schemes [4] consider *(node id)* as a unique function of all the key ids. These key or, node id are used during key establishment.
- Key establishment: The preloaded keys are established by a two steps process, as below:
  - (i) *Shared key discovery phase* establishes the shared key(s) among the participant nodes. This may be achieved by broadcasting the key ids of all keys contained in the nodes (or node id). On receiving each other's key ids, the sensors tally them to trace their shared key id(s), hence common shared key(s).
  - (ii) *Path key establishment phase* establishes a path key between a pair of nodes that do not share key. This process involves intermediate nodes. Refer to common intersection designs *(μ–CID)* in Section 4.

Depending on whether the above processes are probabilistic or deterministic, such schemes are classified into two types: (a) *random* and (b) *deterministic*. Sections 3.1 and 3.2 present a brief overview of individual type of schemes.

### 1.2. Summary of our Contributions

Observing the significant advantages of deterministic KPS during key management for low cost distributed networks, we set out to propose one such scheme. Our proposal uses net of partial spreads (or, nets) in a finite vector space that have been well studied combinatorically and as such, we name the scheme as *Sensornet*. Later we extend our protocol to a resiliency enhanced version, *Hash Chains Sensornet* (*HC(Sensornet)*). The process eliminates dangerous "smart attacks" (defined below in Section 2).

### 1.3. Paper Organization

Prior to the proposal of our protocols, we introduce in Section 2, all threat models that we consider in our work. Section 3 conducts a brief literature survey on KPS and then presents some preliminary theory related to combinatorial set systems that are required to construct such KPS. Construction of net of partial spreads or nets is reviewed in Section 4 and our scheme *Sensornet* is presented in Section 5. This is followed by proposal of a generic approach that eradicates "smart attacks" in Section 6. A generic Hash Chains based approach of Bechkit et al. [5, 6] that enhances resilience of any KPS is reviewed in Section 7. Our basic *Sensornet* scheme is then extended to *Hash Chains Sensornet* (*HC(Sensornet)*) by applications of these generic approaches. We analyze *Sensornet* and *HC(Sensornet)* in terms of performance metrics in Section 9 and thereby establish our scheme's efficiency in comparison to prominent proposals. We infer that our protocols adhere to the desirable criteria set out in Section 3.4. We summarize our work in Section 10 and state related future research directions in Section 11.

## 2. Threat Models

Passive eavesdroppers have little effect on KPS systems. So, our system's resilience is analyzed against two active adversarial attacks. They are *random node compromise* and *smart attacks*.

*'Random node (compromise) attacks'*, as the name suggests, is the random compromise of nodes by an

adversary. This leads to partial disclosure of key pool ($\mathcal{K}$) of existing devices; thereby restricting the use of links that were secured by these keys.

*"Smart Attack"* [6] is essentially selective or "smart" compromise of (other) nodes that share the same key(s) as a pair of communicating nodes. This attack occurs because the key establishment process requires exchange of *key ids* in (unencrypted) plain text. This reveals the *key sharing graph* that can be beneficial to an adversary to selectively (or smartly) target specific nodes (see Pietro et al. [7] for details).

Most KPS solution exchange unencrypted set of their key ids or a unique function of this set, aka node ids during key establishment. Thus the sets of key ids (or the node ids) of these nodes becomes a public information. So, an attacker can easily compute the shared key ids by 'equating' them (like any node) and successfully launch a "smart attack" [7]. Few prominent examples where this happens are [5, 6, 8-13]. However, there are works [14, 15] that encrypt these secondary node ids or set of key ids before transmissions and thereby seal this attack.

Due to its predictable nature, selective capture of (uncompromised) nodes may have more devastating effects than the random model. This situation is rectified by exchange of encrypted sets of key ids; thereby eliminating "smart attack" (see Section 6).

## 3. A Brief Survey of KPS

This section presents a state–of–the–art survey of prominent KPS. We split survey into three stage: (i) random KPS (RPKS), (ii) deterministic KPS (DKPS), and (iii) advantages of later type over former. Thereby, we justify proposal of our new deterministic KPS adhering to design criteria set out in Section 3.4.

### 3.1. Random Key Predistribution Schemes

First generation KPS rely on random graph theory pioneered by Erdos and Renyi [16] to preload SKC keys into the sensors. Therefore, keyrings are formed randomly. This leads to probabilistic key sharing and establishment. Later is achieved by either broadcast of key ids or challenge and response. Refer to [3, Section 2.1]. Earlier, Blom proposed the first key distribution scheme [17] in public key settings meant for resourceful ad hoc networks. Blom's schemes uses pairs of public private matrices for key distribution. It cannot be applied to resource constraint sensor networks due to its heavy memory requirement to store huge vectors. Several researchers use variants of Blom's schemes to propose both random and deterministic KPS for DSN.

### 3.2. Deterministic Key Predistribution

Deterministic KPS were simultaneously proposed by [9, 10, 13] in 2004. The work [13] combines subset based schemes with existing key distribution schemes such as [17] to obtain multiple key spaces. The scheme of Lee and Stinson [10] uses quadratic equation solving and can be viewed as a scalable extension of their later proposal [11] that uses Transversal Design *(TD(k,p))*. This later work further summarizes the necessary conditions for a combinatorial design to yield a deterministic KPS. The work [9] exploits combinatorial designs like symmetric Balanced Incomplete Block Designs (BIBD), generalized quadrangles and projective planes [11, 12, 18]. Certain KPS [4] exploit special Algebraic structures like Reed Solomon code based KPS. However, these protocols permit alternate combinatorial descriptions that have been well studied in [14, 15, 18].

### 3.3. Advantages of DKPS over RKPS

Deterministic schemes have advantages over their random counterparts. For instance, a desired property of a randomized scheme may occur only with a certain probability whereas they can be proven to hold in a deterministic scheme [11, 12, 18]. This led to proposals of numerous deterministic KPS using various combinatorial tricks. Further the predictable nature of these combinatorial structures has been efficiently exploited to address design weaknesses of certain prominent KPS. For instance, the schemes [14, 15] addresses the connectivity and resilience aspect of [4, 11] by a deterministic design specific approach.

Contrary to these observations, Ruj and Pal [19] state that random graph models are well suited for 'scalability' and 'resilience'. Thereby, they justify their proposals of random graph based preferential attachment models with degree bounds. They design networks using their model. Their designs suffer from highly skewed load distribution, poor connectivity and resiliency; and so, are inappropriate for applications to (distributed) IoT.

In fact, sensitive IoT applications require protocols to yield equal distribution of tasks among peers. Moreover, to reduce hops and hence potential risks from node capture, it is more important to have connected networks that cannot be guaranteed by random schemes. So we opt for deterministic protocols for security applications in low cost IoT networks that assure predictable (high) connectivity; despite them having restricted scaling operations. This is a major area of study for most (deterministic) KPS proposals, including ours (recalled in Section 11).

Structure of the combinatorial objects used to design deterministic KPS cannot directly model networks of any specified size $N$. Usually, such structures result in designs having a specific pattern in the number of resultant blocks, viz. a prime power etc. Since $N$ can be any number, a standard strategy is to consider the least prime power that is greater than the network size (i.e., $p^z \geq \mathcal{N}$). Then $\mathcal{N}$ subset are randomly selected to form the key rings of the network nodes. Bose et al. [8] speculate that random removal of blocks may have a disadvantageous effect on the underlying design's properties and hence become an issue of concern. Fortunately, this claim of Bose et al. [8] has been successfully challenged by Henry et al. [20]. Through practical experiments, they establish that random removal of key rings of a combinatorial KPS has negligible effect with overwhelming probability. This work reestablishes the importance of combinatorial schemes. Further these excess block may be useful in (restricted) scaling of a network.

### 3.4. Desirable Design Criteria

Devices of a low cost IoT (example: sensors of a DSN) are highly prone to damage and/or physical capture. This is a crucial consideration during the design of an (energy) efficient KPS. Prime objectives of any KPS is to ensure that the resulting network:

1. Has less number of keys per node, i.e., sizes of individual keyrings are less;
2. Have large *node support*, i.e., support large number of network nodes;
3. Has good (ideally full secure) *connectivity*. Secure connectivity (or, simply *connectivity*) is the ratio of number of (secure) links in eventual network to all possible links. A pair of nodes are said to be connected by a (secure) link if there exists at least one secret key between them;
4. is *resilient* against various types of adversarial attacks. A prevailing method adopted in most existing works [9-15,18] is to show that a standard resiliency coefficient *fail(t)* is minimized. We follow suit. The quantifier *fail(t)* measures the ratio of links broken after compromise of *t* sensors to the total number of links in the remaining network. Formally, *fail(t)* = $u_t/b_t$, where $b_t$ is the number of links broken when *t* nodes are compromised and $u_t$ is the total number of links in the remaining network of uncompromised nodes.

Ideally a KPS should have small keyrings, large network support with appreciable resiliency, scalability and (secure) connectivity. However, prominent researches prove the impossibility of construction of a *perfect KPS* that meets all these criteria [11, 18]. This motivates proposal of designs that are robust for specific purpose. In the same spirit, we propose our schemes, *Sensornet* (in Section 5) and its resiliency enhanced version, *Hash Chains Sensornet (HC(Sensornet))* in Section 8 that are derived from net of partial spreads (or nets). Our schemes have a good balance of these combinatorial properties and so, are useful to design a deterministic KPS.

### 3.5. Deteriorated Resilience of KPS

Any KPS assigns multiple sensors to a given key. For a deterministic scheme, this value is the (regular) degree *r* of the design (refer to Section 4 below). Therefore, compromise of a node exposes partial key rings of uncompromised ones that affect their secure communication. This makes the resultant system vulnerable to various node capture attacks; thereby affecting the system's *resilience*. Several researches develop tricks to reduce the effect of this attack. We discuss a prominent effort in a Section 7 while extending our *Sensornet* to *Hash Chains Sensornet*.

## 4. Preliminaries

This section introduces definitions and notations that are required to describe our scheme, *Sensornet*.

### 4.1. Combinatorial Set Systems and KPS

Construction of generic deterministic KPS by the use of a combinatorial designs is presented in the paper [11]. A unified treatment of prominent combinatorial designs in terms of partially balanced *t – design* is present in [18]. We present below the basic design theoretic concepts:

Let $\mathcal{X}$ be a finite set. Elements of $\mathcal{X}$ are called varieties. Each subset of $\mathcal{X}$ is termed as a block. Consider $\mathcal{A}$ to be a collection of blocks of $\mathcal{X}$. Then $(\mathcal{X},\mathcal{A})$ is said to be a *set system or, a design*. $(\mathcal{X},\mathcal{A})$ is regular of degree *r* if each point is contained in *r* blocks.

$(\mathcal{X},\mathcal{A})$ is uniform (of rank *k*) if all blocks have the same size, say *k*. A design $(\mathcal{X},\mathcal{A})$ is said to form a *(v,b,r,k)–design* if: $|\mathcal{X}| = v$, $|\mathcal{A}| = b$; it is regular of degree *r* and of uniform of rank *k*. A *(v,b,r,k)–design* forms a *(v,b,r,k)–configuration* if any arbitrary pair of blocks intersect in *at most* one point. Moreover, if any pairs of varieties occur in exactly one block, then a *(v,b,r,k)–design* forms a *(v,b,r,k)–BIBD* (Balanced Incomplete Block Designs). They can be used to construct various KPS [11] by mapping:

a) *v* varieties of $\mathcal{X}$ to the set of keys in the scheme (i.e., $|\mathcal{X}|$ = *v* =*size of key pool*),
b) *b* to the number of network nodes in the system (*:=network size*),
c) *k* to the number of keys per node (*:=size of key rings*), and
d) *r* to the number of nodes that share a given key (*:=degree of the resultant KPS*).

The target is to construct KPS with identical burden on each sensor. This leads to opting for design with uniform rank (*k*) and regular degree (*r*), so that every *key ring* is of equal size (*k*) and same number of nodes (*r*) share each key for the resultant network.

*Block graph* $G_{\mathcal{A}}$ of the set design $(\mathcal{X},\mathcal{A})$ is defined with the vertex set $\mathcal{A}$ and edge set $E_{\mathcal{A}} = \{(A, B) : A, B \in \mathcal{A}$ *and* $A \cap B \neq \emptyset\}$. If the set design is regular of degree *r* and uniform of rank *k*, then the block graph $G_{\mathcal{A}}$ is *k(r 1)–regular*. A *(v,b,r,k)−configuration* $(\mathcal{X},\mathcal{A})$ is said to form a *μ−common intersection design* (*μ−CID*) in case for every pairwise empty intersection of blocks, there exists *μ* other blocks that share common keys with these blocks. It is important to construct designs that maximize the value of *μ*.

### 4.2. Net of Partial Spread or Nets

Let $IF_p$ be the finite field on $p$ elements where $p$ is a prime. Denote by $V_n(= IF_p^n)$ to be the vector space of dimension $n$ over the field $IF_p$ with zero vector **0**. Since the finite field $IF_p^n$ is a vector space over $IF_p$ and is isomorphic to $IF_{p^n}$ [21], we interchange the notation as per its suitability. This isomorphism mapping can be considered as a mapping from a basis set of $IF_p^n$ to a basis set of $IF_{p^n}$. We consider $n=2m$ to be an even integer in this work.

A partial spread Σ of order s in $V_n$ is a set of pairwise supplementary *m–dimensional* subspaces $E_1,E_2,\ldots,E_s$ of $V_n$ i.e., $E_i \cap E_j = \{0\}$, $1 \le i < j \le s$. A partial spread $\Sigma$ forms a spread if $\bigcup_{i=1}^{s} E_i = V_n$. It is known that a spread of $V_n$ exists since $m$ divides $n$ [22], then $|\Sigma|=p^m+1$. Therefore, from a given spread Σ each of the choices of s members of $\Sigma$ provides a partial spread of $V_n$. Note that a partial spread might not be a subset of a spread (refer to Eisfeld and Storme [23]). An interested reader can refer to the books [24, 25].

Let $E$ be a subspace of the vector space $V_n$. A coset of $E$ in $V_n$ is of the form $\alpha+E = \{\alpha+ v: v \in V_n\}$ for an $\alpha \in V_n$. The set of cosets makes a disjoint partition of $V_n$. The element α is called a coset representative of the coset $\alpha+E$. Since $E$ is an additive group, any element from the coset $\alpha+E$ can be a coset representative. Given a partial spread $\Sigma = \{E_1,E_2,\ldots,E_s\}$ in $V_n$, let $E_i$ be a set of coset representatives of subspace $E_i$ for $1\le i \le s$. Then the set $A = \{\alpha+ E_i : \alpha \in E_i, 1\le i \le s \}$ i.e., set of all cosets of subspaces $E_i$, $1\le i \le s$ forms a *net* in $V_n$. An interested reader is referred to the book by Johnson et al. [25] for a detailed study of nets.

### 4.3. Examples of Partial Spreads

There are numerous constructions of spreads and partial spreads that can be found in the literature [25]. Now we present a few spreads S in $IF^n_p$, where $p$ is a prime. For a given s, any $\Sigma \subseteq S$ such that $|\Sigma| = s$ forms a partial spread of order $s$. By Theorem 1 (see below in Section 5), this partial spread yields a KPS.

**Spread I**: This is a classic example of a spread from the additive group of the finite field $IF_p^m$. Since $n=2m$, $IF_p^m$ is a subspace of $IF_p^n$ with respect to a basis. Let $\{\alpha_i : 1 \le i \le p^m+1\}$ be a set of coset representative of the cosets of the subgroup $IF_p^m$ in the multiplicative group $IF^n_p$. Then the set $S_I = \{S_i = \alpha_i IF_p^m, 1 \le I \le p^m+1\}$ is a spread in $IF_p^m$.

**Spread II:** This example of spread is represented in bivariate form (see [26]). For each $\alpha \in IF_p^m$, define a subspace $U_\alpha$ of $IF^m_p \times IF^m_p$ by $U_\alpha = \{(\alpha u,u) |\ u \in IF_p^m\}$ and for sake of the consistency $U_\infty = \{(u,0) | u \in IF_p^m\}$. The set $S_{II} = \{U\alpha : \alpha \in IF_p^m\} \cup U_\infty$ constitute a spread in $IF_p^m \times IF_p^m \simeq IF_p^n$.

**Spread III**: This example of spread is generated from pre-quasifield, which is defined as following. A system $Q=(V,+,\circ)$, with finite $|V|$, is a pre-quasifield if the following axioms holds:

1. $(V, +)$ is an abelian group, with identity **0**.
2. $(V^*,\circ)$ is a quasigroup where $V^* = V\setminus\{0\}$. That is, for any $a \in V^*$, the left multiplication operator $a \circ x$ and the right multiplication operator $x \circ a$ are both bijective from $V^*$ to $V^*$.
3. $\forall x, y, z \in V$, $(x + y) \circ z = x \circ z + y \circ z$.
4. $x \circ \mathbf{0} = \mathbf{0}$, $\forall x \in V$.

Now assuming $(IF_p^m,+,\circ)$ is a pre-quasifield, set $E_a = \{(x,a \circ x): x \in IF_p^m\}$ for any $a \in IF_p^m$ and $E_\infty = \{(0,x) : x \in IF_p^m \}$. Then it can be checked that $S_{III} = \{E_a : a \in IF_p^m \cup \{\infty\}\}$ is a spread in $IF_p^m \times IF_p^m$ [25]. Many pre-quasifields are available in literature. Refer to [27] for three types of pre-quasifields on set $IF_2^m$ and [28] for a pre-quasifield on set $IF_p^m$.

**Example 1 (of NETS)**: Here, we present a simple KPS from the spread of type $S_I$. Take $V_n = IF_9 = IF_3[x]/(x^2+1)$. Consider the subspace $IF_3 = \{0,1,2\}$ and $\{1,x,x+1,x+2\}$ as a set of coset representatives of $IF_3^*$ in $IF_9$. Then $S_I = \{\{0,1,2\},\{0,x,2x\},\{0,x+1,2x+2\},\{0, x+2,2x+1\}\}$ is a spread in $IF_9$. Consider a partial spread $\Sigma = \{E_1 = \{0,1,2\}, E_2 = \{0, x,2x\}\}$ where $\overline{E_1} = E_2$ and $\overline{E_2} = E_1$. So, by Theorem 1, the set $\mathcal{X}$= $IF_9$ and the net $\mathcal{A}$={{0,1,2},{x,x+1,x+2},{2x,2x+1, 2x+2}, {0,x,2x},{1,x+1,2x+1},{2,x+2,2x+2}} forms a KPS **($\mathcal{X}$,$\mathcal{A}$)**. The block graph of **($\mathcal{X}$,$\mathcal{A}$)** is the $K_{3,3}$.

## 5. Sensornet

*Sensornet* is a class of KPS for distributed sensor networks. The design of *Sensornet* is based on partial spread or nets and is a consequence of the forthcoming set design and Theorem 1.

Given a partial spread $S = \{E_1,E_2,\ldots, E_s\}$ in $V_n$, let $\overline{E_i}$ be a supplementary subspace of $E_i$ in $V_n$ (their direct sum $E_i \oplus \overline{E_i} = V_n$ and $E_i \cap \overline{E_i} = \{\mathbf{0}\}$). One can check that $\overline{E_i}$ is a set of coset representatives of $E_i$ for $1 \le i \le s$. Note that the subspaces $E_i$'s in a partial spread are pairwise supplementary. So, any $E_j$, $j\neq i$ can be chosen as $\overline{E_i}$. Consider a set system **($\mathcal{X}$,$\mathcal{A}$)** with $\mathcal{X}$=$V_n$ and the set of blocks, $\mathcal{A}$= $\{\alpha+E_i: \alpha \in E_i, 1\le i \le s\}$, which is a net in $V_n$. Then we have the results below:

**Theorem 1**. Given any partial spread Σ, the set design **($\mathcal{X}$,$\mathcal{A}$)** is $\mu(p^n,sp^m,s,p^m)$–CID for $\mu=(s-1)p^m$.

*Proof.* Here $v=|X|=p^n$. Consider two blocks $\alpha+ Ei$ and $\beta+Ej$, then we have the following cases:

1. If $i = j$, then
   a. $\alpha+E_i = \beta+E_j$ if $\alpha=\beta$ or,
   b. $(\alpha+Ei) \cap (\beta+Ej) = \emptyset$ if $\alpha \neq \beta$.
2. If $i \neq j$, then we shall show that: $|(\alpha+Ei) \cap (\beta+Ej)|=1$.

$E_i$ and $E_j$ are supplementary to each other. So, the element $\alpha-\beta \in V_n$ can be uniquely expressed as $-u+v$ where $u \in E_i$ and $v \in E_j$. That is, $\alpha-\beta = -u+v \Rightarrow \alpha+u=\beta+v$ is a unique element in $(\alpha+Ei) \cap (\beta+Ej)$.

So, the number of blocks i.e., the number of cosets is $b=sp^m$ and each block contains $k=p^m$ elements. Given a subspace $E_i$, $i \in \{1,2,\ldots,s\}$, each element $u \in V_n$ belongs to exactly one coset of $Ei$. So, each $u \in V_n$

belongs to exactly ***s*** many blocks in A. The set design ***(𝒳,𝒜)*** is regular with r = s. Every two distinct blocks intersect each other by at most one element which implies that ***(𝒳,𝒜)*** is a $(p^n,sp^m,s,p^m)$– configuration.

We observe that two blocks $\alpha+E_i$ and $\beta+E_j$ do not intersect if $i = j$ and $\alpha \neq \beta$, i.e., both are distinct cosets of same subspace $E_i$. For the case of non–intersecting blocks $\alpha+E_i$ and $\beta+E_i$, $\alpha \neq \beta$, both blocks intersect all other blocks of the form $\gamma+E_j$ where $j \neq i$. Since there are$\mu=(s–1)p^m$ such blocks $\gamma+E_j$ in A, ***(𝒳,𝒜)***) is a $(s-1)p^m(p^n,sp^m,s,p^m)$–*CID*.

Here, the set of block ***𝒜*** of the scheme ***(𝒳, 𝒜)*** forms a net in a vector space. It can be checked then the block graph of ***(𝒳,𝒜)*** is a strongly regular graph with parameters $(n=sp^m,\ r=(s-1)p^m,\ \lambda=(s-2)p^m,\ \mu=(s-1)p^m)$. Moreover, the block graph is a complete s–partite graph. In the study of finite geometry, the varieties together with the blocks (i.e., cosets) form the points and lines of an affine plane. Since two non–parallel lines (i.e., $\alpha+E_i$ and $\beta+E_j$ for i ≠j) intersects at one point, this set of cosets is called a net. We name the scheme as *Sensornet*.

## 6. Eradication of Smart Attacks

*Sensornet*, like all combinatorial KPS is susceptible to "smart attacks" (formalized in Pietro et al. [7]). We briefed this attack in Section 2. This section devises a generic method to reduce it to random node compromise attacks.

Our approach requires locally or group-wise random of nodes with one (ordinary) node in each group acting as its lead. Each group member is assigned exact one extra key for secure communication with its lead. While these leads stores an additional $g + l – 1$ keys for its $g$ children and $l–1$ co-leads. Therefore $\mathcal{N}(= b) = gl$ and so, $g = l =\sqrt{\mathcal{N}}$ is an optimized value. Association of keys and primary addresses (IP/MAC) can be stored in a table and preloaded in these group leads, so that *key establishment is not required for these extra keys*. These extra keys are thus be available to secure communication between these leads and their children. We use these extra keys for secure exchange of node ids during initial or subsequent (seldom) key establishment phases during a network's life time. Recall that these node ids are unique function of the set key ids of the preloaded keys for a given node and so are *secondary* ids. They are consequences of the combinatorial design construction meant for a deterministic KPS.

We propose a modified key establishment protocol as below. Sensors securely transmit their own node ids to their leads by the extra key. The leads circulates these node ids among themselves (in communication range). Node ids being linear functions are easy to "equate" and hence their collective computation burden is not abrupt. Moreover leads can apply a distributed algorithm to reduce the mutual burden on these leads. A table of shard key-primary address (IP/MAC) is formed during run time for each node. This tabular representation is returned to individual sensors and destroyed instantaneously. Node ids of children and co-leads are also flushed. So these leads retain data required only for their own conversations. Their children also gets only those information that concerns themselves.

Being orders less in number ($=\sqrt{b}$), we assume non compromise of the leads during the (short lived) key establishment phase(s). In fact, all existing works assume absolute trust on all system devices that includes their non-compromise during this phase. We are less *restrictive* and allow (random) "children compromise attacks" during key establishment. Refer to Section 2. Our construction does not reveal key sharing graph for "children compromise attacks" during key establishment as they do not ever possess other nodes' ids. Therefore, it is reasonable to assume concealment of the key sharing graph during key establishment. Post key establishment capture of any children does not reveal the key sharing graph due to same reason. Moreover, destruction of relevant information from the leads means that their capture at a late stage does reveal these node ids. Therefore, compromise of any sensor (lead or children) during (later) run time of the system does not reveal node ids, though all information about keys of that sensors are exposed. Thus cycles of keys that are not contained in exposed nodes are not disclosed. In effect "smart attack" cannot be launched.

The remedial construction may seem to convert a distributed system into a hierarchical one. Well, the hierarchy is required only during key establishment. We do not advocate use of these extra keys of the leads and/or their children for message exchange later on. (This is because "single point attack" on the leads may reveal the message (exchange) of their children.) Once a key establishment phase is over, the leads have no role in their children's conversations; these leads are not gateways for future conversations of their children. Therefore, our seemingly hierarchical construction retains its distributed flavor for the entire life time of the network barring the short lived key establishment phases(s). This is contrary to an inherent hierarchy in most key management protocols like [14, 15]. [2]

## 7. Lightweight Resilience Improvement

Degree of any KPS lead to a security deterioration due to node capture attacks (defined to Section 2). A

[2] We do not advocate use of the extra keys in leads and their children for message exchange after key establishment phases. Our system remains distributed later; as opposed to an inherent hierarchy in key management protocols [14, 15].

network's resilience against such attacks is of vital importance. Many work aim to improve this aspect.

One such approach, due to Bechkit et al. [5, 6], presents a cute application of (recursive) hash function in a generic fashion. Their Hash Chains scheme $HC(x)$ successfully improves resilience of any KPS $x$ without affecting other parameters and is briefed below:

- node ids of nodes vary from 0 to $b - 1$ where $b$ is the number of blocks of the underlying combinatorial design. Observe that $b \approx N$.
- given a key $K$, let us inductively define $H^i(K) := H(H^{i-1}(K))$. That is, $H^i(K)$ denotes the $i$ times use of the hash function $H$ on key $K$ for $i \in Z_+$.
- due to resource constraints, let the maximum number of times that we can repeat this (recursive) hash function computation in any sensor be $N$-$1$, ($1 \leq N \leq b \approx N$).
- node ids are used to discriminate initial preloaded KPS keys as described below:
  (i) instead of original keys, $K$, they preloaded a node with id $i$ with the key $H^{(i \bmod N)}(K)$, for each key $K$ in the $i$-th node $(0 \leq i < N)$;
  (ii) thus, two nodes with id $i$ and $j$ that shared the same key $K$ in original KPS $x$ end up with $H^{(i \bmod N)}(K)$ and $H^{(j \bmod N)}(K)$;
  (iii) if $(j - i \bmod N) > 0$ then node $i$ calculates $H^{(j-I \bmod N)}(k^i)$. *Preimage resistant property* of the (cryptographic) hash function $H$ *implies* node $j$ cannot find $H^{(i \bmod N)}(K)$.
- key establishment of these nodes uses set of key ids or node ids and is same as the original KPS.
- nodes $i, j$ establishes their shared key, $SK= H^l(K)$, where $l = \max(i \bmod N; j \bmod N)$. This $SK$ can be computed at either end, in case they possess either key $H^{(i \bmod N)}(K)$ or $H^{(j \bmod N)}(K)$. Node $a$ computes

$$H^{l-a}(H_1{}^a(K)), \text{ for } a = i, j:$$

- capture of $i$-th node exposes all its keys $H^{(i \bmod N)}(K)$ to the adversary, who:
  (i) cannot establish links with the nodes that possess the keys $H^{(j \bmod N)}(K)$, for any key $K$ in the $i$-th node and $j>i$ (mod $N$);
  (ii) can establish link with nodes that possess a key $H^{(j \bmod N)}(K)$ for $j < i$ (mod $N$).
  (iii) resilience dip is 30% for node capture.

So, in this Hash Chains based schemes $HC(x)$, connectivity, storage overhead and communication overhead remains same as the original scheme $x$. Application of our generic "smart attack" removal method (devised in Section 6) reduces any "smart attack" to random node compromise attack. Therefore the combination of the two approaches improves the original system's resilience by 50% against any node compromise attack (see Theorem 3). This combined technique is suitably adapted to enhance resilience of our *Sensornet* scheme and yield a new protocol: *Hash Chains Sensornet* in Section 8.[3]

## 8. *Hash Chains Sensornet (HC(Sensornet))*

Prototyped application of Bechkit et al.'s idea to our *Sensornet* scheme produces a resilience enhanced scheme that we term as *Hash Chains Sensornet*. Therefore considering $x$ = *Sensornet*, the shared key between the node $i, j$ with ($j \bmod N$) > ($i \bmod N$) is computed as described below.

Let $K$ is the distributed keys between the $i^{\text{th}}$ and $j^{\text{th}}$ nodes for $j > i \bmod N$. Then the nodes $i, j$ compute shared secret key as

$$SK = H^j(K):$$

$i$-th node computes this shared secret key as

$$SK = H^l(H^i(K)$$

where $l = (j–i \bmod N) > 0$ So, in this Hash Chains based schemes $HC(x)$, connectivity (range), storage overhead and communication overhead remains same as the original scheme $x$.

Key establishment process is similar to *Sensornet* with modification. We store the basis vectors in leads, say $E_i$ ($\alpha$= 0 is natural choice). So its children $\alpha+E_i$ contains only $\alpha$*'s*. These $\alpha$*'s* are exchanged securely using the extra keys. Therefore the leads $E_i$'s gets the entire node ids $(\alpha; \beta_i^1, \beta_i^2, \beta_i^3,\ldots,\beta_i^m)$ of each of their children $\alpha+E_i$, $\alpha=1,2\ldots m$. So in our *HC(Sensornet)* these leads can only perform the three steps given in Section 9.1 (nodes do not have their own id). Of course our *Sensornet* could have been combined with this unique "smart attack" removal technique. So, key establishment of both these combined schemes require less ($O(\log_p N)$) data transfer; computation complexity is same as original *Sensornet* scheme that was presented in Section 5.

## 9. Analysis of Our Protocols

In this section we compute the values of some important metrics involved in the protocols we have proposed, i.e., *Sensornet* and *HC(Sensornet)*.

### 9.1. Time ($T_k$) and Space Complexities ($M_k$) for Key Establishment

For the key establishment between two nodes, the nodes need to discover a common key stored between them. For this purpose, the nodes need to broadcast some data, which is required to trace the common key between two nodes. Since the sensor nodes have low memory and computation power, data and time requirement for key establishment are two very important factors to design a KPS. In this subsection we discuss the process of key establishment between two nodes and associated time and data requirement of the process. In case of path key establishment (refer to Section 1.1), both the concerned nodes have to find a common neighbor with whom they discover their shared key and establish connection via this neighbor.

[3] We denote a keys by K (CAPITAL LETTER) and its id by k (small letter) throughout this work. A full domain hash function is a suitable candidate for our work (like [5, 6]).

Denote by $T_k$ and $M_k$ to be the time and memory complexity functions for the key establishment.

The blocks of *Sensornet* forms a net, i.e., they are affine spaces. So, nodes can be identified by their basis vectors and key establishment is done using the node id. When nodes $\alpha+E_i$ and $\beta+E_j$ need key establishment between them, they follow the process below:

**Step 1**: The nodes $\alpha+E_i$ and b + $E_j$ compare the last *m* (basis) vectors in their node id. If they are same then follow Step 3 else follow Step 2.

**Step 2**: Here, we have $E_i \neq E_j$ i.e., so they share a common key. Let this key be $\alpha+u = \beta+v$, where $u \in E_i$ and $v \in E_j$. Now we need to find *u* and *v* in terms of the basis vectors of $E_i$ and $E_j$ respectively. Here, $\alpha-\beta$=v–u $\in V_n$. Since $E_i$ and $E_j$ are supplementary subspaces in $V_n$, $\alpha-\beta$ can be uniquely expressed as a linear combination of the basis vectors of $E_i$ and $E_j$. This leads to the fact that this common key is in both $E_i$ and $E_j$. The time complexity in this step is the time complexity to express $\alpha-\beta$ in terms of the basis vectors in a basis i.e., $O(n^3)$. Same is true for *HC(Sensornet).*

**Step 3**: In this case, $E_i = E_j$ i.e., they do not share a common key. So, they have to establish connection through another node with whom they share a key, individually. That is, they have to find a node $\gamma+E_k$ where $k \neq i$. The probability of finding such a node using a random pick up is s–1/s which is very high. Since both $\alpha+E_i$ and $\beta+E_j$ share a key with $\gamma+E_k$, each one does the same process described in Step 2 with $\gamma+E_k$ to discover their common key. After that $\alpha+E_i$ and $\beta+E_j$ can establish connection through $\gamma+E_k$. So, the time complexity in this case is $O(n^3)$.

So, each node spends $M_k = (m+1)*n*(\log_2 p) = O(n^2)$ bits of data for broadcasting of their identity and the time complexity to discover the common key(s) is $T_k=O(n^3)$. Note that in *Sensornet* and so, in *HC(Sensornet),* nodes broadcast only node id, that are only $O(n^2)$ bits instead of $O(rp^{n/2})$ many (all) key ids as broadcast by other prominent schemes.

### 9.2. Key-node Ratio (σ)

The key-node ratio is defined as $\sigma = k/b$. This ratio provides idea about the storage requirement of the scheme at each node with respect to the total number of nodes. With this metric we can compare the storage requirement of the schemes from different designs. It is desirable for this ratio σ to be as small as possible as lesser amount of memory required for key storage at each node. In both our schemes, *Sensornet* and *HC(Sensornet)*, value of key-node ratio is

$$\sigma = \frac{p^m}{sp^m} = \frac{1}{s} = \frac{1}{\Sigma}.$$

If the size of partial spread is larger, then the storage requirement to store keys in *Sensornet* is lesser.

### 9.3. Resiliency *(fail(t))*

Schemes should be equipped to perform against adversarial attacks. To this end, the standard resiliency metric *fail(t)* need to be minimized. This is prevalent method adopted by most existing works [4, 9,-12, 14, 15]. The quantifier *fail(t)* measures the probability that a random link between two sensor nodes is broken due to the compromise of *t* other random nodes. Formally, *fail(t)* = $b_t/u_t$ where $b_t$ is the number of links broken when *t* nodes are compromised and $u_t$ is the total number of links among uncompromised nodes of remaining network. Theorem 2 is due to Lee and Stinson ([11, Section VIII]) provides the formula to compute *fail(t)* for any *(v, b, r, k, 1)–configuration*.

**Theorem 2.** Theorem 2. *For any (v, b, r, k, 1)–configuration, the value of the metric fail(t) on random compromise of t nodes is given by:*

$$fail(t) = 1 - \left(\frac{b-r}{b-2}\right)^t \quad (1)$$

Corollaries 1 is an immediate outcome of substituting the values of b and r in Equation 1, for the scheme *Sensornet*.

**Corollary 1.** The value of the resilience *fail(t)* for the set design ***(X,A)*** of the scheme *Sensornet*, which is a *($p^n$,$sp^m$,s,$p^m$)−configuration* is

$$fail(t) = 1 - \left(\frac{sp^m - s}{sp^m - 2}\right)^t \quad (2)$$

Clearly, the metric *fail(1)* = $O(p^{-m})$ i.e., if a node *N* is compromised, then the probability that a link (which is not incident with *N*) is $O(p^{m})$. Here, the size of the partial spread has no significant effect on *fail*(1). For example, with $p$ = 2, $n$ = 10 (i.e., network size= $2^{10}$ > 1000), the value of *fail*(1) = 0·03.

**Theorem 3.** Value of our resilience metric fail(t) for our HC(Sensornet) design is given by:

$$fail(t) = \frac{1}{2}\left(1 - \left(\frac{sp^m - s}{sp^m - 2}\right)^t\right) \quad (3)$$

*In particular, fail(1) =*

$$fail(1) = \frac{1}{2}\left(\frac{s-2}{sp^m - 2}\right) \approx \frac{1}{2\times p^m} \quad (4)$$

*Proof.* The result follows the observation made in point 6(*c*) during the brief of Bekhkit at al. along with Theorem 2. Of course our *Sensornet* and *HC(Sensornet)* schemes are assumed to be blessed with the generic "smart attack" removal approach. (Lee and Stinson, 2005) like most KPS assume only random node compromise attack in their analysis. Therefore, we rectify this half analyzed situation.

### 9.4. Connectivity (*ρ*)

We say two blocks in a set system are connected by d links (or, are at a distance *e*) if the shortest path between them in the block graph includes 'e' edges. Hence, we define the metric connectivity (or, connection probability) $\rho_e$ of the network to be the probability that two nodes (placed in physical neighborhood) are connected by $e$ links for a positive integer '$e$'.

Observe that the value of d for a $\mu$ CID with $\mu > 1$ is either 1 (if they share a key) or 2 (if they do not share a key). The formula for $\rho_1$ and $\rho_2$ are provided in (Lee and Stinson, 2005, Section VI), which are being formally restated in the following theorem. Let $\tau$ denote the number of nodes in the intersection of the physical neighborhood of two given nodes.

**Theorem 4.** *The value of the matric connectivities of a $\mu(v,b,r,k)-CID$ are: (i)* $\rho_1 = k \times \left(\frac{r-1}{b-1}\right)$; and

*(ii)* $\rho_2 = (1 - \rho_1) \times \left(1 - \left(\frac{b-\mu-2}{b-2}\right)^{\tau}\right)$.

The following corollary is an immediate outcome for our scheme by substituting the values of $b$, $r$, $k$ and $\lambda$ in Theorem 4.

**Corollary 2.** *The value of the connectivity metrics for the set system $(X,A)$, which is a $(s-1)p^m-(pn, spm, s, pm)-CID$ are* $\rho_1 \approx 1 - \frac{1}{s}$ and $\rho_2 = \frac{s^{\tau}-1}{s^{\tau+1}}$.

The metric $\rho_1 \approx 1 - \frac{1}{s}$, i.e., connectivity increases if the size of spread increases. Here, the size of base field (i.e., the value of the prime $p$) has no significant effect on (direct) connectivity. As an example, if $n = 10$, $p = 2$ (i.e., there are $2^{10} \approx 1000$ many nodes) and s=25 then the value of $\rho_1 = 1-2^{-5}$.

## 9.5. Comparative Study

This section presents a comparative study of *our schemes (Sensornet* and *HC(Sensornet))* with existing works with respect to connectivity, resilience and network scaling. Performance of our schemes with respect to other prominent metric like storage, etc. has been discussed in previous section.

Schemes with high connectivity *(i.e., $\rho_1$)* and resiliency as small as possible are preferred. Unfortunately, both these metrics are inversely related to each other. So, it is a fundamental problem of trading off connectivity verses resiliency. The works [18, 29] considers the ratio $\rho = \frac{\rho_1}{fail(1)}$ for comparison of several combinatorial designs. Therefore, the larger $\rho$ value confirms higher connectivity and lower resiliency. It is desirable that the ratio $\rho$ be as large as possible for the basic combinatorial designs. If necessary, resilience improvement tricks like Bechkit et al. [5, 6] can applied like we did to construct *HC(Sensornet)* from our basic scheme, *Sensornet*.

There have been several proposals for deterministic key predistribution schemes for wireless sensor networks based on various types of combinatorial structures such as designs and codes. The paper [18] proposes a general framework by unifying those structures into a new design, termed as "*partially balanced t-designs (PBtD)*". Although, our scheme *Sensornet* falls into $2-(v, k, \lambda_0=b, \lambda_1=r)-PBtD$ as a configuration, the generalization does not consider $\mu-CID$s. Hence, being a $\mu-CID$, *Sensornet* does not classify as PBtD by their description [18]. There are few comparison tables of different schemes are provided in [18]. In the following, we take data of *TD(t,k,Q)* with intersection threshold $\eta = 1$ from the paper [18] along with other designs to compare with the scheme *Sensornet*.

Let consider the number of nodes in all the compared scheme is N. Now we shall compare the asymptotic behavior of metrics $\rho_1$, *fail(1)* and the ratio $\rho$. The comparison is displayed in Table 1.

**Table 1.** Comparison of asymptotic behavior of different schemes. Refer to Remark 1 for discussions about the parameter $\boldsymbol{\rho}$.

| **Schemes** | **Number of nodes ($\mathcal{N}$)** | $\rho_1$ | ***fail(1)*** | $\rho = \frac{\rho_1}{fail(1)}$ |
|---|---|---|---|---|
| (Sensornet)<br>(devised in Section 5) | $\mathcal{N} = sp^m$ | $1 - \frac{1}{s}$ | $p^{-m} = \mathcal{N}^{-1/2}$ | $\left(1 - \frac{1}{s}\right)\mathcal{N}^{1/2}$ |
| *HC(Sensornet)*<br>(devised in Section 8) | $\mathcal{N} = sp^m$ | $1 - \frac{1}{s}$ | $p^{-m} = \mathcal{N}^{-1/2}$ | $\left(1 - \frac{1}{s}\right)\mathcal{N}^{1/2}$ |
| *TD(2; k; q); k = cq;* [18] | $\mathcal{N} = q^2$ | $C$ | $q^{-1} = \mathcal{N}^{-1/2}$ | $c\mathcal{N}^{1/2}$ |
| *TD*(3; *k; q*); *k = cq; c* < 1 | $\mathcal{N} = q^3$ | $\frac{c(2-c)}{2}$ | $\frac{2(1-c)}{2-c}\mathcal{N}^{-1/3}$ | $\frac{c(2-c)^2}{4(1-c)}\mathcal{N}^{1/3}$ |
| *TD(3; k; q); k = q* | $\mathcal{N} = q^3$ | $\frac{1}{2}$ | $5\mathcal{N}^{-2/3}$ | $\frac{1}{10}\mathcal{N}^{2/3}$ |
| *TD(4; k; q); k = cq; c* > 1 | $\mathcal{N}=q^4$ | $\frac{c(c^2-3c+6)}{6}$ | $\frac{3(c^2-2c+2)}{c^2-3c+6}\mathcal{N}^{-1/4}$ | $\frac{c(c^2-3c+6)^2}{18(c^2-2c+2)}\mathcal{N}^{1/4}$ |
| Symmetric BIBD [9] | $\mathcal{N}=q^2+q^2+1$ | 1 | $\mathcal{N}^{-1/2}$ | $\mathcal{N}^{1/2}$ |
| RS code based [4] | $\mathcal{N}=q^2$ | $\frac{q-1}{q+1}$ | $\mathcal{N}^{-1/2}$ | $\mathcal{N}^{1/2}$ |
| MB designs [15] over TD(k,q) KPS [11, 12] or RS code KPS [4] | $\mathcal{N} = \frac{q^2}{2}$ | 1 | $(2\mathcal{N})^{-1/2}$ | $(2\mathcal{N})^{1/2}$ |

**Remark 1.** For parity with the existing works that we consider in in Table 1, stated *r* values are independent of "smart attack" removal technique devised in Section 6. In case, this novel technique is applied, the *l* LEAD nodes gets $g+l-1$ extra keys. So their memory is a bit strained. However Moore's law ascertains that memory expansion with time is easy as compared to other hardware. Therefore this excess memory overload *(O(k),k : original key rings sizes)* is reasonable to assume in front of existing computation and transreceiver overheads.

From this comparison table it is clear that the asymptotic behavior of the ratio r of *Sensornet* and *HC(Sensornet)* is similar or better than all other schemes except the scheme $TD(3;k;q);k=q$ and Merging Block (MB) design of [14, 15].

Former scheme needs computation of number theoretic problems during key agreement; while the later supports significantly less (merged) blocks. Moreover shared key discovery in *Sensornet* scheme requires $O((\log_p \mathcal{N})^3)$ time complexity and the amount of broadcast data is $O((\log_p \mathcal{N})^2)$. Broadcast data in both *Sensornet* and *HC(Sensornet)* is $O(\log_p\sqrt{\mathcal{N}})$ since transfer of basis vectors not required. This is advantageous over many KPS that require more data broadcast and complex establishment mechanism.

### 9.5.2. Scalability Comparison

*Sensornet* and thus *HC(Sensornet)* can support large networks. This is because the choice *n* and respectively *m* and/or *s* are unbounded in theory. Thus networks designed by our schemes are scalable.

Scalability is a major challenge in most deterministic KPS. For instance, the schemes [4, 9-12] have restricted scaling. This owes to the fact that key establishment for these network require general solutions of polynomials. Therefore, the complexity of the key establishment process increases with increment in degree of these polynomials. Random schemes can scale arbitrarily [19]; but at the expense of desirable parameters like connectivity, resilience, storage (key-node ratio), etc. Therefore, we opt deterministic schemes while designing KPS [18]. Also refer to Section 3.3.

## 10. Conclusion

Realizing the need of deterministic KPS with desirable properties (set out in Section 3.4) to address the problem of key management in low cost networks, we propose one such scheme. Since the scheme is constructed using nets in a vector space, we named it as *Sensornet*. The scheme is later improved using Hash Chains trick of Bechkit et al. [5,6] to obtain *HC(Sensornet)*, a resilience improved version. Key establishment of both *Sensornet* and *HC(Sensornet)* is a great advantage over other schemes, but still exploits the network's key sharing graph. This is overcome by a generic approach devised in Section 6 that removes "smart attacks". Resilience of *Sensornet* can be improved by a Hash Chains approach (along with the above approach) to yield *HC(Sensornet)*.

## 11. Related Future Works

Although both or schemes, *Sensornet* and *HC(Sensornet)* suffer from lack of full connectivity, it is very close to full connectivity for large size of partial spread. Moreover, the generic computations in Section 9.4 establish that connectivity of *Sensornet* is good (either direct or 1–hop path). It is preferable to have full connectivity or at least a deterministic path in case of 1–hop connectivity. The sophisticated MB designs of [14, 15] establishes a deterministic 1– hop connectivity for the Reed Solomon code based KPS (Ruj and Roy [4]). These heavily design dependent works can certainly open the doors for future research by considering similar constructions over *Sensornet* in place of other combinatorial design based schemes.

Efficient deterministic protocols for security applications in low cost IoT networks have restricted scaling. We adopt them due to their predictable connectivity and resilience. Scalable deterministic security protocols with flat topology using SKC techniques is a major area of study. Our "smart attack" removal approach may give interesting leads.

## References


[1]. Dalai, D. K. and Sarkar, P., Sensornet: A key predistribution scheme for distributed sensors using nets, in *Proceedings of the 6th International Conference on Sensor Networks SENSORNETS*, Porto, Portugal, Vol. 1, 2017, pp. 49-58.

[2]. Zhu, S., Setia, S., and Jajodia, S., LEAP - efficient security mechanisms for largescale distributed sensor networks, in *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems, (SenSys'03)*, Los Angeles, California, USA, November 5-7, 2003, pp. 308–309.

[3]. Eschenauer, L. and Gligor, V., A key-management scheme for distributed sensor networks, in *Proceedings of 9th ACM Conference on Computer and Communications Security,* 2002, pp. 41–47.

[4]. Ruj, S. and Roy, B. K., Key predistribution schemes using codes in wireless sensor networks, in Information Security and Cryptology, Inscrypt 2008, *Lecture Notes in Computer Science*, 5487, pp. 275–288.

[5]. Bechkit, W., Bouabdallah, A., and Challal, Y., Enhancing resilience of probabilistic key pre-distribution schemes for WSNs through hash chaining, in *Proceedings of the 17th ACM Conference on Computer and Communications Security (CCS'10),* Chicago, Illinois, USA, October 4-8, 2010, pp. 642–644.

[6]. Bechkit, W., Challal, Y., and Bouabdallah, A., A new class of hash-chains based key pre-distribution schemes for WSN, *Computer Communications*, 36, 3, 2013,
pp. 243–255.

[7]. Pietro, R. D., Mancini, L. V., and Mei, A., Energy efficient node-to-node authentication and communication confidentiality in wireless sensor networks, *Wireless Networks,* 12, 6, 2006, pp. 709–721.

[8]. Bose, M., Dey, A., and Mukerjee, R., Key pre-distribution schemes for distributed sensor networks via block designs, *Design, Codes and Cryptography,* 67, 1, 2013, pp. 111–136.

[9]. Camtepe, S. A. and Yener, B., Combinatorial design of key distribution mechanisms for wireless sensor networks, *IEEE/ACM Transactions on Networking,* Vol. 15, Issue 2, April 2007, pp. 346 - 358.

[10]. Lee, J. and Stinson, D. R., Deterministic key pre-distribution schemes for distributed sensor networks, in *Selected Areas in Cryptography (SAC'04), Lecture Notes in Computer Science*, 3357, 2004, pp. 294–307.

[11]. Lee, J. and Stinson, D. R., A combinatorial approach to key predistribution for distributed sensor networks, in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC'05),* 2005, pp. 1200-1205.

[12]. Lee, J. and Stinson, D. R., On the construction of practical key predistribution schemes for distributed sensor networks using combinatorial designs, *ACM Transactions on Information and System Security,* 11, 2, 2008, pp. 854–867.

[13]. Wei, R. and Wu, J., Product construction of key distribution schemes for sensor networks, in *Selected Areas in Cryptography (SAC) 2004, Lecture Notes in Computer Science,* 3357, 2004, pp. 280–293.

[14]. Bag, S., Dhar, A., and Sarkar, P., 100 % connectivity for location aware code based KPD in clustered WSN: Merging blocks, in *Proceedings of the Information Security Conference (ISC'12), Lecture Notes in Computer Science,* 7483, 2012, pp. 136–150.

[15]. Sarkar, P., Rai, B. K., and Dhar, A., Connecting, scaling and securing RS code and TD based KPDs in WSNs: deterministic merging, in *Proceedings of the Fourteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '13),* Bangalore, India, July 29 - August 01, 2013, pp. 301–304.

[16]. Erdos, P. and Renyi, A., On the evolution of random graphs, *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, 1960.

[17]. Blom, R., An optimal class of symmetric key generation systems, in *Advances in Cryptology - Eurocrypt 1984, Lecture Notes in Computer Science,* 209, 1985, pp. 335–338.

[18]. Paterson, M. B. and Stinson, D. R., A unified approach to combinatorial key predistribution schemes for sensor networks, *Design, Codes and Cryptography,* 71, 3, 2014, pp. 433-457.

[19]. Ruj, S. and Pal, A., Preferential attachment model with degree bound and its application to key predistribution in WSN, in *Proceedings of the IEEE Conference on Advanced Information Networking and Applications (AINA'16),* 2016, pp. 677–683.

[20]. Henry, K. J., Paterson, M. B., and Stinson, D. R., Practical approaches to varying network size in combinatorial key predistribution schemes, in *Selected Areas in Cryptography (SAC) 2013, Lecture Notes in Computer Science,* 8282, 2014, pp. 89-117.

[21]. Lidl, R. and Niederreiter, H., Finite fields. Encyclopaedia of mathematics and its applications. *Cambridge University Press*, 1997.

[22]. Eisfeld, J. and Storme, L., (partial) t-spreads and minimal t-covers in finite projective spaces, in Lecture notes for the Socrates Intensive Course on Finite Geometry and its Applications, *University of Ghent*, 2000.

[23]. Lu, H. Y., Partial spreads and hyperbent functions in odd characteristic, Master's Thesis, *Simon Fraser University*, 2008.

[24]. Johnson, N. L., Combinatorics of Spreads and Parallelisms, *CRC Press*, 2010.

[25]. Johnson, N. L., Jha, V., and Biliotti, M., Handbook of Finite Translation Planes, volume 289 of Pure and Applied Mathematics, *Chapman & Hall/CRC*, 2007.

[26]. Bu, T., Partitions of a vector space, *Discrete Mathematics*, 31, 1980, pp. 79–83.

[27]. Wu, B., Ps bent functions constructed from finite prequasifield spreads, 2013. Available at http://arxiv.org/pdf/1308.3355.pdf

[28]. Çeşmelioğlu A., Meidl, W., and Pott, A., Bent functions, spreads, and o-polynomials, *SIAM Journal of Discrete Mathematics,* 29, 2, 2015, pp. 854–867.

[29]. Dong, J. W., Pei, D. Y., and Wang, X. L., A class of key predistribution schemes based on orthogonal arrays, *Journal of Computer Science and Technology*, 23, 2011, pp. 825–831.

*Sensors & Transducers, Vol. 212, Issue 5, May 2017, pp. 50-56*

**Sensors & Transducers**
Published by IFSA Publishing, S. L., 2017
http://www.sensorsportal.com

# CO Gas Adsorption on $SnO_2$ Surfaces: Density Functional Theory Study

**[*] Hayk Zakaryan and Vladimir Aroutiounian**
Yerevan State University, 1 Alex Manoogian, 0025, Armenia
Tel.: +37410555240, fax: +37410554641
[*] E-mail: zhayk91@gmail.com

*Received: 3 May 2017 /Accepted: 29 May 2017 /Published: 31 May 2017*

**Abstract:** This research is devoted to the investigation of the toxic carbon monoxide gas adsorption mechanisms on the tin dioxide semiconductor. We used density functional theory to describe adsorption processes and found out that the Mars-van Krevelen adsorption mechanism is not responsible for adsorption on (101) and (001) surface orientations of tin dioxide. For (110) and (100) surfaces, after adsorption carbon dioxide molecule forms and desorbs from the surfaces. For (101) surface orientation, carbon monoxide adsorb to the surface's oxide by carbon atom and stay bonded to it. Charge transfer from the molecule to the surface, which equal to 1.9e calculated by Bader charge analysis. In the case of (001) surface orientation, carbon monoxide adsorb to surface's oxygen and stay bonded too. Here, we consider half and full surface coverages. It was shown, that during full surface coverage, only one molecule can adsorbs and transfer 2e charge. Electronic density of states calculation was done to explain the increase of surface conductance.

**Keywords:** Gas, Sensor, DFT, CO, Adsorption, Mars-van Krevelen.

## 1. Introduction

Carbon monoxide (CO) is tasteless and transparent substance, which is known as an "invisible killer" due to the high level of toxicity. It is extremely poisonous and can even cause health effects up to a certain extent – 9 ppm, therefore, the detection of CO in the environment is quite vital. For the above reasons, chemical sensors are used to precisely monitor the concentration of the target compounds in the air. The main part of chemical sensors is sensing or active material, which interact with gas molecules. Usually, for sensing materials are used metal oxides like $TiO_2$, $SnO_2$, ZnO, $In_2O_3$.

In modern world, where technology is developing every day, it is crucial to have precise and functional devices. For sensor technology, it means to have a sensing material, which has different high parameters like high conductance, wide band gap, and high charge carriers' mobility. All these parameters have the $SnO_2$ semiconductor. The gas sensors made by it, has high sensitivity and low response time for wide variety of molecules, such as CO, $H_2$, $CH_3OH$, $NO_x$, etc. [1-8]. Sensitivity is the one of the most important parameters of sensors; it defines how much signal change when sensors detect the target gas. In order to increase the sensitivity of the detectors, it is crucial to deeply investigate adsorption mechanism of CO compounds.

Several models, which are named as Langmuir Hishelwood [10-11], Eley Redeal [12] and the MvK [13], have been developed to describe the mechanisms that are responsible for CO adsorption on oxide semiconductors. The MvK adsorption mechanism is in a good agreement with the experiments on $SnO_2$–(110) surface [14-15]. Apparently, the MvK mechanism consists of following steps:



http://www.sensorsportal.com/HTML/DIGEST/P_RP_0224.htm

1) Adsorbed CO molecule is reacting with tin dioxide oxygen and forms $CO_2$ compound, leaving oxygen vacancy in the material;

2) Then remained vacancy fills by adsorbed $O_2$ molecule from the environment;

3) And finally, another CO molecule is reacting with already bonded $O_2$ and forms $CO_2$, leaving material in its initial undisturbed state.

We would like to stress that there are plenty of manuscripts dedicated to the investigation of CO absorption on $SnO_2$ surfaces, using first principle and ab initio DFT calculations [16-17]. Some calculations were done particularly for pristine (110) surface orientation [18], oxygen reach [15] and tainted surfaces [16]. However, very few are devoted to other surface orientations (100), (101) [5] and (001) [6]. In practice, surface orientation immensely influences sensor parameters such as sensitivity, time response etc.

In this manuscript, we explore CO adsorption on various surface orientations (110), (100), (101), (001) (Fig. 1) of $SnO_2$ and determine the most optimal configurations for adsorption, using ab initio DFT calculations in order to shed the light on the atomic scale processes that still remains elusive and unclear.

Here, we will rise following issues:

1) Does MvK mechanism similarly describing adsorption processes for all surface orientations of $SnO_2$?

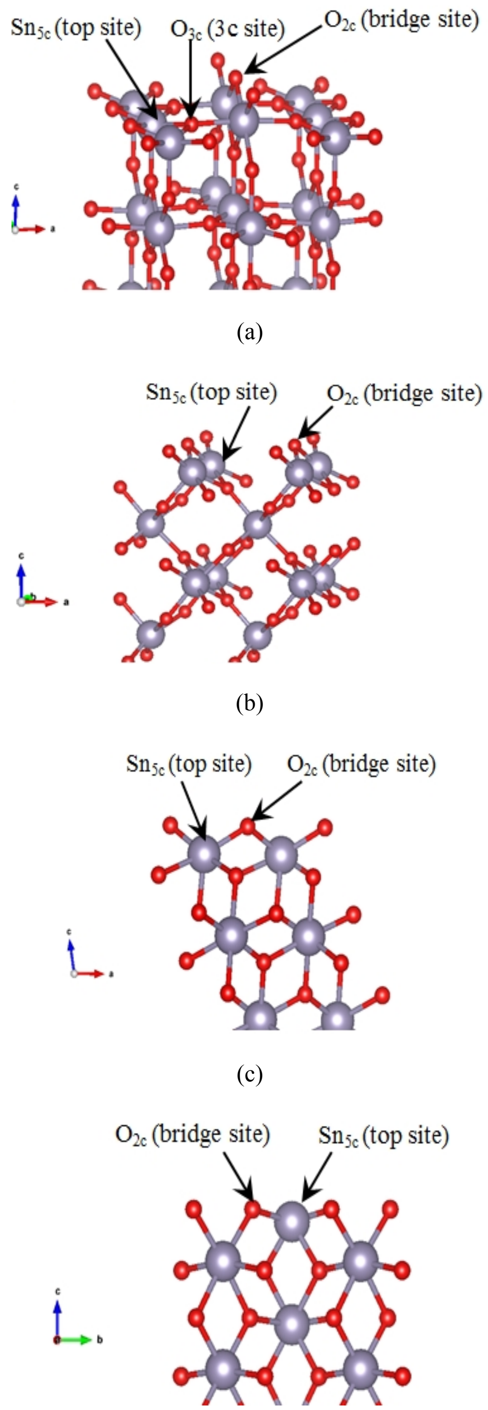
2) What is the exact amount of charge transferred between the surface and adsorbed CO molecule?

3) Which surface orientation is more prominent to the interaction with the CO molecule?

The paper is organized in the following way: Section 2 is devoted to computational methods and surface/bulk structure of $SnO_2$ sensing material. Section 3 is devoted to the adsorption processes, eDoS calculations for the various surface orientations of $SnO_2$. Section 4 is dedicated to discussions and conclusions.

## 2. Models and Computational Methods

Calculations were done using conventional ab initio DFT [21-22] method implemented in Vienna ab initio Simulation Package (VASP) [23-25]. DFT relaxations were done within Perdew-Burke-Ernzerhof (PBE) generalized gradient approximation (GGA) [26]. The 500 eV cutoff energy was chosen due to the total energy convergence from that value. Surface structures were relaxed until the threshold net force on atoms become less than 0.01 eV/A. The Monkhorst-Pack scheme [27] was used to sample the Brillouin zone, using 6×6×1 k-points mesh. During the calculations of adsorbed molecules on surface, each configuration was relax to find the most relaxed structure or the structure were the forces on atoms are the minimum. The ionic relaxation was done using conjugate gradient method implemented in VASP. After that relaxation, the Bader charge analysis was done to find out charge distribution around each atom [28-30]. The difference, between charge distribution of the relaxed surface and pure surface, will be the amount of charge transferred to the surface after adsorption finished.




(d)

**Fig. 1.** Different surfaces of $SnO_2$ and possible adsorption sites on it. (a) - (110), (b) - (100), (c) - (101), (d) - (001).

Stable adsorbed configurations were found using the following equation:

$$E_{surf} = E_{ads} - E_{clean} - E_{CO}, \quad (1)$$

where $E_{ads}$ is the adsorption energy, $E_{surf}$ is the total energy of $SnO_2$ surface and adsorbed CO, $E_{clean}$ is the total energy of pure surface without CO and $E_{CO}$ is the total energy of the gas molecule. It is clear from this equation that if $E_{ads}$ is negative the configuration of adsorbed site is stable, in other words, the process is exothermic. In the case of positive $E_{ads}$, the process is endothermic: the molecule will not adsorb to the surface and will remain in the non-interacting state.

Bulk $SnO_2$ has rutile, tetragonal structure, corresponding to the P42/mnm space group. The lattice parameters of $SnO_2$ from [31-32] are a = 4.82 Å, c = 3.23 Å and u = 0.307. In our calculations (see Fig. 1) each (110), (100), (101), (001) surface consists of 4 layers and relative stability has the following sequence (110), (100), (101), (001) [6]. Here, number of layers was tuned to check the convergence of surface energy. However, after getting all results, we double checked the obtained data by recalculating stable structures with a big substrate of 12 atoms of tin and 24 atoms of oxygen and make sure that the results are reliable. For all calculations, we choose vacuum thickness of 15 Å, which is greater than substrate thickness. For each surface, the electronic density of states (eDoS) was calculated and established that the gap between valence and conduction bands was underestimated, because of the self-consistency of DFT calculations. From experiments, it is known that value of band gap is 3.6 eV, but we get 1 eV, which is in accordance with previous DFT calculations [31]. Despite of the gap underestimation, it doesn't influence on the value of the adsorption energies and bond lengths of the calculation. Even the gap is narrower than in experiment; the trend of band gap change after gas molecule adsorption would be the same, because charge transfer or charge density calculated by PBE has minimal error between known functionals.

## 3. Results and Discussion

In this section, each surface orientation would be described separately. Bader charge analysis, eDoS and stable structures would be presented as well.

### 3.1. $SnO_2$ (110) Surface

This surface orientation consists of 4 layers, each comprising of 3 subsequent layers. Top sublayer represents oxygen, where each atom ($O_{2c}$) is connected by the covalent bond to 2 Sn atoms. The Second sublayer has 2 Sn atoms and 2 oxygen atoms. Third one is similar to the first layer. So, together these 3 sublayers can be treated as one layer which is continually repeated 4 times (see Fig. 1(a)). There are 3 possible adsorption sites: top site (t), bridge site (br), three-coordinated oxygen site (3c). For each site, we consider 2 configurations:

1) C atom in a CO molecule is closer to the surface (C down configuration),

2) The vice versa configuration (O down configuration). Thus, we end up with 6 possible configurations for (110) surface.

For each separate configuration calculated adsorption energies are given in Table 1. As we can clearly see the O down configuration for all surfaces is completely unstable.

In the case of C down configuration, (110) surface has 3 stable sites. However, there is only one that is transferring the charge form molecule to the surface, and it is corresponding to the event when molecule approaches to br-site, reacts with oxygen and takes it away forming a $CO_2$ molecule and leaves oxygen vacancy on the surface. Here, vacancy could be an adsorption site for $O_2$ or CO molecule, as described in [16, 19]. The eDoS of surface reveals that conductivity of the surface increases by decreasing of the band gap. The Bader charge analysis shows charge transport to the surface of 1.7e. Such processes examined in a number of experimental and theoretical studies [14], [16, 19]. It is important to note, that for this particular configuration and surface orientation the first step of MvK mechanism is preserved, and the distance between $CO_2$ molecule and surface is 3.1 Å.

**Table 1.** Adsorption energies, distances and bader charge analysis of (110) surface.

| Ads. sites | $E_{ads}$ (eV) | CO distance d surface (Å) | Charge transferred to surface ( e ) |
|---|---|---|---|
| br | -0.48 | 3.1 | 1.7 |
| br (O down) | 0.31 | 2.58 | 0.0 |
| top | -0.16 | 2.465 | 0.0 |
| top (O down) | 0.01 | 2.69 | 0.0 |
| 3c | -0.03 | 2.61 | 0.0 |
| 3c (O down) | 0.5 | 2.62 | 0.0 |

For other sites (top and 3c) the CO molecule does not exchange electrons, which serves as a confirmation that we have physisorption process onto the surface. In these cases, band gap does not change and distances from carbon to tin atoms are 2.46, 2.61 Å, respectively

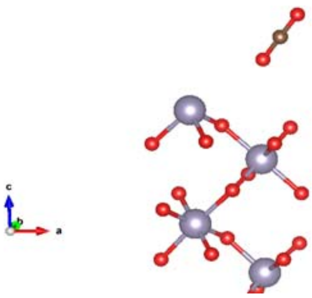
### 3.2. $SnO_2$ (100) Surface

For (100) surface orientation the unit cell also consists of same 4 layers, where each layer can be divided into 3 sublayers: O-Sn-O layers. The top atom of the surface is 2 coordinated oxygen as in (110) surface, see Fig. 1(b). Adsorption energies for possible two site configurations are given in Table 2. According to the calculations interaction of CO with the $O_2$c site of oxygen leads to the formation of $CO_2$

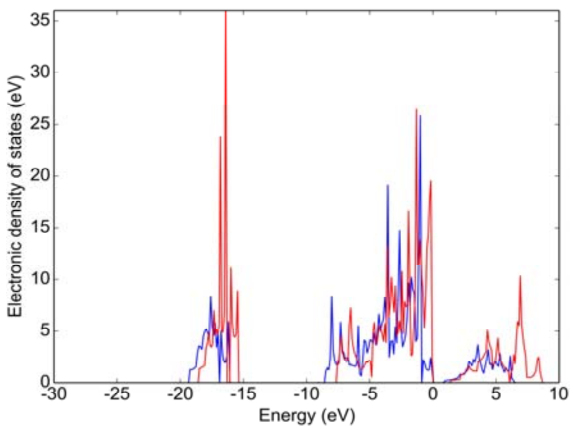
molecule and also leaves a vacancy on the surface (Fig. 2(a)).

**Table 2.** Adsorption energies, distances and bader charge analysis of (100) surface.

| Ads. sites | $E_{ads}$ (eV) | CO distanced surface (Å) | Charge transferred to surface ( e ) |
|---|---|---|---|
| br | -0.48 | 3.25 | 1.6 |
| top | -0.16 | 2.65 | 0.0 |



(a)




(b)

**Fig. 2.** Structure and eDoS of (100) surface. (a) Desorbed $CO_2$ molecule from (100) surface (red circle is oxygen, grey is Sn, brown is C atoms). (b) eDoS of substrate before adsorption (red line), and after $CO_2$ desorption (blue line).

In this case, the first step of MvK mechanism is also conserved and it leads to charge transfer of 1.6e to the surface. Therefore, the surface conductance is increasing, because of transferred charge leading to the band gap reduction by 0.1 eV in comparison to the uncharged surface, see Fig. 2(b). Here, the blue lines represent the eDoS after desorption that correspond to 3.25 Å distance from CO molecule to surface.

The Second possible adsorption site is top (t) on Sn atom. Here as for (110), the CO is physisorbed and no charge transfer has been observed. Due to physisorption, the distance from a carbon atom of CO molecule to the oxygen atom of the surface is 2.65 Å.

### 3.3. $SnO_2$ (101) Surface

In the case of (101) surface orientation, each of 4 layers consists of 3 subsequent layers of 2O, 2Sn, 2O. There are 3 possible sites of adsorptions O2c, Sn atom and 3 coordinated O, which is located in third sub layer as it is shown in Fig. 1(c). Here, only one configuration has negative adsorption energy and it is $O_{2c}$ (Table 3).

**Table 3.** Adsorption energies, distances and bader charge analysis of (101) surface.

| Ads. sites | $E_{ads}$ (eV) | CO distanced surface (Å) | Charge transferred to surface ( e ) |
|---|---|---|---|
| $O_{2c}$ | -0.47 | 1.18 | 1.9 |

The CO molecule adsorbs by the surface and remains connected to it in $O_{2c}$ site with a distance of 1.18 Å (see Fig. 3). During that process, the distance between $O_{2c}$ oxygen and Sn atom increases up to 2.5 Å. The Bader charge analysis shows that there is 1.9e charge transferred to the surface, which make it more conductive. Thus, adsorption mechanism on (101) surface differs from MvK, because no $CO_2$ desorption observed. Conductivity increases and band gap decreases due to transferred charge, see Fig. 3(b), where blue lines represent eDoS after adsorption and have more states around Fermi level (Energy = 0 in the horizontal axis). To find the whole adsorption mechanism it is necessary to do molecular dynamics simulation (MD), which should be done in future researches.

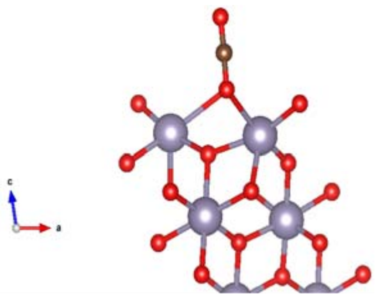
### 3.4. $SnO_2$ (001) Surface

In this case, we have a completely different situation. Instead of 4 layers that consist of 3 subsequent layers, we have got only one that consists of one atom of Sn and 2 oxygen atoms. There are two possible sites: on top of Sn (t) site and $O_{2c}$ (2c) site Fig. 1 (d). Moreover, there are 2 equivalent 2c sites in one unit cell, thus we should take into account two possible coverages. The first is when both sites are occupied by the CO molecule and form one mono layer (ML=1). The second possible coverage is when only one site is occupied and half mono layer of CO forms (ML = 0.5).

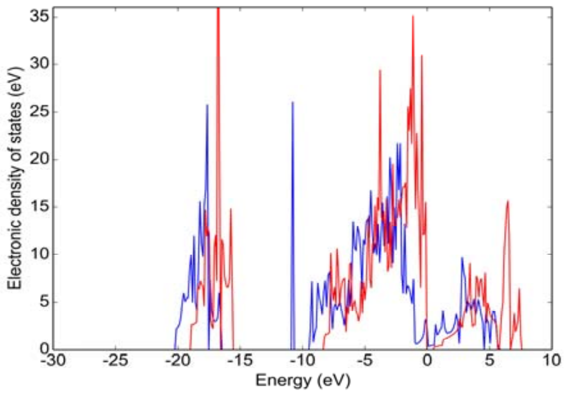
For ML = 0.5, carbon monoxide adsorbed and stay bounded with a distance of 1.16 Å (Fig. 4) to the $O_{2c}$ atom, transferring 2e charge to the substrate (see Table 4).

For ML = 1, one CO molecule adsorbed and one physisorbed in a 1.255 and 2.6 Å distances, respectively. In fact, for (001) only ML=0.5 coverage can happen, because when one CO adsorbed, the second one will be physisorbed. Bader charge analysis shows that 1.83e was transferred to the surface.

In both cases, surface conduction increases and band gap decreases due to charge transfer from the CO molecule.



(a)




(b)

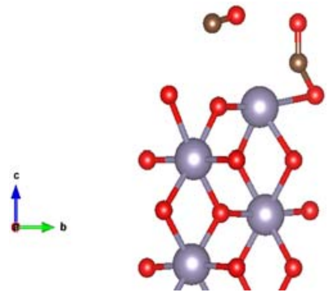
**Fig. 3.** (a) Adsorbed CO on (101) surface (red circle is oxygen, grey is Sn, brown is C atoms). (b) eDoS of clean (101) surface (red line) and CO adsorbed on it (blue line).

**Table 4.** Adsorption energies, distances and bader charge analysis of (001) surface.

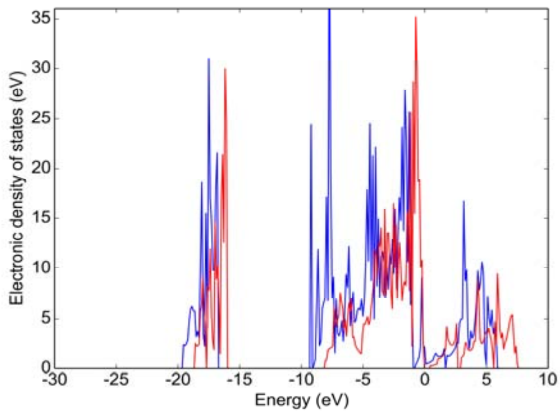
| **Ads. sites** | **$E_{ads}$ (eV)** | **CO distance d surface (Å)** | **Charge transferred to surface ( e )** |
|---|---|---|---|
| 2c, ML = 0.5 | -1.19 | 1.16 | 2.0 |
| 2c, ML = 1 | -1.24 | 1.25 | 1.8 |
| top | -0.19 | 2.41 | 0.0 |

On Fig. 4 (b)(d), the red lines represent eDoS of undistorted surface and blue correspond to the eDoS after adsorption, which has more states around Fermi level compared to undistorted surfaces eDoS. Physisorption occurs for the top site with distance 2.41 Å.

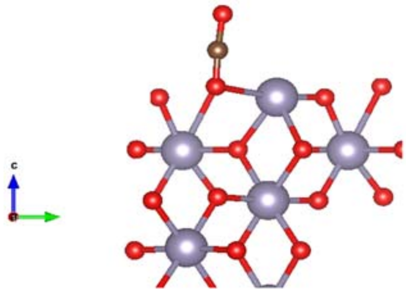
So, MvK mechanism is not responsible for adsorption on (001) surface orientation, because CO molecule adsorb and remain bonded to the surface instead of forming $CO_2$.



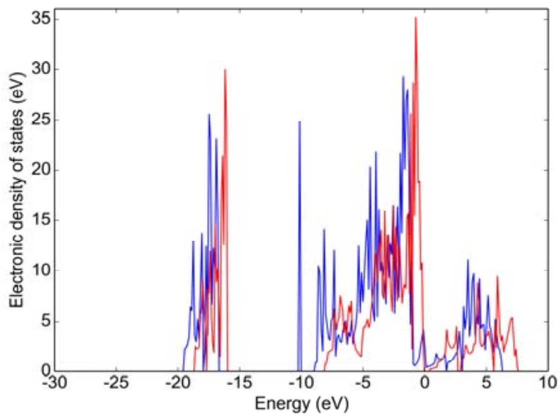
(a)




(b)



(c)




(d)

**Fig. 4.** Configurations and eDoS of (001) surface (red circle is oxygen, grey is Sn, brown is C atoms). (a) Adsorption when ML=1. (b) eDoS before and after adsorption, when ML = 1. (c) Adsorption when ML= 0.5. (d) eDoS before and after adsorption, when ML = 1.

## 4. Conclusions

Previously, it was considered that adsorption on $SnO_2$ surfaces follows MvK mechanism; however, here we have proven that CO adsorption on $SnO_2$ (101), (001) is different. Here, the C atom of the CO molecule remains bonded to surface's O atom. Also, we find out that in the case of (001) surface orientation CO coverage can be only half mono layer (ML = 0.5). For (110) and (100) surfaces, we establish that adsorption obeys MvK mechanism, where its first stage $CO_2$ molecules are forming during CO-surface interaction. For all adsorption cases, O down configuration was not stable, due to positive adsorption energy. To find whole adsorption mechanism for (101) and (001) surface orientations, it is necessary to do MD calculation and find next steps of reaction between CO and surface. This will be done in the future research.

The Bader charge analysis reveals that charge transfer to (101), (001) surfaces are 1.9e, 2e respectively and 1.7e, 1.6e for (110), (100) surfaces. The eDoS combined with Bader analysis shows that (101), (001) surface orientations gather more electrons than the rest orientations, thus, those should be considered as a better platform for the interaction of the CO molecules with $SnO_2$ surfaces. We believe that our findings will pave the way for the fabrication of $SnO_2$ based CO sensors with higher sensitivity and lower response time.

## Acknowledgment

This work was supported by NATO project SfP-984597. We are grateful to Prof. Artem Oganov for opportunities to perform the DFT calculations on the Rurik supercomputer at Moscow Institute of Physics and Technology (State University). We are also thankful to Davit Ghazaryan and Dr. Alexander Kvashnin for the useful discussions and significant amendments of this manuscript

## References

[1]. V. Aroutiounian, Metal oxide hydrogen, oxygen, and carbon monoxide sensors for hydrogen setups and cells, *International Journal of Hydrogen Energy*. Vol. 32, Issue 9, 2007, pp. 1145-1158.

[2]. V. M. Aroutiounian, Metal oxide gas sensors decorated with carbon nanotubes, *Lithuanian Journal of Physics*, Vol. 55, Issue 4, 2015, pp. 319-329.

[3]. V. M. Aroutiounian, V. M. Arakelyan, E. A. Khachaturyan, G. E. Shahnazaryan, M. S. Aleksanyan, L. Forro, *et al*., Manufacturing and investigations of i-butane sensor made of $SnO_2$/multiwall-carbon-nanotube nanocomposite, *Sensors and Actuators B: Chemical*, Vol. 173, No. 7, 2012, pp. 890-896.

[4]. M. Batzill, U. Diebold, The surface and materials science of tin oxide, *Progress in Surface Science*, Vol. 79, Issue (2–4), 2005, pp. 47-154.

[5]. S. Gong, J. Liu, J. Xia, L. Quan, H. Liu, D. Zhou, Gas sensing characteristics of $SnO_2$ thin films and analyses of sensor response by the gas diffusion theory, *Materials Science and Engineering: B*, Vol. 164, Issue 2, 2009, pp. 85-90.

[6]. A. Adamyan, Z. Adamyan, V. Aroutiounian, A. Arakelyan, K. Touryan, J. Turner, Sol–gel derived thin-film semiconductor hydrogen gas sensor, *International Journal of Hydrogen Energy*, Vol. 32, Issue 16, 2007, pp. 4101-4108.

[7]. V. G. Paremuzyan, V. M. Aroutiounian, Influence of reduction in the heat of adsorption on noise characteristics of sensors, *Sensors & Transducers*, Vol. 137, Issue 2, February 2012, pp. 95-103.

[8]. A. Z. Adamyan, Z. N. Adamyan, V. M. Aroutiounian, Study of sensitivity and response kinetics changes for $SnO_2$ thin-film hydrogen sensors, *International Journal of Hydrogen Energy*, Vol. 34, Issue 19, 2009, pp. 8438-8443.

[9]. I. Langmuir, The mechanism of the catalytic action of platinum in the reactions $2CO + O_2 = 2CO_2$ and $2H_2 + O_2 = 2H_2O$, *Transactions of the Faraday Society*, Vol. 17, 1922, pp. 621-654.

[10]. R. H. Nibbelke, M. A. J. Campman, J. H. B. J. Hoebink, G. B. Marin, Kinetic Study of the CO Oxidation over Pt/$\gamma$-$Al_2O_3$ and Pt/Rh/$CeO_2$/$\gamma$-$Al_2O_3$ in the Presence of $H_2O$ and $CO_2$, *Journal of Catalysis*, Vol. 171, Issue 2, 1997, pp. 358-373.

[11]. C. C. Cheng, S. R. Lucas, H. Gutleben, W. J. Choyke, J. T. Yates, Atomic hydrogen-driven halogen extraction from silicon (100): Eley-Rideal surface kinetics, *Journal of the American Chemical Society*, Vol. 114, Issue 4, 1992, pp. 1249-1252.

[12]. P. Mars, D. W. van Krevelen, Oxidations carried out by means of vanadium oxide catalysts, *Chemical Engineering Science*, Vol. 3, No. 1, 1954, pp. 41-59.

[13]. Z. Lu, D. Ma, L. Yang, X. Wang, G. Xu, Z. Yang, Direct CO oxidation by lattice oxygen on the SnO 2(110) surface: a DFT study, *Phys. Chem. Chem. Phys.*, Vol. 16, Issue 24, 2014, pp. 12488-12494.

[14]. M. Melle-Franco, G. Pacchioni, CO adsorption on SnO2(110): cluster and periodic ab initio calculations, *Surface Science*, Vol. 461, Issue 1–3, 2000, pp. 54-66.

[15]. X. Wang, H. Qin, Y. Chen, J. Hu, Sensing Mechanism of $SnO_2$ (110) Surface to CO: Density Functional Theory Calculations, *The Journal of Physical Chemistry C*, Vol. 118, Issue 49, 2014, pp. 28548-28561.

[16]. Y. B. Xue, Z. A. Tang, Density functional study of the interaction of CO with undoped and Pd doped SnO2(110) surface, *Sensors and Actuators B: Chemical*, Vol. 138, Issue 1, 2009, pp. 108-112.

[17]. P. Hohenberg, W. Kohn, Inhomogeneous Electron Gas, *Physical Review*, Vol. 136, Issue 3B, 1964, pp. B864–B871.

[18]. W. Kohn, L. J. Sham, Self-Consistent Equations Including Exchange and Correlation Effects, *Physical Review*, Vol. 140, Issue 4A, 1965, pp. A1133–A1138.

[19]. G. Kresse, J. Hafner, Ab initio molecular dynamics for liquid metals, *Physical Review B*., Vol. 47, Issue 1, 1993, pp. 558-561.

[20]. G. Kresse, J. Hafner, Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium, *Physical Review B*, Vol. 49, Issue 20, 1994, pp. 14251-14269.

[21]. G. Kresse, J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, *Physical Review B*, Vol. 54, Issue 16, 1996, pp. 11169-11186.

[22]. J. P. Perdew, K. Burke, M. Ernzerhof, Generalized Gradient Approximation Made Simple, *Physical Review Letters*, Vol. 77, Issue 18, 1996, pp. 3865-3868.
[23]. H. J. Monkhorst, J. D. Pack, Special points for Brillouin-zone integrations, *Physical Review B*, Vol. 13, Issue 12, 1976, pp. 5188-5192.
[24]. G. Henkelman, A. Arnaldsson, H. Jónsson, A fast and robust algorithm for Bader decomposition of charge density, *Computational Materials Science*, Vol. 36, Issue 3, 2006, pp. 354-360.
[25]. E. Sanville, S. D. Kenny, R. Smith, G. Henkelman, Improved grid-based algorithm for Bader charge allocation, *Journal of Computational Chemistry*, Vol. 28, Issue 5, 2007, pp. 899-908.
[26]. W. Tang, E. Sanville, G. Henkelman, A grid-based Bader analysis algorithm without lattice bias, *Journal of Physics: Condensed Matter*, Vol. 21, Issue 8, 2009, pp. 1-7.
[27]. T. T. Rantala, T. S. Rantala, V. Lantto, Surface relaxation of the (110) face of rutile $SnO_2$, *Surface Science*, Vol. 420, Issue 1, 1999, pp. 103-109.
[28]. I. Manassidis, M. J. Gillan, First-principles study of $SnO_2$(110), *Surface Review and Letters*, Vol. 01, Issue 04, 1994, pp. 491-494.
[29]. T. T. Rantala, T. S. Rantala, V. Lantto, Electronic structure of $SnO_2$ (110) surface, *Materials Science in Semiconductor Processing*, Vol. 3, Issue 1–2, 2000, pp. 103-107.
[30]. C. Xu, Y. Jiang, D. Yi, S. Sun, Z. Yu, Environment-dependent surface structures and stabilities of $SnO_2$ from the first principles, *Journal of Applied Physics*, Vol. 111, Issue 6, 2012, pp. 063504-1 – 063504-9.
[31]. Y. Duan, Electronic properties and stabilities of bulk and low-index surfaces of SnO in comparison with $SnO_2$ : A first-principles density functional approach with an empirical correction of van der Waals interactions, *Physical Review B*, Vol. 77, Issue 4, 2008, pp. 045332-1 – 045332-22.
[32]. Y. Sun, F. Lei, S. Gao, B. Pan, J. Zhou, Y. Xie, Atomically Thin Tin Dioxide Sheets for Efficient Catalytic Oxidation of Carbon Monoxide, *Angewandte Chemie International Edition*, Vol. 52, Issue 40, 2013, pp. 10569–10572.

*Sensors & Transducers, Vol. 212, Issue 5, May 2017, pp. 57-69*

**Sensors & Transducers**

Published by IFSA Publishing, S. L., 2017
http://www.sensorsportal.com

# The Novel Artificial Intelligence Based Sub-Surface Inclusion Detection Device and Algorithm

**Jong-Ha LEE**
Keimyung University, School of Medicine, South Korea
Tel.: (82)-53-580-3736, fax: (82)-53-580-3746
E-mail: segeberg@kmu.ac.kr

*Received: 10 May 2017 /Accepted: 29 May 2017 /Published: 31 May 2017*

**Abstract:** We design, implement, and test a novel tactile elasticity imaging sensor to detect the elastic modulus of a contacted object. Emulating a human finger, a multi-layer polydimethylsiloxane waveguide has been fabricated as the sensing probe. The light is illuminated under the critical angle to totally reflect within the flexible and transparent waveguide. When a waveguide is compressed by an object, the contact area of the waveguide deforms and causes the light to scatter. The scattered light is captured by a high resolution camera. Multiple images are taken from slightly different loading values. The distributed forces have been estimated using the integrated pixel values of diffused lights. The displacements of the contacted object deformation have been estimated by matching the series of tactile images. For this purpose, a novel pattern matching algorithm is developed. The salient feature of this sensor is that it is capable of measuring the absolute elastic modulus value of soft materials without additional measurement units. The measurements were validated by comparing the measured elasticity of the commercial rubber samples with the known elasticity. The evaluation results showed that this type of sensor can measure elasticity within ±5.38 %.

**Keywords:** Artificial intelligence, Tactile elasticity imaging sensor, Pattern matching algorithm

## 1. Introduction

Traditionally, physicians have used palpation to detect breast tumors or prostate tumors, which is based on the observation that the tissue abnormalities are usually associated with localized changes in mechanical properties such as low elasticity and stiffer tissues [1-2]. To help physicians detect tumors more efficiently, various imaging techniques utilizing different imaging modalities such as computer tomography, ultrasonic imaging, nuclear magnetic resonance imaging, and x-rays have been developed [3-5]. However, each of these techniques has limitations, including the radiation to the body, low specificity, complicated system, etc. Moreover, these techniques can only provide the spatial information of the tumor. They do not measure mechanical properties directly. The absolute material properties are very important to measure the severity of the tumor. Identifying a stiff region relative to the surrounding region does not lead to diagnosing tissue abnormalities completely. Therefore it is desirable to measure the absolute elastic modulus directly using tactile elasticity imaging technique.

In fact, different tactile sensors using diverse approaches have already been investigated in robotic systems and medical tools for surgery. They are based on piezoelectric [6-8], piezoresistive [9-10], or capacitive sensing [11]. Some sensors provide good spatial resolution through the use of microelectromechanical systems (MEMS) technology. However, its small measurable force range due to the brittle sensing elements such as silicone based diaphragms has not proven to be a reliable biomedical tool. In addition, most of them are in the form of an array of distributed pressure sensors on a flat plate and merely detect the applied force at that point. Without the ability to measure the displacement of the tissue deformation, the sensor cannot estimate the elasticity. Elasticity is used in cancer detection. Recently, some research groups use force sensing resistors and a super-resolution algorithm for a neck palpation device [12]. However, this approach can only detect the relative stiffness, not the absolute elastic modulus. Even though spatial resolution of the tactile sensor has been improved by the super-resolution algorithm, it has still fairly low resolution compared to human fingers, which were millions of mechanoreceptors per square inch of the skin. Some tactile sensors use piezoelectric cantilevers for the absolute elasticity measurement [13-14]. However, this method requires auxiliary instruments such as oscilloscope or voltage generators. This scheme also has relatively low spatial resolution due to its large size of the probe. Therefore in order for tactile sensors to be successfully developed as the palpation tool, high tactile spatial resolution is necessary for the precise elasticity measurement.

In this paper, we present a newly designed tactile elasticity imaging sensor. The rigid waveguide transduction based optical tactile sensors are already investigated in [15-16]. Our system is inspired by this system with important differences. In the current design, a polydimethylsiloxane (PDMS) are used to make a multi-layer flexible transparent waveguide. The mechanical properties (i.e. elastic modulus) of each layer have been matched with the three human finger layers, dermis, epidermis, and subcutanea, to maximize the sensitivity of touch. In order to have high tactile spatial resolution, we utilize the total internal reflection principle in the waveguide. A force applied to a waveguide causes the light to change the critical angle of internally reflected lights, and results in light scattering which can be captured by a camera. The salient feature of this sensor compared to the other tactile sensors is the capability of measuring the elasticity of the contacted object without any external force sensor. In the current design, the force distribution has been measured through the integration of tactile image pixel values. In order to accurately estimate 3-D displacements of the contacted object deformation, a non-rigid pattern matching algorithm is developed. This technique relies on matching the random patterns recorded in tactile images to obtain the surface displacements and gradients from which the strain field can be determined. The obtained stress and strain information are finally used to identify the elasticity of the contacted object.

In the following section, the background of human tactile perception is given. Next, the design and characteristic of the tactile elasticity imaging sensor is presented and its sensing principle is introduced. Then, the stress estimation and non-rigid pattern matching algorithm for the strain estimation are discussed. Finally, the experimental results and conclusions are presented.

## 2. Background

The tactile elasticity imaging sensor emulates a human finger. Here, we briefly review the human tactile perception and human finger biology.

### 2.1. Human Tactile Perception

Touch sensation is perceived via physical contact mainly through the skin. Human skin has about five million sensory cells, however, the cells are not evenly distributed. Areas such as fingertips and lips are more sensitive to touch because they have more nerve endings. Fingers can perceive a wide variety of tactile information such as roughness, softness, humidity, temperature, friction, pain, vibration and hardness. Human fingers also have the amazing ability to detect inclusion, such as tumors inside the tissues. In general, sensory receptors can be classified by their functions: chemoreceptors (chemical stimuli), nociceptors (pain), osmoreceptors (osmolarity of fluids), photoreceptors (light stimuli), and mechanoreceptors (mechanical stimuli) [17]. There are also two sensory systems that react via contact with a physical object: exteroceptive and proprioceptive sensory systems. Proprioceptive system is the sense of the relative position of neighboring parts of the body. Exteroceptive system is the response to external stimuli such as temperature, deformation of the skin and mechanical stimuli. The proposed sensor is mainly dealing with mechanoreceptors and exteroceptive sensor.

### 2.2. Biology of the Human Finger

A human finger has an oval shaped cross section, composed of tissue, and nail. The curved surface allows consistent and precise grasping and manipulation. Nails are effective in enlarging the stimuli on mechanoreceptors by sandwiching the tissue between the surface and the nail. Human tissues are made up of multiple layers: epidermis, dermis and subcutanea. Each layer has different physical properties. The outmost layer is the epidermis (elastic modulus: $1.4\times10^2$ Pa); beneath it is the dermis layer ($8.0\times10^4$ Pa) and the layer closest to the bone is the subcutanea ($3.4\times10^4$ Pa) [16]. The epidermis is the hardest layer, with the smallest elasticity at approximately 1 mm thickness. The dermis is a softer layer with more elasticity, usually 1 to 3 mm thick. The subcutanea, which fills the space between the dermis and bone, is mainly composed of fat and functions as a cushion when shock load is applied to the finger. Due to the differences in elastic coefficients, there is greater deformation of the inner layers, dermis and subcutanea, than the outmost layer, epidermis, when the finger presses into or moves along a surface. The multi-layer structure enhances the effective texture and hardness perception, which is why we have emulated this multi-layer structure for the proposed sensor.

## 3. Sensor Design and Sensing Principle

In this section, we present the concept, fabrication, and characterization of our sensor in detail.

### 3.1. Design Requirement for Emulating Human Finger

The tactile elasticity imaging sensor emulates the structure of a human finger. The design requirement is as follows.

1) *Human tissues*: Polydimethylsiloxane (PDMS) is used for emulating human tissue. PDMS creates a soft contact surface, which has proven to be effective in detecting the texture of material.

2) *Three-layered structure*: Emulating the structure of human tissue, three types of the PDMS with different elasticity are stacked together. This allows for more sensitive perception.

3) *Distribution of bone and nail elements*: In order to effectively obtain tactile sensory data, parts that function as the bone and nail are situated at the base of the sensor. In the current design, a heat-resistant borosilicate glass plate is used as the substrate for the stacked PDMS.

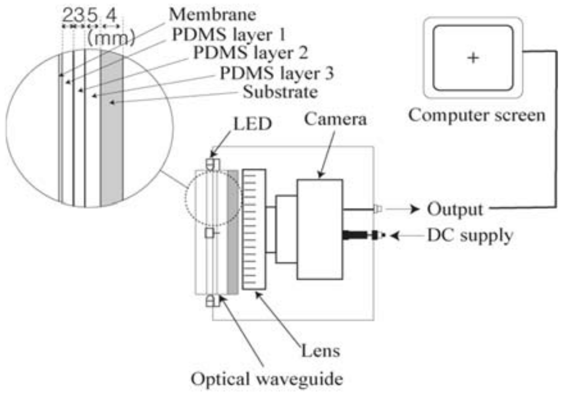
4) *Distributed sensor elements*: To emulate mechanoreceptors of a human finger, an optical based sensing method using a light reflection pattern and a digital imager is used. This is to obtain high spatial distribution of contact force.

### 3.2. Sensor Design

Fig. 1 (a) shows the schematic of the tactile sensor module and Fig. 1 (b) shows the integrated tactile sensor. The elasticity imaging sensor comprises of an optical waveguide, light sources, and a digital imager.

The optical waveguide is the main sensing probe of the device.

The optical waveguide is composed of PDMS (PDMS, $Si(CH_3)_2$), which is a high performance silicone elastomer [18-19]. The optical waveguide needs to be transparent and PDMS meets this requirement. In the current design, one of the Hydroxyl-terminated PDMS, RTV6186 has been used (R. S. Hughes, Baltimore, MD). The PDMS is produced through a process of pouring viscous fluid silicone and a catalyst into a mold cavity. Here, the viscous fluid silicone used is vinyl-stopped phenylmethypolyer and the catalyst is a mixture of components, including methylhydrogen polysiloxane, dimethyl, methyvinyl siloxane and dimethylvinyl terminated. The viscous fluid silicone is hardened by the catalyst. The hardness is dependent on the ratio of silicone and catalyst. The elastic moduli of three PDMS layers are set as the modulus values of epidermis, dermis, and subcutanea. The height of each layer is 2 mm for epidermis layer (PDMS layer 1), 3 mm for dermis layer (PDMS layer 2) and 5 mm for subcutanea layer (PDMS layer 3), respectively. The fabricated PDMS optical waveguide is shown in Fig. 2.




(a)

(b)

**Fig. 1.** (a) Schematic of the tactile elasticity imaging sensor. (b) The integrated version.

**Fig. 2.** Fabricated PDMS optical waveguide. The waveguide is elastic and flexible.

The digital imager is a mono-cooled complementary camera with 4.65 µm (H) × 4.65 µm (V) individual pixel size (FLEA2, Point Grey Research, British Columbia).

The maximum lens resolution is 1392 (H) × 1042 (V) with the angle of view is 60º. The camera is placed below an optical waveguide. A heat-resistant borosilicate glass plate is placed between the camera and the optical waveguide to sustain an optical waveguide without losing camera resolution. The glass also functions as the bone and nail in a human finger. The internal light source is a micro-LED (Unique-Leds, Newalla, OK) with a diameter of 1.8 mm. There are four LEDs on the four sides of the waveguide to provide illumination. The direction and incident angle of the LED light has been calibrated with the cone of acceptance angle and is described in the next section.

### 3.3. Sensing Principle

Fig. 3 illustrates the conceptual diagram of the sensing principle. The tactile elasticity imaging sensor is developed based on the optical phenomenon known as total internal reflection (TIR) of light within an optical waveguide.

If two mediums have different indices of refraction, and the light is shone through those two mediums, then a fraction of light is transmitted and the rest is reflected. The amount of reflection is dependent on the angle of incidence. There is a critical angle above which the ray is completely reflected. The basic principle of the sensor system lies in the monitoring of the reflected light caused by the changing of the critical angle by contact. The intensity of the reflected light is related to the applied force and the strain on the optical waveguide. Here we investigate TIR in the multi-layer optical waveguide using ray optics approximation.

Consider light trapped inside the waveguide in the geometry shown in Fig. 4.

The basic design of the optical waveguide plate consists of three different refractive indices of PDMS. Consider three PDMS layers that are non-absorbing mediums (refractive index: $n_1$, $n_2$, $n_3$) on a heat resistant borosilicate glass plate (refractive index: $n_4$). The magnitude of refractive index is set as the highest at medium 4 and decreases toward to medium 1, i.e., $n_1 < n_2 < n_3 < n_4$. Medium 0 and medium 5 are air which is the absorption-free medium, i.e., $n_0 = n_5 = 1$. Assume that LED light sources are placed around the middle of the PDMS layers. Light is incident from the outside of PDMS layers and strikes each layer of the PDMS stack. Due to Snell's law the propagation angles $\gamma_i$ in each layer $i$, $i$ =1, 2, 3, 4 are bound by the following relations.

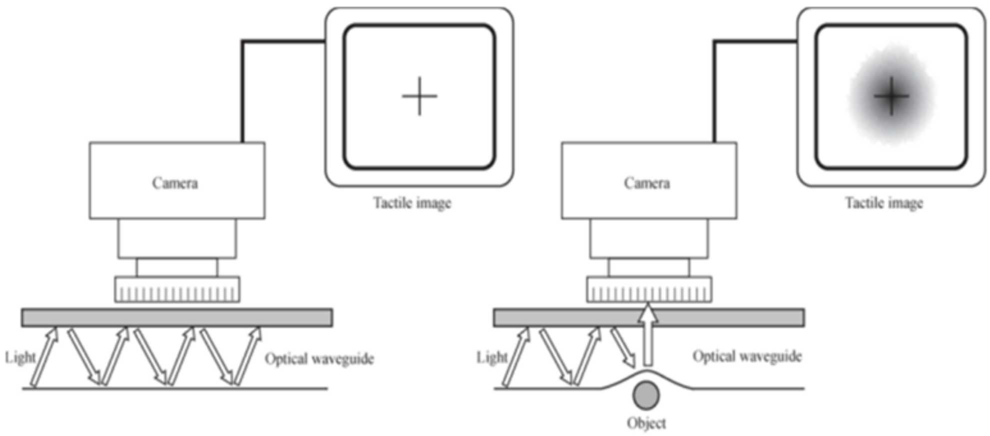



**Fig. 3.** Schematic diagram of the sensing principle. The light strays outside the waveguide as the optical waveguide deforms according to the applied force.

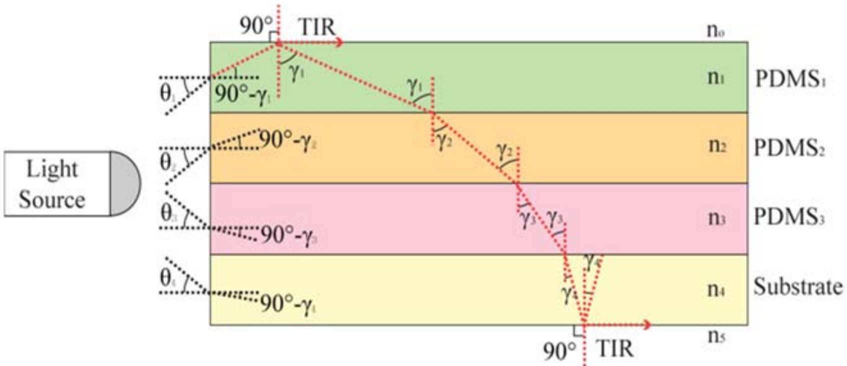



**Fig. 4.** Ray propagation inside the multi-layer optical waveguide under total internal reflection (TIR).

$$\begin{aligned} n_1 \sin\gamma_1 &= n_0 \sin\gamma_0 \\ n_2 \sin\gamma_2 &= n_1 \sin\gamma_1 \\ n_3 \sin\gamma_3 &= n_2 \sin\gamma_2 \\ n_4 \sin\gamma_4 &= n_3 \sin\gamma_3 \\ n_5 \sin\gamma_5 &= n_4 \sin\gamma_4 , \end{aligned} \tag{1}$$

where $n_0$ and $n_5$ are the refractive indices of air $n_0 = n_5 = 1$, and the critical TIR condition has been achieved when $\gamma_0 = \gamma_5 = 90^o$ at the boundaries with air. Light propagating in the waveguide with angles $\gamma_1$, $\gamma_2$, $\gamma_3$, $\gamma_4$ or higher in their respective layers will remain trapped inside the waveguide. The critical angle indicates the minimum propagation angle for TIR. To make the propagation angle below the critical angle in a waveguide, the acceptance angle for light sources has been calculated.

The acceptance angle is the maximum angle, under which light directed into the waveguide remains trapped inside. Angles $\gamma_1$, $\gamma_2$, $\gamma_3$, $\gamma_4$ are related to the acceptance angle $\theta_i$ in each layer *i* by the Snell's Law:

$$\begin{aligned} \sin\theta_i &= n_i \sin(\gamma_0 - \gamma_i) = n_i \sin(\gamma_5 - \gamma_i) = \\ &= n_i \sin(90^o - \gamma_i) = n_i \cos\gamma_i . \end{aligned} \tag{2}$$

Further, transforming Eq. (2), we obtain

$$\begin{aligned} \sin\theta_i &= n_i \cos\gamma_i = n_i (1 - \sin^2\gamma_i)^{1/2} = \\ &= (n_i^2 - n_i^2 \sin^2\gamma_i)^{1/2} . \end{aligned} \tag{3}$$

It follows from Eq. (1) that all $n_i \sin\gamma_i$ are equal to $n_0 \sin 90^o$, which is equal to 1 for air. Therefore, we finally have

$$\theta_i = \sin[(n_i^2 - 1)^{1/2}]^{-1} \tag{4}$$

for each layer $i$. Light, incident on layer $i$ under angle $\theta_i$ will be trapped inside the waveguide. For instance, if $n_1$, $n_2$, $n_3$, $n_4$ are measured approximately 1.38, 1.39, 1.40, 1.41, the acceptance angles, $\theta_i$, are calculated as $71.98^o$, $74.89^o$, $78.46^o$, $83.73^o$, respectively. Thus, for the TIR in the waveguide, the spatial radiation pattern of LED should be smaller than $71.98^o \times 2 = 143.96^o$. Fig. 5 shows the total internal reflection in a three layer PDMS optical waveguide using four LED light sources.

## 3.4. Sensor Specification

**Spatial resolution between sensing points:** The resolution of the tactile elasticity imaging sensor is going to be the pixel size of the camera. The spatial resolution between sensing points of the fingertip is at least 0.1 mm, which translates into an approximately 200×300 elements grid on a fingertip-sized area (20 mm×30 mm) [18]. In the current design, the pattern discrimination ability of the elasticity-imaging sensor is 4.65 µm and translates into an approximately 4301×6415 elements grid on a fingertip-sized area.

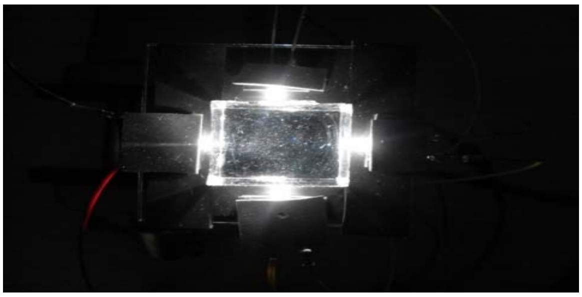


**Fig. 5.** The total internal reflection in a three layer PDMS optical waveguide using four LED light sources.

**Temporal resolution:** With regard to the human fingertip temporal resolution, the vibration bandwidth reported at the fingertips is a few Hz for separate touches and hundred Hz for continuous sensing. The camera that we chose had a 1392×1042 resolution at 80 frames per second (80 Hz). However, this temporal resolution can be improved depending on the camera.

**Force sensitivity:** Sensitivity is described in terms of the relationship between the physical signal (input) and the electrical signal (output) and is generally the ratio between a small change in the input to a small change in the output signal. The force sensitivity of the proposed tactile sensor force is approximately $2.5\times10^{-3}$ N compared to the fingertip force sensitivity of $2.0\times10^{-2}$ N [21].

**Linearity/hysteresis:** The human skin response has hysteresis. It has been noted that, for example, the force required maintaining a given indentation on the skin decreases as the probe is held against the skin [16]. The skin relaxes with time, with an observed length of up to 8 seconds. The proposed sensor is stable, repeatable and continuous in its variable output signal. The response of the sensor is non-hysteric and verified in the experiments. Table 1 summarizes the sensory specification of the human fingertip and our tactile elasticity imaging sensor.

**Table 1.** Sensory specification of the human fingertip and tactile elasticity imaging sensor.

| Design criteria | Human fing ertip | Tactile elasti city imaging sensor |
|---|---|---|
| Spatial resolution betw een sensing points | 0.1 mm | 4.65 µm |
| Temporal resolution | 0~100 Hz | 0~80 Hz |
| Force sensitivity | $2.5\times10^{-3}$ N | $2.0\times10^{-2}$ N |
| Hysteresis | High | Low |

## 4. Elasticity Estimation

The elasticity is obtained using stress and strain information from tactile images. Since the stress is measured as force per unit area, we estimate the applied force by contacting the object using the integrated pixel value of tactile image. For the strain information, the displacement of a material deformation under the applied force has been measured by tracking the control points extracted from two different tactile images. In this section, stress and strain estimation algorithms are been discussed in detail.

### 4.1. Stress Emotion

In this section, we present the stress estimation method using the obtained tactile image. If the optical waveguide of a sensor is compressed by the contacting object, it is deformed in both compressive and shearing directions. Because the light scatters at the contact area, the pixel value of the tactile image acquired by the camera distributes as a Gaussian function, in which the pixel intensity is the highest at the centroid and decreases with increased distance from the centroid [15]. Let $I(x, y)$ be the pixel value of the image plane. Since $I(x, y)$ is proportional to the contact stress, $P(x, y)$, caused by the contact between optical waveguide and the object, it can be expressed as follows:

$$P(x, y) = f(I(x, y)), \tag{5}$$

where $f$ is the conversion function which is determined by an experiment. If $S$ is designated as the contact area between the optical waveguide and the contact object, then the vertical force $F_z$ is obtained by integrating the stress over the contact area as follows

$$F_z = \int_S P(x, y)dS. \tag{6}$$

In order to determine horizontal force vectors $F_x$ and $F_y$, the $x$- and $y$- coordinates of the centroid $(X_c, Y_c)$, are calculated by

$$X_c = \int_S I(x, y)xdS / \int_S I(x, y)dS, \tag{7}$$

$$Y_c = \int_S I(x, y)ydS / \int_S I(x, y)dS. \tag{8}$$

The movements of the $x$- and $y$- components of the centroid are denoted as $u_x$ and $u_y$ and expressed as

$$u_x = X_c^{(t)} - X_c^{(t-1)}, \tag{9}$$

$$u_y = Y_c^{(t)} - Y_c^{(t-1)}. \tag{10}$$

where $t$ and $t-1$ represent current and prior steps, respectively. If friction between the optical waveguide and contact object is ignored, then $x$- and $y$- directional forces $F_x$ and $F_y$ are calculated as follows

$$F_x = K_x u_x, \tag{11}$$

$$F_y = K_y u_y., \tag{12}$$

where $K_x$ and $K_y$ are the $x$- and $y$- directional spring constants of the optical waveguide, respectively.

Spring constants are determined experimentally.

If we calculate the applied force, then the stress is calculated from the applied force per unit contact area.

### 4.1. Strain Estimation using Non-rigid Pattern Matching Algorithm

The strain measurement of a material under loading is achieved by tracking the displacement of control points extracted from series of tactile images. This concept is attractive, but manual measurement of control point positions is tedious and subject to error. This limits the number of control points that can be used and measured, and the spatial resolution of displacement fields. In this paper, to tracking the control points efficiently and automatically, non-rigid pattern matching algorithm is developed. The essence of this algorithm is to automatically measure displacements by tracking the change in position of control points. Fig. 6 represents the concept of tracking control points between tactile images. Considering a point of interest, $p$, in the image of the reference configuration. It is desired to determine $q$, the point to which it has moved in the image of the deformed configuration. Since the contacted object is a 3-D object and the tactile images are provided in 2-D, the 3-D surface image has been reconstructed from 2-D image using "Shape from Shading" method [22]. The control points are extracted from the reconstructed 3-D image.

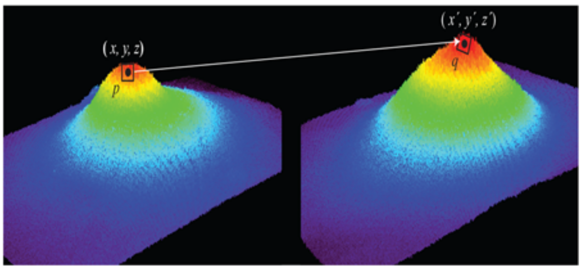



**Fig. 6.** The estimation of displacements of the contacted object deformation by tracking control points on the surface of the tactile image.

The non-rigid pattern matching algorithm we propose, uses iterative algorithm to find appropriate correspondence and transformation function between control points. The displacement of the object deformation is obtained from the transformation function. This displacement function is used to estimate the strain information. The algorithm is described in more detail below.

If we consider the first 3-D tactile image as a reference image and the second 3-D tactile image with different loading ratio, as the target image, then we can use the pattern matching method to obtain the strain information. From the surface of each 3-D tactile image, a random number of control points are extracted. Let $P=\{p_1, p_2, ..., p_I\}$ be a set of points in the model and $Q=\{q_1, q_2, ..., q_J\}$ be a set of points in the target. For a given point, $p_i \in P$, one can select neighboring points $N_a(p_i)$, $a$=1, 2, …, $A$, which reside in the circle centered at $p_i$. We set the radius of a circle as the median value of all Euclidean distances between point pairs in $P$. Similarly, for a point, $q_j \in Q$, adjacent points are $N_b(q_j)$, $b$=1, 2, …, $B$. In this paper, the pattern matching problem is formulated as a graph matching problem. Each point is a node of a graph, and a given point and its adjacent point constitute the edges of the graph. The problem then is to maximize the number of matched edges between two graphs. For this purpose, we determine the fuzzy correspondence matrix $M$. Each entry of $M$ has continuous value between [0,1] that indicates the weight of the correspondence between $p_i$ and $q_j$. The optimal match $\hat{M}$ is found by maximizing the energy function as follows:

$$\hat{M} = \arg\max_{M} C(P,Q,M), \tag{13}$$

where

$$C(P,Q,M) = \sum_{i=1}^{I}\sum_{b=1}^{B}\sum_{j=1}^{J}\sum_{a=1}^{A} M_{p_i q_j} M_{\mathcal{N}_a(p_i)\mathcal{N}_b(q_j)}. \tag{14}$$

The above equations are subject to the following constraints $\sum_{j=1}^{J+1} M_{p_i q_j} = 1$ for $i = 1,2,...,I$ and $\sum_{i=1}^{I+1} M_{p_i q_j} = 1$ for $j = 1,2,...,J$.

In the following section, we discuss how the optimal correspondence and transformation function between control points are obtained. For this purpose, the algorithm uses an iterated estimation framework to find appropriate correspondence and transformation.

### 4.2.1. Point Correspondence

Initially, each point $p_i \in P$ is assigned with a set of matching probability based on the shape context distance [22]. After the initial probability assignment, the relaxation labeling process updates the matching probability. The relaxation labeling is an iterative procedure that reduces local ambiguities and achieves global consistency by exploiting contextual information. The process is to assign a matching probability that maximizes $C(P,Q,M)$ under the relaxed condition of $M_{s_i t_j} \in [0,1]$. At the end of the relaxation labeling process, it is expected that each point will have one unambiguous matching probability. The determination of the compatibility coefficients is crucial because the performance of the relaxation labeling process depends on them. We propose new compatibility coefficient that quantifies the degree of agreement between the hypothesis that $p_i$ matches to $q_j$ and $\mathcal{N}_a(p_i)$ matches to $\mathcal{N}_b(q_j)$.

In the non-rigid degradation of point sets, we note that a point set is usually distorted; however, the neighboring structure of a point is generally preserved due to physical constraints. The displacement of a point and its adjacent point between two point sets constrain one another. Thus, if the distance and angle of a point pair $(p_i, \mathcal{N}_a(p_i))$ in the reference image and its corresponding point pair $(q_j, \mathcal{N}_b(q_j))$ in the target image are similar, we say that they have high correlation. This is further strengthened if a point pair $(p_i, \mathcal{N}_a(p_i))$ in the model shape is closer to each other. To quantify this knowledge, we introduce the similarity constraint $\alpha$, $\beta$ as well as the spatial smoothness constraint $\gamma$.

The first constraint is the similarity that is related to the differences between the distances and angles of $(p_i, \mathcal{N}_a(p_i))$ and $(q_j, \mathcal{N}_b(q_j))$. This first constraint imposes that if $(p_i, \mathcal{N}_a(p_i))$ has smaller distance and angle differences with $(q_j, \mathcal{N}_b(q_j))$, then they are more compatible. The disparities between $(p_i, \mathcal{N}_a(p_i))$ and $(q_j, \mathcal{N}_b(q_j))$ are defined as follows.

$$\alpha(p_i, \mathcal{N}_a(p_i); q_i, \mathcal{N}_b(q_j)) = \\ = \left(1 - \left|(d_i(p_i, \mathcal{N}_a(p_i)) - d_j(q_i, \mathcal{N}_b(q_j))) / \max_{i,j}\{d_i(p_i, \mathcal{N}_a(p_i)), d_j(q_j, \mathcal{N}_b(q_j))\}\right|\right), \tag{15}$$

$$\beta(p_i, \mathcal{N}_a(p_i); q_i, \mathcal{N}_b(q_j)) = \\ = \left(1 - \left|(l_i(p_i, \mathcal{N}_a(p_i)) - l_j(q_i, \mathcal{N}_b(q_j))) / \max_{i,j}\{l_i(p_i, \mathcal{N}_a(p_i)), l_j(q_j, \mathcal{N}_b(q_j))\}\right|\right), \tag{16}$$

The second constraint, spatial smoothness, is measured by the distance between $p_i$ and $\mathcal{N}_a(p_i)$.

$$\gamma(p_i, \mathcal{N}_a(p_i)) = \left(1 - d_i(p_i, \mathcal{N}_a(p_i)) / \max_i\{d(p_i, \mathcal{N}_a(p_i))\}\right), \tag{17}$$

where $\max_i\{d(p_i, \mathcal{N}_a(p_i))\}$ is the longest edge of point-adjacent point pairs. We define a total compatibility coefficient by

$$r_{p_i q_j}(\mathcal{N}_a(p_i), \mathcal{N}_b(q_j)) = \alpha(p_i, \mathcal{N}_a(p_i); q_j, \mathcal{N}_b(q_j)) \cdot \beta(p_i, \mathcal{N}_a(p_i); q_j, \mathcal{N}_b(q_j)) \cdot \gamma(p_i, \mathcal{N}_a(p_i)). \tag{18}$$

The support function $S^k_{p_i q_j}$ in the *k*-th iteration is given by

$$\begin{aligned} S^k_{p_i q_j} &= \sum_{i=1}^{I} \sum_{j=1}^{J} r_{p_i q_j}(\mathcal{N}_a(p_i), \mathcal{N}_b(q_j)) M^k_{\mathcal{N}_a(p_i)\mathcal{N}_b(q_j)} \\ &= \sum_{i=1}^{I} \sum_{j=1}^{J} \alpha(p_i, \mathcal{N}_a(p_i); q_j, \mathcal{N}_b(q_j)) \cdot \beta(p_i, \mathcal{N}_a(p_i); q_j, \mathcal{N}_b(q_j)) \cdot \gamma(p_i, \mathcal{N}_a(p_i)) \cdot M^k_{\mathcal{N}_a(p_i)\mathcal{N}_b(q_j)}. \end{aligned} \tag{19}$$

Finally, the fuzzy correspondence matrix element $M_{p_i q_j}$ in Eq. (14) is updated according to

$$M^{k+1}_{p_i q_j} = M^k_{p_i q_j} S^k_{p_i q_j} / \sum_{j=1}^{J} M^k_{p_i q_j} S^k_{p_i q_j}, \tag{20}$$

Traditionally, the sum of the rows (or columns) of the matrix $M$ is used as a constraint in the relaxation labeling process. In this paper, we use the sum of the rows and columns as a two-way constraint. In order to meet these constraints, alternated row and column normalization of the matrix $M$ is performed after each relaxation labeling update. This procedure is known as Sinkhorn normalization, which shows that the procedure always converges to a doubly stochastic matrix [25].

### 4.2.2. Transformation Function

Since the strain is determined by the displacements of the contact object deformation, we estimate a transformation function $T: \Re^3 \to \Re^3$ to find the displacements between tactile images obtained under different loading values. In this study, we use the thin-plate spline (TPS) model, which is used in non-rigid pattern matching method for representing flexible coordinate transformations.

Let $v_i$ denote the target function values at corresponding locations $p_i = (x_i, y_i, z_i)$ in the plane, with $i = 1, 2, ..., n$. In particular, we will set $v_i$ equal to $x_i'$, $y_i'$, $z_i'$ in turn to obtain one continuous transformation for each coordinate. We assume that the locations $(x_i, y_i, z_i)$ are all different and are not collinear. In 3-D interpolation problem, the TPS interpolant $f(x, y, z)$ minimizes the bending energy

$$I_f = \iiint_{\mathbb{R}^3} \left[ \left(\frac{\partial^2 f}{\partial x^2}\right)^2 + \left(\frac{\partial^2 f}{\partial y^2}\right)^2 + \left(\frac{\partial^2 f}{\partial z^2}\right)^2 + 2\left( \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 + \left(\frac{\partial^2 f}{\partial x \partial z}\right)^2 + \left(\frac{\partial^2 f}{\partial y \partial z}\right)^2 \right) \right] dxdydz, \tag{21}$$

and the interpolant form is

$$\begin{aligned} f(x, y, z) &= a_1 + a_x x + a_y y + a_z z + \\ &= \sum_{i=1}^{n} w_i U(\|(x_i, y_i, z_i) - (x, y, z)\|), \end{aligned} \tag{22}$$

where $a_1$, $a_x$, $a_y$, $a_z$ are the affine transformation coefficients and $w_i$ is the non-affine deformation coefficients. The kernel function $U(r)$ is defined by $U(r) = r^2 \log r^2$ and $U(0) = 0$ as usual. In order for $f(x, y, z)$ to have square integrable second derivative, we require the boundary condition as $\sum_{i=1}^{n} w_i = 0$ and $\sum_{i=1}^{n} w_i x_i = \sum_{i=1}^{n} w_i y_i = \sum_{i=1}^{n} w_i z_i = 0$.

Together with the interpolation conditions, $f(x_i, y_i, z_i) = v_i$, this yields a linear system for the TPS coefficients:

$$\begin{pmatrix} K & G \\ G^T & 0 \end{pmatrix}\begin{pmatrix} W \\ A \end{pmatrix} \triangleq L\begin{pmatrix} W \\ A \end{pmatrix} = \begin{pmatrix} V \\ 0 \end{pmatrix} = Y,$$

where $K = \begin{bmatrix} 0 & U(r_{12}) & \cdots & U(r_{1n}) \\ U(r_{21}) & 0 & \cdots & U(r_{2n}) \\ \cdots & \cdots & \cdots & \cdots \\ U(r_{n1}) & U(r_{n2}) & \cdots & 0 \end{bmatrix}_{n\times n}$

$$\text{and } G = \begin{bmatrix} 1 \; xx_1 & y_1 & z_1 \\ 1 \; xx_2 & y_2 & z_2 \\ \cdots \; \cdots & \cdots & \cdots \\ 1 \; xx_n & y_n & z_n \end{bmatrix}_{n\times 4} \cdot, \tag{23}$$

where, $r_{ij} = \| p_i - p_j \|$ is the Euclidean distance between points $p_i$ and $p_j$. $W$ and $A$ are column vectors formed from $W = [w_1, w_2, ..., w_n]^T$ and $A = [a_1, a_x, a_y, a_z]^T$, respectively. $V = [v_1, v_2, ..., v_n]^T$ is an arbitrary n-vector. We denote the $(n+4)\times(n+4)$ matrix $\begin{pmatrix} K & G \\ G^T & 0 \end{pmatrix}$ by $L$. Define the vector $Y = (V \mid 0 \;\; 0 \;\; 0)^T$, then we can find the coefficients of TPS by $(W \mid a_1 \;\; a_x \;\; a_y \;\; a_z)^T = L^{-1}Y$ [26].

In the point matching problem, it is necessary to relax the exact interpolation by means of regularization. This is accomplished by minimizing the bending energy as follows.

$$H[f] = \sum_{i=1}^{n}(v_i - f(x_i, y_i, z_i))^2 + \lambda I_f \tag{24}$$

The regularization parameter $\lambda$, a positive scalar, controls the amount of smoothing; the limiting case of $\lambda = 0$ reduces to exact interpolation. As demonstrated in [27-28] we can solve for the TPS coefficients in the regularized case by replacing the matrix $K$ with $K + \lambda I$.

After several relaxation labeling updates, the parameters of TPS deformation model is estimated from the matched control points. The estimated parameters are then used to transform the reference image bringing it as close as possible to the target image. The relaxation labeling process then starts again between the transformed model set and the target set. The processes for identifying correspondence and transformation are alternatively iterated until the stopping criterion is met.

The resulting function $f(x, y, z) = [f_x(x, y, z), f_y(x, y, z), f_z(x, y, z)]$ is a vector-valued, and this maps each point $(x_i, y_i, z_i)$ to $(x_i', y_i', z_i')$ and is the least bent of all such functions. These vector-valued functions $f(x, y, z)$ are the thin-plate spline mappings.

### 4.3. Elasticity Measurement from Stress and Strain

Now we obtain the elasticity from the tactile images using previous stress and strain estimation method. The final TPS model $f(x, y, z) = [f_x(x, y, z), f_y(x, y, z), f_z(x, y, z)]$ provides a continuous displacement field for performing strain analysis. The nonlinear Lagrangian strain tensor components are then determined from the equations as follows

$$e_{xx} = \frac{\partial f_x(x,y,z)}{\partial x} + \frac{1}{2}\left[\left(\frac{\partial f_x(x,y,z)}{\partial x}\right)^3 + \left(\frac{\partial f_y(x,y,z)}{\partial x}\right)^3 + \left(\frac{\partial f_z(x,y,z)}{\partial x}\right)^3\right], \tag{25}$$

$$e_{yy} = \frac{\partial f_y(x,y,z)}{\partial y} + \frac{1}{2}\left[\left(\frac{\partial f_x(x,y,z)}{\partial y}\right)^3 + \left(\frac{\partial f_y(x,y,z)}{\partial y}\right)^3 + \left(\frac{\partial f_z(x,y,z)}{\partial y}\right)^3\right], \tag{26}$$

$$e_{zz} = \frac{\partial f_z(x,y,z)}{\partial z} + \frac{1}{2}\left[\left(\frac{\partial f_x(x,y,z)}{\partial z}\right)^3 + \left(\frac{\partial f_y(x,y,z)}{\partial z}\right)^3 + \left(\frac{\partial f_z(x,y,z)}{\partial z}\right)^3\right], \tag{27}$$

$$\begin{aligned} e_{xy} &= \frac{1}{2}\left(\frac{\partial f_x(x,y,z)}{\partial y} + \frac{\partial f_y(x,y,z)}{\partial x}\right) + \\ &+ \frac{1}{2}\left[\left(\frac{\partial f_x(x,y,z)}{\partial x}\right)\left(\frac{\partial f_x(x,y,z)}{\partial y}\right) + \left(\frac{\partial f_y(x,y,z)}{\partial x}\right)\left(\frac{\partial f_y(x,y,z)}{\partial y}\right) + \left(\frac{\partial f_z(x,y,z)}{\partial x}\right)\left(\frac{\partial f_z(x,y,z)}{\partial y}\right)\right], \end{aligned} \tag{28}$$

$$e_{yz} = \frac{1}{2}\left(\frac{\partial f_z(x,y,z)}{\partial y} + \frac{\partial f_y(x,y,z)}{\partial x}\right) + \\ + \frac{1}{2}\left[\left(\frac{\partial f_x(x,y,z)}{\partial y}\right)\left(\frac{\partial f_x(x,y,z)}{\partial z}\right) + \left(\frac{\partial f_y(x,y,z)}{\partial y}\right)\left(\frac{\partial f_y(x,y,z)}{\partial z}\right) + \left(\frac{\partial f_z(x,y,z)}{\partial y}\right)\left(\frac{\partial f_z(x,y,z)}{\partial z}\right)\right], \tag{29}$$

$$e_{zx} = \frac{1}{2}\left(\frac{\partial f_x(x,y,z)}{\partial z} + \frac{\partial f_z(x,y,z)}{\partial x}\right) + \\ + \frac{1}{2}\left[\left(\frac{\partial f_x(x,y,z)}{\partial x}\right)\left(\frac{\partial f_x(x,y,z)}{\partial z}\right) + \left(\frac{\partial f_y(x,y,z)}{\partial x}\right)\left(\frac{\partial f_y(x,y,z)}{\partial z}\right) + \left(\frac{\partial f_z(x,y,z)}{\partial x}\right)\left(\frac{\partial f_z(x,y,z)}{\partial z}\right)\right]. \tag{30}$$

To determine the elastic property of the contacted object from the uniaxial loading configuration, the strain components are averaged $e_{zz}$ along the $x$ and $y$ directions to yield the average strain, $\overline{e}_{zz}$. Given the applied normal stress $S_{zz}$ acting on the loading surface of the optical waveguide, the elastic modulus $E$ is then determined from

$$E = S_{zz} / \overline{e}_{zz}. \tag{31}$$

The elasticity estimation using the non-rigid pattern matching algorithm is based on the NP-hard problem and has similar computational complexity of $O(N^3)$ for matching in $\mathbb{R}^3$. For a Young's modulus calculation using a $105 \times 105$ point matching, the algorithm takes about 1.69 seconds on a desktop PC with Core 2 Duo CPU with 2.13 GHz and 2 GB RAM.

## 5. Experiment Results

### 5.1. Normal Force Estimation Experiment

In this section, the relationship between the normal force and the integrated pixel value is established via experiments with a loading machine. The loading machine has a force/torque gauge (Mecmesin, West Sussex, UK) to detect the normal force. This machine is shown in Fig. 7.

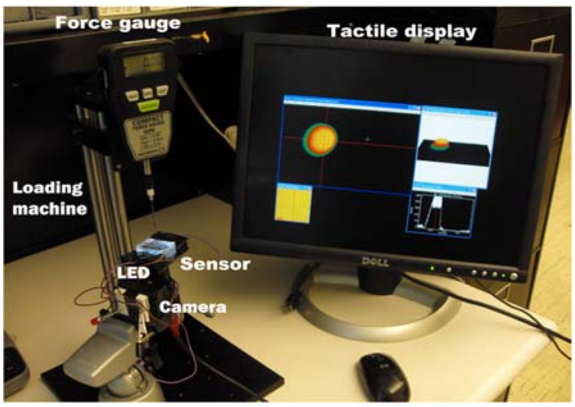



**Fig. 7.** Measurement setup for elasticity-imaging sensor characterization.

The force gauge has a probe to measure the force from 0 to 50 N with a resolution of $1.0 \times 10^{-3}$ N. Since the camera is an 8-bit digital imager, each pixel value is between of 0 and 255. A circular tip with 2 mm radius is attached to the force/torque gauge and this is used to contact the sensor. To validate the normal force detection, we start from the initial load of 0 N, then the normal force is increased in a stepwise manner. When the applied force reaches around 2.0 N, the applied normal force is decreased in a stepwise fashion until it returns to 0 N. The resulting diffused light is captured by the camera, and the corresponding contact force is measured by the force gauge.

Fig. 8 shows the pixel value along the contact area's horizontal line passing through the centroid of tactile image. As we expected, the graph is the Gaussian like bell shaped graph and the maximum value is on the centroid of the tactile image. The plot of integrated pixel value change as the applied force changes is shown in Fig. 9 (a). The relationship between the integrated pixel value and the applied force is found to be approximately linear as shown in Fig. 9 (b). The approximated curve shows a monotone increasing relationship between the normal force and the integrated pixel value of the tactile image. The hysteresis loop is not observed in the graph, indicating that the proposed sensor functions as the precise load cell in the region from 0 to 2 N. Using this approximation, we can approximate the normal forces from the integrated pixel values.

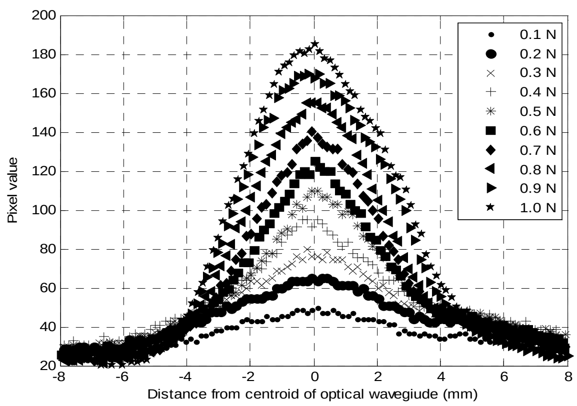
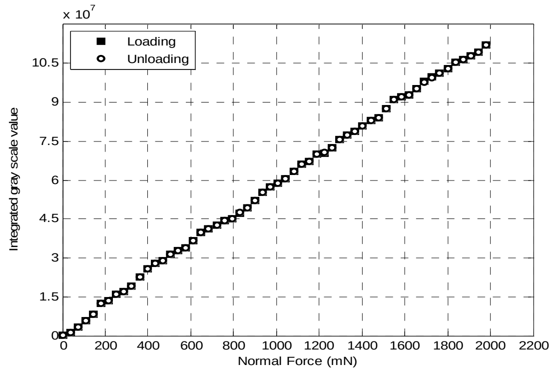
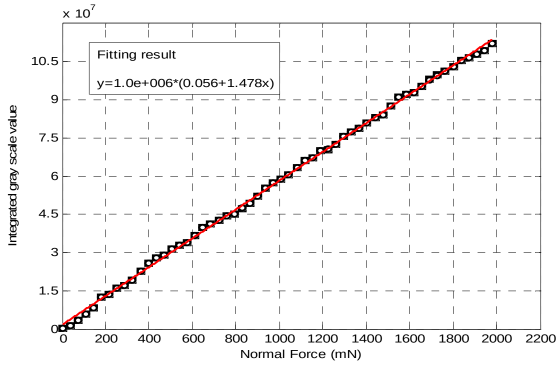



**Fig. 8.** Pixel value along the contact area's horizontal line passing through the centroid of tactile image.




(a)

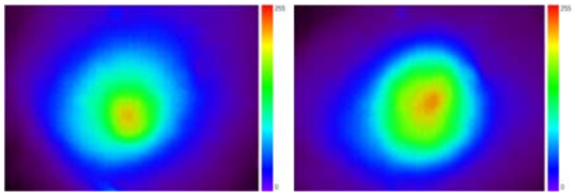



(b)

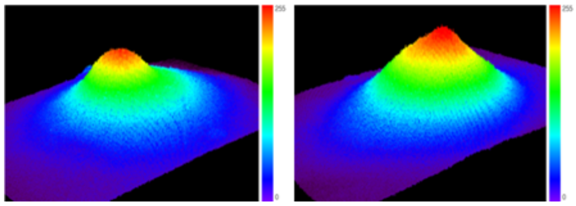
**Fig. 9.** The relationship between normal force and integrated gray scale value: (a) Loading and unloading experimental results; (b) the approximated fitting curve.

## 5.2. Strain and Elasticity Measurement Using Soft Polymers

To validate the elasticity measurement using the proposed sensor, Versaflex CL2000X and CL2003X (GLS, McHenry, Illinois) soft polymers with known Young's moduli of 103 kPa and 62 kPa have been used. The objects was 3 mm in radius and spherical in shape. The tactile elasticity imaging sensor compressed the polymer samples. The compression ratio was gradually increased. At 0.7 N and 1.2 N applied forces, tactile image has been taken. Fig. 10(a) shows two 2-D tactile images under the 0.7 N and 1.2 N normal forces. In the images, a color scale replaced the original grayscale for better visualization. A purple color indicates grayscale value 0 and a red color indicates grayscale value 255. The two obtained 2-D tactile images were rendered to 3-D images using "shape from shading" method [23]. The 3-D rendered tactile images are represented in Fig. 10(b). The 200 control points were then sampled from the surface of 3-D tactile images. In this the equally spaced control points are extracted automatically. The point correspondence and transformation between control points are iteratively estimated. Fig. 11(a) represents control point distributions from 0.7 N and 1.2 N 3-D tactile images of Vesaflex CL2000X. The final matching result is represented in Fig. 11(b).

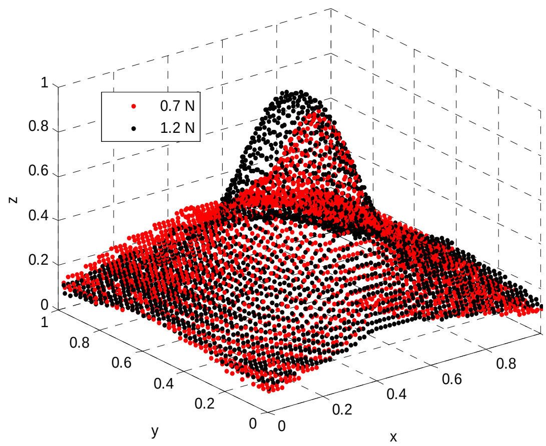


(a)

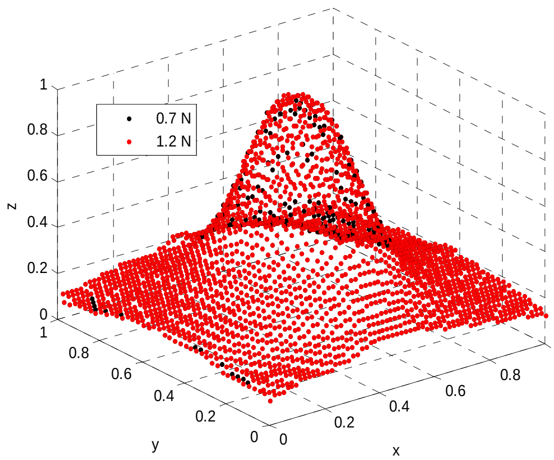


(b)

**Fig. 10.** 2-D tactile images and 3-D rendered tactile images: a) 2-D tactile images under 0.7 N (left) and 1.2 N (right) loading value; (b) 3-D recovery image of 0.7 N (left) and 1.2 N (right).




(a)




(b)

**Fig. 11.** Non-rigid pattern matching under different tactile image of CL2000X: (a) Control points from 3-D tactile images under the loading values of 0.7 N and 1.2 N to polymer sample, CL2000X; (b) The non-rigid pattern matching results.

The TPS transformation functions from the final matching result are used for the elasticity determination. Fig. 12 represents the experimental verification. The solid line represents the gold standard of CL2000X and CL2003X moduli, and the square represents measurement values from a tactile elasticity imaging sensor. The errors of the estimated moduli were within 4.23 % for CL2000X and 5.38 % for CL2003X.

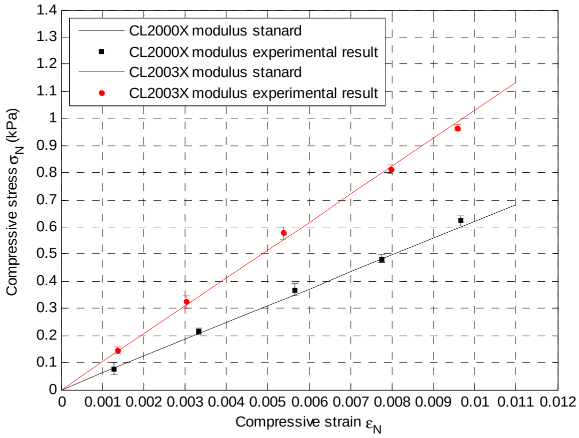



**Fig. 12.** Polymer samples CL2000X and CL2003X moduli measurements using tactile elasticity imaging sensor.

## 6. Conclusions

In this paper, a tactile elasticity imaging sensor using the total internal reflection principle is designed and experimentally evaluated. To increase the sensing range, an optical waveguide consisting of three different densities of PDMS with different elastic modulus was fabricated. In order to obtain the elasticity of the sensed object, the strain is estimated by a non-rigid pattern matching technique. The performance of the proposed sensor is experimentally verified. The results show that the elastic moduli are estimated within 5.38 % of the true value.

## References


[1]. Y. C. Fung, Biomecanics: Mechanical properties of living tissues, 2nd ed., *Springer-Verlag*, New York, 1993.

[2]. T. A. Krouskop, T. M. Wheeler, F. Kallel, B. S. Garra, T. Hall, Elastic moduli of breast and prostate tissues under compression, *Ultrasonic Imaging*, Vol. 20, 1998, pp. 260-274.

[3]. H. Shojaku, H. Seto, H. Iwai, S. Kitazawa, W. Fukushima, K. Saito, Detection of incidental breast tumors by noncontrast spiral computed tomography of the chest, *Radiation Medicine*, Vol. 26, No. 6, 2008, pp. 362-367.

[4]. H. Degani, V. Gusis, D. Weinstein, S. Fields, S. Strano, Mapping pathophysiological features of breast tumors by MRI at high spatial resolution, *Nature Medicine*, Vol. 3, 1997, pp. 780-782.

[5]. C. R. Gentle, Mammobarography: a possible method of mass breast screening, *Journal of Biomedical Engineering*, Vol. 10, 1988, pp. 124-126.

[6]. S. Omata, Y. Terunuma, New tactile sensor like human hand and its applications, *Sensors and Actuators A: Physical*, Vol. 35, No. 1, 1992, pp. 9-15.

[7]. M. F. Barsky, D. K. Lindner, R. O. Claus, Robot gripper control system using PVDF piezoelectric sensors, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. 36, No. 1, Jan 1989, pp. 129-134.

[8]. K. Motoo, F. Arai, T. Fukuda, Piezoelectric vibration-type tactile sensor using elasticity and viscosity change of structure, *IEEE Sensors Journal*, Vol. 7, 2007, pp. 1044-1051.

[9]. L. Liu, X. Zheng, Z. Li, An array tactile sensor with piezoresistive single crystal silicon diaphragm, *Sensors and Actuators A: Physical*, Vol. 32, No. 3, 1993, pp. 193-196.

[10]. B. J. Kane, A high resolution traction stress sensor array for use in robotic tactile determination, Ph.D. Dissertation, *Stanford Univ.*, Stanford, CA, 1999.

[11]. H. Morimura, S. Shigematsu, K. Machinda, A novel sensor cell architecture and sensing circuit scheme for capacitive fingerprint sensors, *IEEE Journal Solid-State Circuits*, Vol. 35, No. 5, 2000, pp. 724-731.

[12]. D. J. van den Heever, K. Schreve, C. Sheffer, Tactile sensing using force sensing resistors and a super-resolution algorithm, *IEEE Sensors Journal*, Vol. 9, No. 1, 2009, pp. 29-35.

[13]. H. Yegingil, W. Y. Shih, W.-H. Shih, All-electrical indentation shear modulus and elastic modulus measurement using a piezoelectric cantilever with a tip, *Journal of Applied Physics*, Vol. 101, 2007.

[14]. S. Najarian, J. Dargahi, V. Mirjalili, Detecting embedded objects using haptics with applications in artificial palpation of tumors, *Sensors and Materials*, Vol. 18, No. 4, 2006.

[15]. M. Ohka, H. Kobayashi, J. Takata, T. Mitsuya, An experimental optical three-axis tactile sensor featured with hemispherical surface, *Journal of Advanced Mechanical* Design, *Systems, and Manufacturing*, Vol. 2, No. 5, 2008, pp. 860-873.

[16]. S. Saga, H. Kajimoto, S. Tachi, High-resolution tactile sensor using the deformation of a reflection image, *Sensor Review*, Vol. 27, No. 1, 2007, pp. 35-42.

[17]. W. R. Uttal, The Psychology of Sensory Coding, *Harper and Row Publishing Co*, 1973.

[18]. E. Kandel, J. Schwartz, T. Jessell, Principles of Neural Science, *McGraw-Hill Medical*, 2000.

[19]. G. S. Rajan, G. S. Sur, J. E. Mark, D. W. Schaefer, G. Beaucage, Preparation and characterization of some unusually transparent poly (dimethylsiloxane) nanocomposites, *Journal of Polymer Science*, Vol. 41, No. 16, 2003, pp. 1897-1901.

[20]. D. A. Chang-Yen, R. K. Eich, B. K. Gale, A monolithic PDMS waveguide system fabricated using soft-lithography techniques, *Journal of Lightwave Technology*, Vol. 23, No. 6, 2005, pp. 2088-2093.

[21]. K. L. Johnson, Contact Mechanics, *Cambridge University Press*, 1985.

[22]. J. Petit, Y. Galifret, K. Ichimaru, Artificial softness sensing - an automatic apparatus for measuring viscoelasticity, *Mechanism and Machine Theory*, 12, 1977, pp. 11–26.

[23]. P. Tsai, M. A. Shah, Shape from shading with variable albedo, *Optical Engineering*, Vol. 37, No. 4, 1998, pp. 1212-1220.

[24]. S. Belongie, J. Malik, J. Puzicha, Shape matching and

object recognition using shape contexts, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 24, No. 4, Apr. 2002, pp. 509-522.
[25]. R. Sinkhorn, A relationship between arbitrary positive matrices and doubly stochastic matrices, *The Annals of Mathematical Statistics*, Vol. 35, No. 2, 1964, pp. 876-879.
[26]. M. J. D. Powell, A Thin Plate Spline Method for Mapping Curves into Curves in Two Dimensions, *Computational Techniques and Applications*, 1995.
[27]. H. Chui, A. Rangarajan, A new point matching algorithm for non-rigid registration, *Computer Vision and Image Understanding*, Vol. 89, No. 23, 2003, pp. 114-141.
[28]. F. L. Bookstein, Principal warps: Thin-plate Splines and the decomposition of deformations, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 11, No. 6, Jun. 1989, pp. 567-585.

*Sensors & Transducers, Vol. 212, Issue 5, May 2017, pp. 70-78*

**Sensors & Transducers**

Published by IFSA Publishing, S. L., 2017
http://www.sensorsportal.com

# A Highly Selective Room Temperature $NH_3$ Gas Sensor based on Nanocrystalline α- $Fe_2O_3$

**Priyanka A. PATIL, Dhanashri G. PATIL, Vinita V. DEO, and * Lalchand A. PATIL**

Nanomaterials Research Laboratory, Department of Physics Pratap College Amalner,
Dist.: Jalgaon, Maharashtra, 425401, India
Tel.: +91-2587-223101/223103, fax: (02587) 223101
E-mail: plalchand_phy_aml@yahoo.co.in

*Received: 1 May 2017 /Accepted: 29 May2017 /Published: 31 May 2017*

**Abstract:** Nanocrystalline α-$Fe_2O_3$ powder was synthesized by simple, inexpensive sol-gel method. The obtained powder was calcined at 700 $^0$C in air atmosphere for 2 hours. The structural and morphological properties of calcined powder were studied by X-ray diffraction (XRD) and Field Emission Scanning Electron Microscopy (FESEM) respectively. Thermal properties of dried gel were studied by Thermogravimetric Analysis/Differential Scanning Calorimetry (TGA/DSC). The XRD pattern of the powder confirmed the α-$Fe_2O_3$ (hematite) phase of iron oxide with average crystalline size of 30.87 nm calculated from Scherrer equation. The FESEM images showed uniform wormlike morphology of α-$Fe_2O_3$ powder. TGA result indicated that α-$Fe_2O_3$ is thermodynamically stable. Room temperature $NH_3$ sensing characteristics of α-$Fe_2O_3$ were studied for various concentration levels (250-2500 ppm) of $NH_3$ at various humid conditions. The sensor based on α-$Fe_2O_3$ exhibited good selectivity and excellent sensitivity (S=92) towards 1000 ppm of $NH_3$ with quick response of 4 sec and fast recovery of 9 sec. Room temperature sensing mechanism is also discussed.

**Keywords:** α-$Fe_2O_3$, Sol-gel, $NH_3$ sensing, Room temperature.

## 1. Introduction

The huge industrial development all over the world leads to increase in environmental pollution by releasing hazardous gases and vapors in the surrounding. $NH_3$ plays an important role in all forms of life. It is naturally produced in living organisms. $NH_3$ is widely used in the production of nitrogenous fertilizers, chemicals and petrochemical industries etc. It is also used as an industrial refrigerant. $NH_3$ is colorless gas with pungent smell. It is very toxic and can cause burns on the skin, in the eyes, in the respiratory system and in the lungs which can lead permanent blindness and lung diseases [1-4]. Therefore, it is necessary to develop $NH_3$ sensors with fast response time and low detection limit at room temperature.

Previous research shows $NH_3$ can be detected by using potentiometric electrode, infrared devices, optical fibers, conducting polymers and metal oxides [5, 6]. Among all these, metal oxide gas sensors are widely used because of their simplicity of production and use, low cost and capability of detecting wide range of toxic gases and vapors under different conditions. Various metal oxide semiconductors such as $SnO_2$, ZnO, $V_2O_5$, ITO, $WO_3$ and $TiO_2$ are being used as $NH_3$ sensors [7].

α-$Fe_2O_3$ (hematite) is one of the promising metal oxides due to its potential applications in many fields such as gas sensors, lithium batteries, catalysis, pigment and magnetic devices [8-12]. It is thermodynamically most stable phase of iron oxides. α-$Fe_2O_3$ exhibits n-type semiconducting properties with an energy gap of 2.2 eV [13]. Various techniques have been employed for the synthesis of α-$Fe_2O_3$ nanocrystalline powder such as chemical coprecipitaion, hydrothermal method and sol-gel technique etc. Here sol-gel method is used for synthesis of α-$Fe_2O_3$ nanocrystalline powder because it is simple and cost effective [8, 14, 15].

Most of the metal oxide $NH_3$ gas sensors are operated effectively only at higher temperatures (150-400$^0$C), which results in high power consumption. Hence it is necessary to develop gas sensors that would have good results at room temperature [16]. But from available literature it is rare to find unmodified α-$Fe_2O_3$ as $NH_3$ sensor at room temperature with excellent sensitivity, good selectivity and quick response and recovery times.

In the present investigation, experiments have been carried out for the fabrication of room temperature $NH_3$ sensor based on α-$Fe_2O_3$. The sensing characteristics of α-$Fe_2O_3$ are reported.

## 2. Materials and Methods

### 2.1. Materials

All chemicals were used of analytical grade. Chemicals used for synthesis of α-$Fe_2O_3$ powder were ferric nitrate (Fe $(NO_3)_3.9H_2O$), ethanol, triethylamine (TEA) and glycerol.

### 2.2. Synthesis of α-$Fe_2O_3$ Powder

0.1 M of Fe $(NO_3)_3.9H_2O$ was dissolved in 20 ml ethanol under permanent magnetic stirring at 350-400 rpm. The temperature of the solution was maintained at 60 °C using water bath. After complete dissolution orange colored transparent sol was obtained. Subsequently 6 ml of TEA was added to the solution drop by drop. The color of solution rapidly changed from orange to a reddish brown accompanied by appreciable heat release and the dark brown color gel was set within 30 minutes. After two hours, the gel was completely dried. The dried gel was calcined at 700 °C in air atmosphere for 2 hours to obtain the α-$Fe_2O_3$ powder.

### 2.3. Characterization Techniques

The structural property of the synthesized powder was investigated using X-ray diffractometer (Miniflex 600, Rigaku) with CuKα1 radiation (λ= 1.5406 Å). The morphology of powder was studied using field emission scanning electron microscopy (FESEM model S4800 type II, Hitachi High Technologies). Quantitative analysis was carried out using energy dispersive analysis by X-rays (EDAX model Bruker Nano GmbH X Flash Detector 5030) equipped with FESEM. Thermo gravimetric analysis (TGA) of the dried gel (xerogel) was done by STA-6000 (Perkin Elmer) in the temperature range of 50-800 °C under nitrogen atmosphere.

### 2.4. Fabrication of α-$Fe_2O_3$ Gas Sensor and Gas Sensing System

#### 2.4.1. Fabrication of α-$Fe_2O_3$ Gas Sensor

α-$Fe_2O_3$ gas sensor was fabricated by screen printing technique. The thixotropic paste of α-$Fe_2O_3$ powder was prepared by mixing the fine powder with a temporary organic binder (1-acetoxy 2-butoxy ethane +2-butoxyethanol + terpineole + ethyl cellulose). The ratio of inorganic part to organic part was kept 75:25 in formulating the paste. This paste was screen printed on the glass substrates in desired patterns. Thus obtained thick films were fired at 500 $^0$C for 30 minutes. The obtained thick films were termed as α-$Fe_2O_3$ sensor.

#### 2.4.2. Gas Sensing Setup

To study the changes in electrical properties of the gas sensor, electrical contacts were made on film surface using copper wire and silver paste. The gas sensing set-up is as shown in Fig. 1.

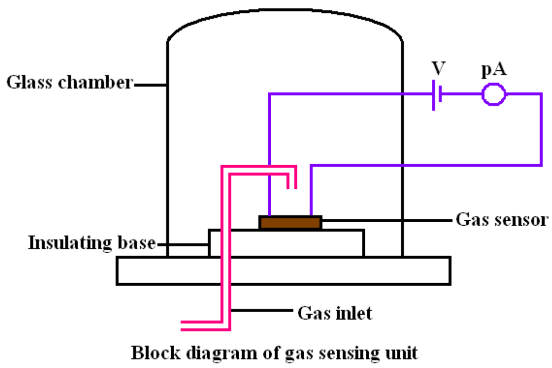



**Fig. 1.** Block diagram of gas sensing system.

The gas sensor was kept inside the glass chamber. A constant voltage (2V) was applied to the electrodes of α-$Fe_2O_3$ sensor and the current flowing through it was measured with picoammeter. The known volume of gas was injected into the glass chamber. The current before exposure of target gas ($I_a$) and the current on exposure of target gas ($I_g$) was measured. The performance of gas sensor was expressed in terms of gas response (S) and is defined as the ratio of change in conductance of the gas sensor on the exposure of target gas to the conductance before exposure of target gas [17]. Gas sensing study is carried out at room temperature ($R_T$) which is about 35 °C.

$$S = \frac{I_g}{I_a} \quad (1)$$

### 2.4.3. Gas Concentration Calculation

The desired gas concentration of the gas was calculated by using equation (2) [18]

$$V = \frac{Cppm \times Va \times M}{24.5 \times 10^9 \times D}, \quad (2)$$

where V is the required liquid volume, $C_{ppm}$ is gas concentration in ppm, $V_a$ is volume of test chamber, M is molecular weight of liquid, D is the density of liquid.

All units were measured in mililitres, cubic centimeters and grams.

The appropriate volume of liquor $NH_3$ calculated from (2) was taken in conical flask by the aid of micro syringe. The flask was slowly heated on magnetic stirrer. $NH_3$ liquid gets converted into $NH_3$ gas. As prepared $NH_3$ gas was observed to be wet as tested by pH paper. The wet $NH_3$ was passed through calcium oxide (CaO) to reduce the water vapors and get dry $NH_3$. The detailed procedure of preparing and drying $NH_3$ is given in supplementary information.

## 3. Results and Discussion

### 3.1. XRD Study

The XRD pattern of α- $Fe_2O_3$ thick film is shown in Fig. 2.

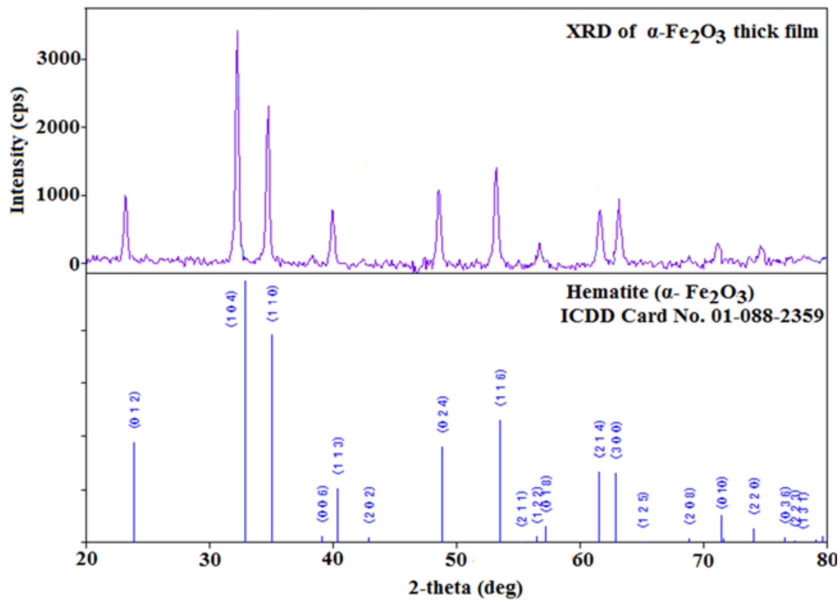



**Fig. 2.** XRD pattern of α- $Fe_2O_3$ thick film.

The strong diffraction peaks were observed to be hematite (α- $Fe_2O_3$) phase of iron oxide (ICDD Card No. 01-088-2359). From Fig. 2 it can be seen that there is significant difference in the peak position between the reference data and presented XRD pattern. This peak shifting occurs due to specimen displacement error. The specimen displacement error is due to the misalignment of the sample [19].

The average crystalline size of the prepared α-$Fe_2O_3$ powder was calculated by Scherrer equation (3).

$$D = \frac{0.9\lambda}{\beta cos\theta} \quad (3)$$

The average crystalline size of α-$Fe_2O_3$ nanocrystalline powder from Scherrer equation is found to be 30.87 nm.

### 3.2. FESEM Study

Fig. 3 shows field emission scanning electron microscopic (FESEM) images of α- $Fe_2O_3$ thick film. Fig. 3 indicates wormlike morphology of α- $Fe_2O_3$ thick film [20].

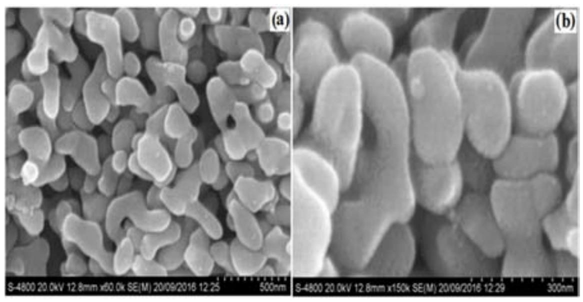



**Fig. 3.** FESEM images of α- thick film.

### 3.3. TGA/DSC Study

Fig. 4 shows the TG and DSC curves of xerogel. The weight loss of 4.94 % below 120 °C resulted from the evaporation of the adsorbed water. The weight loss of 44.02 % from 120°C-350 °C with sharp exothermic peak at 141°C can be ascribed to the decomposition of the residual organic molecules. The small exothermic peak at 355 °C in DSC curve correspondence to the transformation of γ-$Fe_2O_3$ to α-$Fe_2O_3$ phase [21, 22]. Within the temperature range of 355 °C- 800 °C the weight loss is only 1.3 % which indicates that the α-$Fe_2O_3$ is thermodynamically stable above 355 °C.

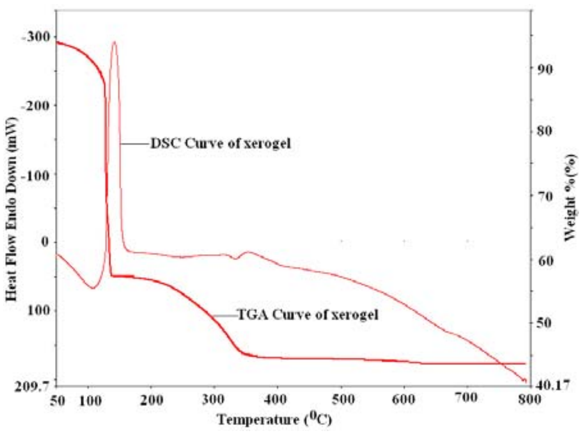



**Fig. 4.** TGA/DSC curve of xerogel.

### 3.4. Gas Sensing Study

Fig. 5 shows the room temperature I-V characteristics of the α-$Fe_2O_3$ thick film in the voltage range -25 V to +25 V. The obtained I-V curve shows the ohmic contacts between the copper wire and the thick film.

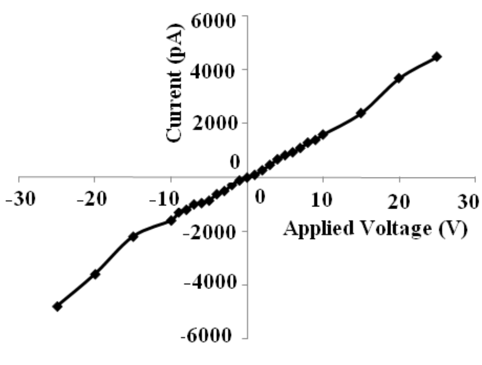



**Fig. 5.** I-V characteristics of the α-$Fe_2O_3$ thick film.

### 3.4.1. Selectivity

Selectivity is an important parameter of gas sensors. Selectivity is defined as the ability of gas sensor to identify a specific gas in presence of other gases [23]. Fig. 6 shows the room temperature response of α- $Fe_2O_3$ to 1000 ppm of different gases such as $NH_3$, methanol, ethanol, isopropanol and acetone.

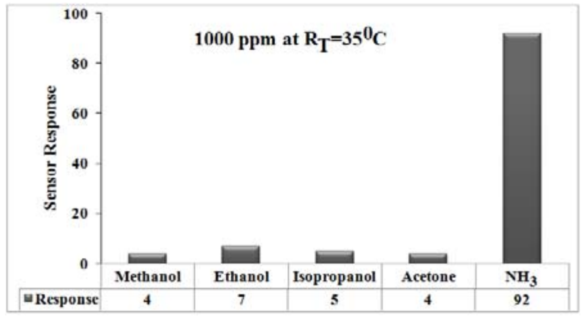



**Fig. 6.** Selectivity of α-$Fe_2O_3$ sensor.

It can be seen from Fig. 6 that α- $Fe_2O_3$ sensor is highly selective to $NH_3$ at room temperature. Selectivity to $NH_3$ can be quantified in terms of selectivity coefficient (K). The selectivity coefficient (K) is defined in equation (4)

$$K = \frac{S_{NH3}}{S_{gas}}, \quad (4)$$

where $S_{NH3}$ is the response of sensor to $NH_3$, $S_{gas}$ is the response of sensor to other gas.

Generally selectivity coefficient should be greater than 5 [13, 24]. Selectivity coefficients were 23 to methanol, 13 to ethanol, 18 to isopropanol and 23 to acetone.

### 3.4.2. Effect of Temperature

Though the aim of this study was to develop room temperature $NH_3$ sensor, we also tested the sensor to 1000 ppm $NH_3$ at different temperatures such as 35, 45, 55 65, 100 and 150 °C to study the effect of temperature on sensitivity of different gases. The obtained responses of sensor are shown in Fig. 7.

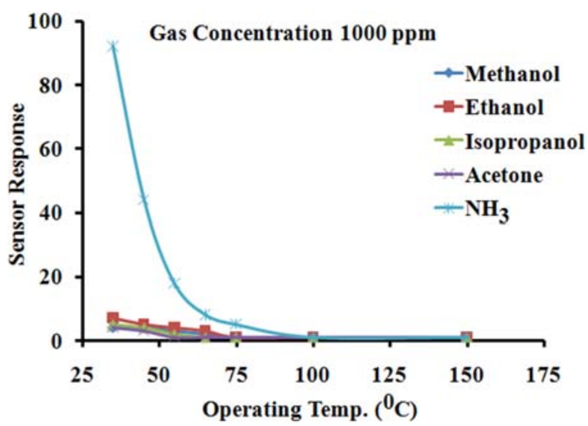



**Fig. 7.** Sensor response as a function of operating temperature.

It is clear from Fig. 7 that the sensor response significantly decreases with increase in operating temperatures.

As the temperature of sensor surface increases, the surface electrolyte layer is evaporated and there is no possibility of ionic conduction resulting in no response towards $NH_3$ at 100 and 150 °C. In this case humidity plays an important role in the sensor response. As the temperature increases, humidity decreases resulting in no adsorption of water molecules on sensor surface and hence no sensor response towards $NH_3$ [25].

### 3.4.3. Effect of Humidity

Relative humidity (RH %) is defined as the ratio of the amount of water vapor in the air at specific temperature to the maximum amount that the air could hold at that temperature. It is expressed in percentage (%). Relative humidity (RH %) is an important parameter in determining the response of metal oxide gas sensors and especially in case of gas sensors operating at room temperature. Hence the effect of RH on response to $NH_3$ has been studied at different humid conditions such as 11 %, 27 %, 37 %, 47 %, 57 %, 67 %, 78 % and 87 % of RH. Fig. 8a shows the effect of humidity on sensor response of different gases.

Fig. 8b shows the variation in sensor response with relative humidity when exposed to 1000 ppm of $NH_3$.

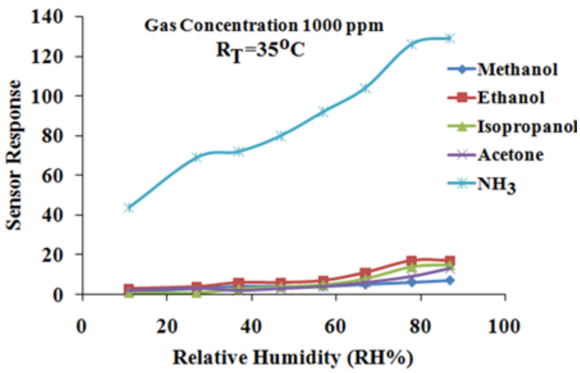



**Fig. 8a.** Effect of humidity on sensor response of different gases.

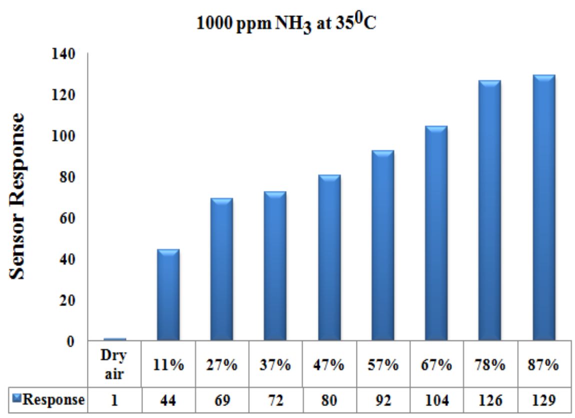



**Fig. 8b.** Effect of humidity on NH3 sensor response.

Under the humid conditions the response of sensor towards 1000 ppm of $NH_3$ was found to be increased. This might be due to the increase in number of adsorbed water molecules on sensor surface and ultimately change in sensing properties of sensor surface [26, 27]. When the $NH_3$ sensing characteristics were carried out in dry air, no response to $NH_3$ was detected due to the absence of thin layer of adsorbed water on sensor surface. If there will be no humidity there will be no response hence humidity plays a significant role in sensing at room temperature.

### 3.4.4. Effect of Gas Concentration

The α-$Fe_2O_3$ sensor response as a function of gas concentrations of different test gases is as shown in Fig. 9a and Fig. 9b.

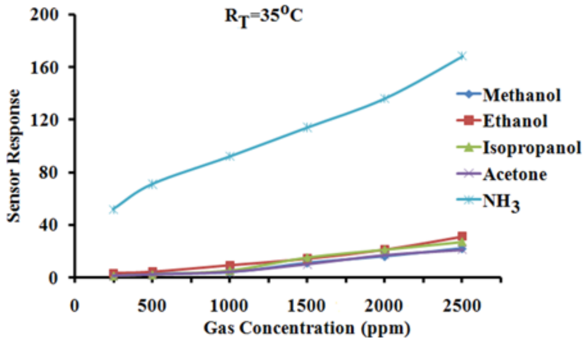



**Fig. 9a.** Sensor response to different concentrations of gases.

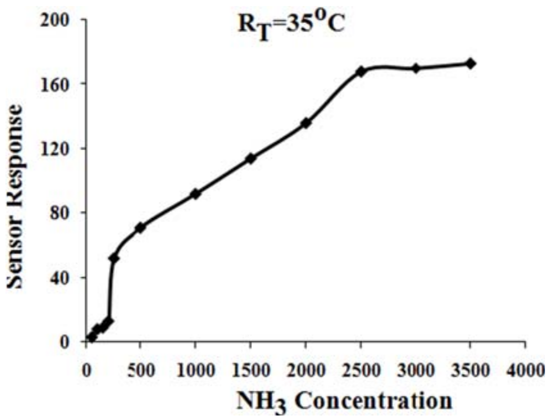



**Fig. 9b.** Sensor response to different concentrations of $NH_3$.

It can be seen from Fig. 9a that α-$Fe_2O_3$ sensor response exhibits a nearly linear increase with increasing concentration of different gases. The α-$Fe_2O_3$ sensor response increased linearly with $NH_3$ concentration from 250 ppm to 2500 ppm and saturates beyond it as shown in Fig. 9b. Even after the further increase of $NH_3$ concentration there was no significant increase in response. The sensor response saturated above 2500 ppm because of formation of multilayer of $NH_3$ over the sensor surface which restricted the further interaction of $NH_3$ molecules with adsorbed water molecules. Hence the sensor response reached at saturation above 2500 ppm. At lower concentration 250 ppm, sensor response was found to be 52. Further decrease in $NH_3$ concentration (<250 ppm) does not produce considerable response. This might be due to less interaction of $NH_3$ molecules with adsorbed water molecules. Hence, the detection limit for $NH_3$ at room temperature was determined to be in the range of 250-2500 ppm [28].

### 3.4.6. Response and Recovery Time

The response time ($T_{Res}$) is defined as the time taken by the sensor to reach maximum steady state current change on the exposure of target gas with the surface of the sensor. The recovery time ($T_{Rec}$) is defined as the time required for the sensor current to reach the base line after the removal of the target gas [29].

Fig. 10 shows the steady state response properties of α- $Fe_2O_3$ gas sensor operated at room temperature under 1000 ppm of $NH_3$. The steady state response properties were measured by introducing constant flow of 1000 ppm of $NH_3$ for 25 sec in glass chamber. The baseline currents before exposure of $NH_3$ were measured. The current through sensor increases on exposure of $NH_3$ and attains to maximum current value in 4 sec and remains in steady state till the exposure of $NH_3$ is on. The current through sensor falls to the baseline current in 9 sec once the exposure of $NH_3$ was off.

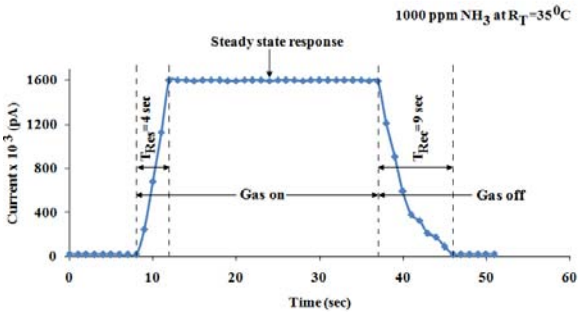



**Fig. 10.** Steady state response properties of α- $Fe_2O_3$ sensor.

It could be seen from Fig. 10 that the α- $Fe_2O_3$ gas sensor has fast response time of 4 sec and quick recovery time of 9 sec. Adsorption and desorption of $NH_3$ molecules occur rapidly because of large surface area of α- $Fe_2O_3$ gas sensor and lower molecular weight of $NH_3$ molecule which results in fast response and quick recovery of sensor [1].

The comparison of present α- $Fe_2O_3$ sensor with other reported sensors for $NH_3$ is summarized in Table 1.

**Table 1.** Comparison of sensor responses toward $NH_3$ of various gas sensors.

| Sr. No | Material | Operating Temp. (°C) | Gas concentration (ppm) | Sensor Response | $T_{Res}$ (S) | $T_{Rec}$ (S) | Reference |
|---|---|---|---|---|---|---|---|
| 1 | α-$Fe_2O_3$ | RT | 1000 | 92 | 4 | 9 | Present work |
| 2 | Sr-SnO2 | RT | 2000 | 54.23% | 6 | - | 28 |
| 3 | PANI/$SnO_2$ | RT | 100 | 29 | - | - | 29 |
| 4 | QDS/$TiO_2$ | RT | 100 | 17.49 | - | - | 30 |
| 5 | $SnO_2$-$SnS_2$ | RT | 500 | 2.2 | 11 | - | 31 |
| 6 | ZnO | RT | 400 | 1.15 | 240 | 290 | 32 |

Table 1 shows that the response towards $NH_3$ of present α- $Fe_2O_3$ sensor is higher with fast response and quick recovery time than other reported sensors.

### 3.4.7. Gas Sensing Mechanism

Metal oxide semiconductor exhibits two kinds of gas sensing mechanisms (i) dissociative gas sensing mechanism which takes place at low temperature ($T_{opt}$ < 100 °C), (ii) combustive gas sensing mechanism which takes place at high temperatures ($T_{opt}$ > 100 °C).

Here we discuss the dissociative gas sensing mechanism at room temperature. At room temperature, metal oxide semiconductor surface is covered with thin layer of adsorbed humidity. In such a condition direct gas sensing mechanism is not possible. At room temperature, gas response in humid atmosphere is controlled by the pH of thin layer of adsorbed water. This electrolytic layer with metal oxide semiconductor surface acts as pH sensor. The pH of the surface electrolyte changes with the interaction of analyte gas with the surface electrolyte. This pH change in surface electrolyte changes the ionic and electronic conductivities of metal oxide semiconductor surface resulting in detectable electrical output signal.

#### 3.4.7.1. Before Exposure of $NH_3$

Water undergoes dissociative adsorption on metal oxide surface as:

$$H_2O \leftrightarrow H^+ + OH^-$$

Chemisorption of water molecules is possible at low humidity level while physisorption is observed at higher humidity levels. At low humidity level electronic conduction is dominant. Fig. 11 and equation (5) correspond to case of water molecule interacting with metal oxide, two Fe-OH dipoles are formed from dissociated hydroxyl ions and by combination of hydrogen ion and lattice oxygen. Two free electrons are produced from each oxygen vacancy ($V_o^{2+}$) in the lattice during formation of Fe-OH bond [26, 35-37].

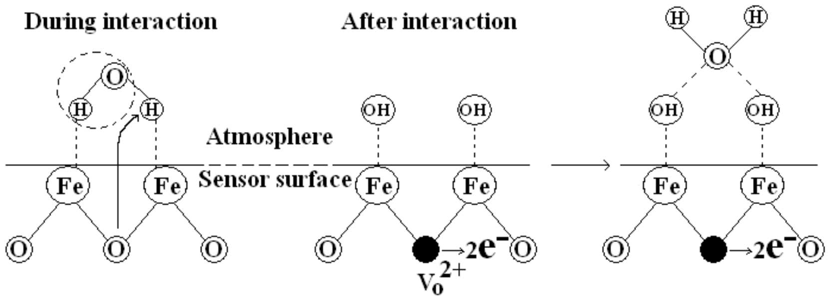



**Fig. 11.** Mechanism of humidity adsorption on α-$Fe_2O_3$ sensor surface.

$$H_2O + 2(Fe_{Fe} + O_o) \leftrightarrow 2\left(Fe_{Fe}^{\delta+} - OH^{\delta-}\right) + V_O^{2+} + 2e^- \quad (5)$$

### 3.4.7.2. After Exposure of $NH_3$

At room temperature layer of water molecules adsorbed on the α-$Fe_2O_3$ sensor surface. This adsorbed water layer acts as liquid electrolyte. $NH_3$ is basic molecule and it has lone pair of electrons. $NH_3$ can easily dissociate electrolytically in water.

When $NH_3$ is exposed on α-$Fe_2O_3$ sensor surface, it penetrates into the liquid electrolyte followed by the electrolytic dissociation of $NH_3$ molecule in $NH_4^+$ and $OH^-$ ions according to the equation (6). The dissociation of $NH_3$ molecules changes the ionic concentration inside the electrolytic layer by increasing the ionic conductivity and change in pH value.

$$NH_{3(gas)} + H_2O_{(aq.)} \leftrightarrow NH_{4(aq.)}^{+} + OH_{(aq.)}^{-} \quad (6)$$

After every such reaction, one electron releases and conductivity of the α-$Fe_2O_3$ sensor increases on exposure of $NH_3$ gas. Interaction of $NH_3$ with α-$Fe_2O_3$ sensor surface through liquid electrolyte is as shown in Fig. 12. At room temperature $NH_3$ sensing may be due to the change in ionic and electronic conductivities. In the present study electronic conduction is dominant over ionic conduction because the humidity level is low in the surrounding [38-41].

### 3.4.8. Stability

Stability is important parameter for room temperature gas sensors. Stability is defined as the ability of the sensor to give reproducible results for certain period of time. The stability and reproducibility of α- $Fe_2O_3$ sensor were studied for 1000 ppm of $NH_3$ at room temperature over the period of 30 days. Fig. 13 showed obtained response and indicates that **α**- $Fe_2O_3$ sensor is stable for 1000 ppm of $NH_3$ over a period of 30 days. The sensing studies were repeated to the α-$Fe_2O_3$ sensor stored over the period of 4 months.

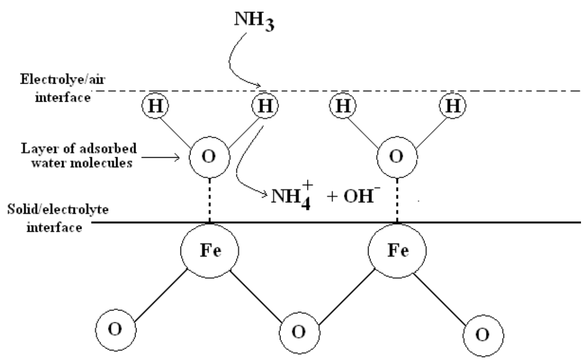



**Fig. 12.** Interaction of $NH_3$ with α-$Fe_2O_3$ sensor surface.

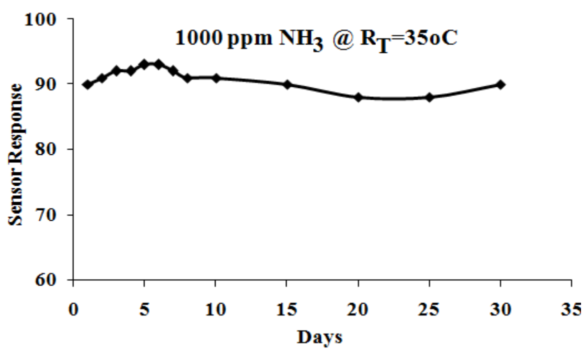



**Fig. 13.** Stability of α- $Fe_2O_3$ sensor towards 1000 ppm of $NH_3$.

## 4. Conclusions

In the summery, we have synthesized nanocrystalline α- $Fe_2O_3$ simple sol gel method. The structural, morphological and thermal properties of α-$Fe_2O_3$ powder were studied by different analytical techniques. Sensing characteristics of α- $Fe_2O_3$ sensor were examined towards the various gases at room temperature under different humid conditions. α-$Fe_2O_3$ sensor was found to be highly selective and showed excellent response towards $NH_3$ at room temperature under humid conditions. In dry air, no response to $NH_3$ was detected. It has been demonstrated that humidity plays a key role in the $NH_3$ sensing at room temperature. The considerable linear change in sensor response was observed over the range of 250-2500 ppm. The lower $NH_3$ detection limit was found to be 250 ppm. The α-$Fe_2O_3$ sensor exhibited excellent sensitivity (S=92) towards 1000 ppm of $NH_3$ with quick response of 4 sec and fast recovery of 9 sec. The α-$Fe_2O_3$ sensor showed good stability and reproducibility.

## Acknowledgements

Authors are heartily thankful to Head, INSPIRE Program, Department of Science & Technology, New Delhi. The authors also would like to thank Head of Physics Department and Principal Pratap College Amalner, Dist: Jalgaon, Maharashtra for providing laboratory facilities.

## References

[1]. C. Balamurugam, D. W. Lee, A selective $NH_3$ gas sensor based on mesoporous p-type $NiV_2O_6$ semiconducting nanorods synthesized using solution method, *Sensors Actuators B: Chemical,* Vol. 192, 2014, pp. 414-422.

[2]. V. Balouria, A. Kumar, S. Samanta, A. Singh, A. K. Debnath, A. Mahajan, R. K. Bedi, D. K. Aswal, S. K. Gupta. Nano-crystalline $Fe_2O_3$ thin films for ppm level detection of $H_2S$, *Sensors Actuators B: Chemical,* Vol. 181, 2013, pp. 471-478.

[3]. P. Dhivya, A. K. Prasad, M. Shridharan. Nanostructured $TiO_2$ films: Enhanced $NH_3$ detection at room temperature, *Ceramics International*, Vol. 40, Issue 1, 2014, pp. 409-415.

[4]. R. Pandeeswari, B. G. Jeyaprakash, High sensing response of β-$Ga_2O_3$ thin film towards ammonia vapours: Influencing factors at room temperature, *Sensors Actuators B: Chemical,* Vol. 195, 2014, pp. 206-214.

[5]. S. K. Dhawan, D. Kumar, M. K. Ram, S. Chandra, D. C. Trivedi, Application of conducting polyaniline as sensor material for ammonia, *Sensors Actuators B: Chemical*, Vol. 40, 1997, pp. 99-103.

[6]. C. Malins, A. Doyle, B. D. MacCraith, F. Kvasnik, M. Landl, P. Simon, L. Kalvoda, R. Lukas, K. Pufler, I. Babusik. Personal ammonia sensor for industrial environments, *Journal of Environment and Monitoring,*, Vol. 4, Issue 5, 1999, pp. 417-422.

[7]. B. Timmer, W. Olthuis, A. V. D. Berg, Ammonia sensors and their applications – a review, *Sensors Actuators B: Chemical,* Vol. 107, 2005, pp. 666-677.

[8]. P. Sun, W. Eang, Y. Liu, Y. Sun, J. Ma, G. Lu, Hydrothermal synthesis of 3D urchin-like α-$Fe_2O_3$ nanostructure for gas sensor, *Sensors Actuators B: Chemical,* Vol. 173, 2012, pp. 52-57.

[9]. P. Zhang, Z. G. Ping, H. K. Liu. Submicron-sized cube-like α-$Fe_2O_3$ agglomerates as an anode material for Li-ion batteries, *Electrochemical Acta*, Vol. 55, Issue 28, 2010, pp. 8521-8526.

[10]. L. Li, Y. Chu, Y. Liu, L. Dong, Template-Free Synthesis and Photocatalytic Properties of Novel $Fe_2O_3$ Hollow Spheres, *Journal of Physics and Chemistry C*, Vol. 111, Issue 5, 2007, pp. 2123-2127.

[11]. F. Bodily, A. M. Ferrari, C. Leonel, T. Manfredini, Hydrothermal Preparation and Characterization of Nanocrystalline Powder of β-Indium Sulfide, *Materials Research Bulletin*, Vol. 33, Issue 5, 1998, pp. 723-729.

[12]. J. Ma, J. Lian, X. Duan, X. Liu, W. Zheng, α-$Fe_2O_3$: Hydrothermal synthesis, magnetic and electrochemical properties, *Journal of Physics and Chemistry C*, Vol. 114, Issue 24, 2010, pp. 10671–10676.

[13]. D. Patil, V. Patil, P. Patil, Highly sensitive and selective LPG sensor based on α-$Fe_2O_3$ nanorods, *Sensors Actuators B: Chemical,* Vol. 152, 2011, pp. 299-306.

[14]. L. Duraes, B. F. O. Costab, J. Vasquesa, J. Camposc, A. Portugala, Phase investigation of as-prepared iron oxide/hydroxide produced by sol–gel synthesis, *Materials Letters,* Vol. 59, Issue 7, 2005, pp. 859-863.

[15]. L. Duraes, A. Moutinho, I. J Seabra, B. F. O. Costa, H. C. de Sousa, A. Portugal, Characterization of iron (III) oxide/hydroxide nanostructured materials produced by sol–gel technology based on the Fe $(NO_3)_3 \cdot 9H_2O$ – $C_2H_5OH$ – $CH_3CHCH_2O$ system, *Materials Chemistry and Physics,* Vol. 130, Issue 1-2, 2011, pp. 548-560.

[16]. N. V. Hieu, N. Q. Dung, P. D. Tam, T. Trung, N. D. Chien, Thin film polypyrrole/SWCNTs nanocomposites-based $NH_3$ sensor operated at room temperature, *Sensors Actuators B: Chemical*, Vol. 140, 2009, pp. 500-507.

[17]. V. V. Deo, D. M. Patil, L. A. Patil, M. P. Kaushik, Ultrasonically sprayed nanostructured $CdSnO_3$ thin films for the detection of VOC's, *Sensors Actuators B: Chemical*, Vol. 196, 2014, pp. 489-496.

[18]. K. Arshak, I. Gaidan, NiO / $Fe_2O_3$ polymer thick films as room temperature gas sensors, *Thin Solid Films*, Vol. 495, Issue 1-2, 2006, pp. 286-291.

[19]. B. D. Cullity, Elements of X-ray Diffraction, 2$^{nd}$ edition, *Addison-Wesley,* Reading, MA, 1978.

[20]. L. Wan, K. Shi, X. Tian, H. Fu, Facile synthesis of iron oxide with wormlike morphology and their application in water treatment, *Solid State Chemistry*, Vol. 181, Issue 4, 2008, pp. 735-740.

[21]. J. S. Xu, Y. J. Zhu, Monodisperse $Fe_3O_4$ and γ-$Fe_2O_3$ magnetic mesoporous microspheres as anode materials for lithium-ion batteries, *ACS Applied Materials Interfaces*, Vol. 4, Issue 9, 2012, pp. 4752−4757.

[22]. E, Darezereshki, One-step synthesis of hematite (α-$Fe_2O_3$) nano-particles by direct thermal-decomposition of maghemite, *Materials Letters*, Vol. 65, Issue 4, 2011, pp. 642–645.

[23]. L. A. Patil, A. R. Bari, M. D. Shinde, Vinita Deo, M. P. Kaushik, Detection of dimethyl methyl phosphonate -a simulant of sarin: The highly toxic chemical warfare using platinum activated nanocrystalline ZnO thick films, *Sensors Actuators B: Chemical*, Vol. 161, 2012, pp. 372– 380.

[24]. X. Zhao, Z. Li, X. Lou, M. Li, N. Zhang. Room-temperature chlorine gas sensor based on $CdSnO_3$ synthesized by hydrothermal process, *Journal of Advanced Ceramics*, Vol. 2, Issue 1, 2013, pp. 31-36.

[25]. A. Helwig, G. Müller, G. Sberveglieri, M. Eickhoff, On the low-temperature response of semiconductor gas sensors, *Journal of Sensors*, 2009, pp. 1-17, Article ID 620720.

[26]. P. Shankar, J. B. B Rayappan, Gas sensing mechanism of metal oxides: The role of ambient atmosphere, type of semiconductor and gases - A review, *Science Letters Journal*, Vol. 126, Issue 4, 2015, pp. 1-16.

[27]. G. K. Mani, J. B. B. Rayappan, A highly selective room temperature ammonia sensor using spray deposited zinc oxide thin film, Sensors *Actuators B: Chemical,* Vol. 183, 2013, pp. 459-456.

[28]. S. T. Navale, D. K. Bandgar, S. R. Nalage, G. D. Khuspe, M. A. Chougule, Y. D. Kolekar, S. Sen, V. B. Patil. Synthesis of $Fe_2O_3$ nanoparticles for nitrogen dioxide gas sensing applications, *Ceramics International*, Vol. 39, Issue 6, 2013, pp. 6453-6460.

[29]. L. A. Patil, M. D. Shinde, A. R. Bari, V. V. Deo, Highly sensitive ethanol sensors based on nanocrystalline $SnO_2$ thin films, *Current Applied Physics*, Vol. 10, Issue 5, 2010, pp. 1249-1254.

[30]. S. Xu, K. Kan, Y. Yang, C. Jiang, J. Gao, L. Jing, P. Shen, L. Li, K. Li, Enhanced $NH_3$ gas sensing performance based on electrospun alkaline-earth metals composited $SnO_2$ nanofibers, *Journal of Alloys and Compounds*, Vol. 618, 2015, pp. 240-247.

[31]. S. Bai, Y. Tian, M. Cui, J. Sun, Y. Tian, R. Luo, A. Chen, D. Li, Polyaniline@$SnO_2$ heterojunction loading on flexible PET thin film for detection of $NH_3$ at room temperature, *Sensors Actuators B: Chemical*, Vol. 226, 2016, pp. 540–547.

[32]. Y. Liu, L. Wang, H. Wang, M. Xiong, T. Yang, G. S. Zakharova, Highly sensitive and selective ammonia gas sensors based on PbS quantum dots/$TiO_2$ nanotube arrays at room temperature, *Sensors Actuators B: Chemical*, Vol. 236, 2016, pp. 529–536.

[33]. K. Xu, N. Li, D. Zeng, S. Q. Tian, S. S. Zhang, D. Hu, C. Xie, Interface bonds determined gas-sensing of $sno_2$–$sns_2$ hybrids to ammonia at room temperature, *ACS Applied Materials Interfaces*, Vol. 7, Issue 21, 2015, pp. 11359–11368.

[34]. A. Kshirsagar, A. B. Joshi, A. Joshi, D. K. Avasthi, T. M. Bhave, S. A. Gangal, Comparative study of irradiated and annealed ZnO thin films for room temperature ammonia gas sensing, IEEE *Sensors Conference 2007*, Atlanta, GA, USA, 28-31 October 2007, pp. 162-165.

[35]. Z. Chen, C. Lu, Humidity Sensors: A Review of Materials and Mechanisms, *Sensors Letters*, Vol. 3, Issue 4, 2005, pp. 274-295.

[36]. Y. C. Yeh, T. Y. Tseng, D. A. Chang, Electrical properties of $TiO_2$-$K_2Ti_6O_{13}$ porous ceramic humidity sensor, *Journal of American Ceramic Society*, Vol. 73, Issue 7, 1990, pp. 1992-1998.

[37]. Y. Shimizu, M. Shimabukuro, H. Arai, T. Seiyama, Humidity-sensitive characteristics of $La^{3+}$-doped and undoped SrSnO3, *Journal of Electrochemical Society*, Vol. 136, Issue 4, 1989, pp. 1206-1210.

[38]. A. Helwig, G. Müller, O. Weidemann, A. Härtl, J. A. Garrido, M. Eickhoff, Gas sensing interactions at hydrogenated diamond surfaces, *IEEE Sensors Journal*, Vol. 7, Issue 9, 2007, pp. 1349–1353.

[39]. A. Helwig, G. Müller, G. Sberveglieri, G. Faglia, Gas sensing properties of hydrogenated amorphous silicon films, *IEEE Sensors Journal*, Vol. 7, Issue 11, 2007, pp. 1506–1512.

[40]. A. Helwig, G. Muller, M. Eickhoff, G. Sberveglieri, Dissociative gas sensing at metal oxide surfaces, *IEEE Sensors Journal*, Vol. 7, Issue 12, 2007, pp. 1675–1679.

[41]. S. Beer, A. Helwig, G. Muller, J. Garrido, M. Stutzmann. Water adsorbate mediated accumulation gas sensing at hydrogenated diamond surfaces, *Sensors Actuators B: Chemical*, Vol. 181, 2013, pp. 894– 903.

___

International Frequency Sensor Association Publishing

www.sensorsportal.com

ISSN 1726- 5479

9 771726 547001