Data Estimation in Sensor Networks Using Physical and Statistical Methodologies

Yingshu Li, Chunyu Ai, Wiwek P. Deshmukh, and Yiwei Wu
Georgia State University

Data Acquisition

- continuous query: SELECT *
- Goal: continuously collect data and prolong network lifetime.
  - In-network aggregation can not be applied.
- Existing solutions
  - Continuous reporting (accurate data)
    - Too much radio transmission.
  - Model-driven acquisition (approximate data)
    - All nodes should be always in working status.

Our Contribution

- Energy-efficient Working Scheme
  - All nodes serve as active working nodes by turns.
  - Conserve energy.
- Estimation models
  - DEPM (Data Estimation using Physical Model)

    take advantages of physical characteristics of light intensity
  - DESM (Data Estimation using Statistical Model)

    temporal and spatial correlations

Outline

- Problem Definition
- Estimation Models
  - DEPM
  - DESM
- Simulation
- Conclusion

Problem Definition

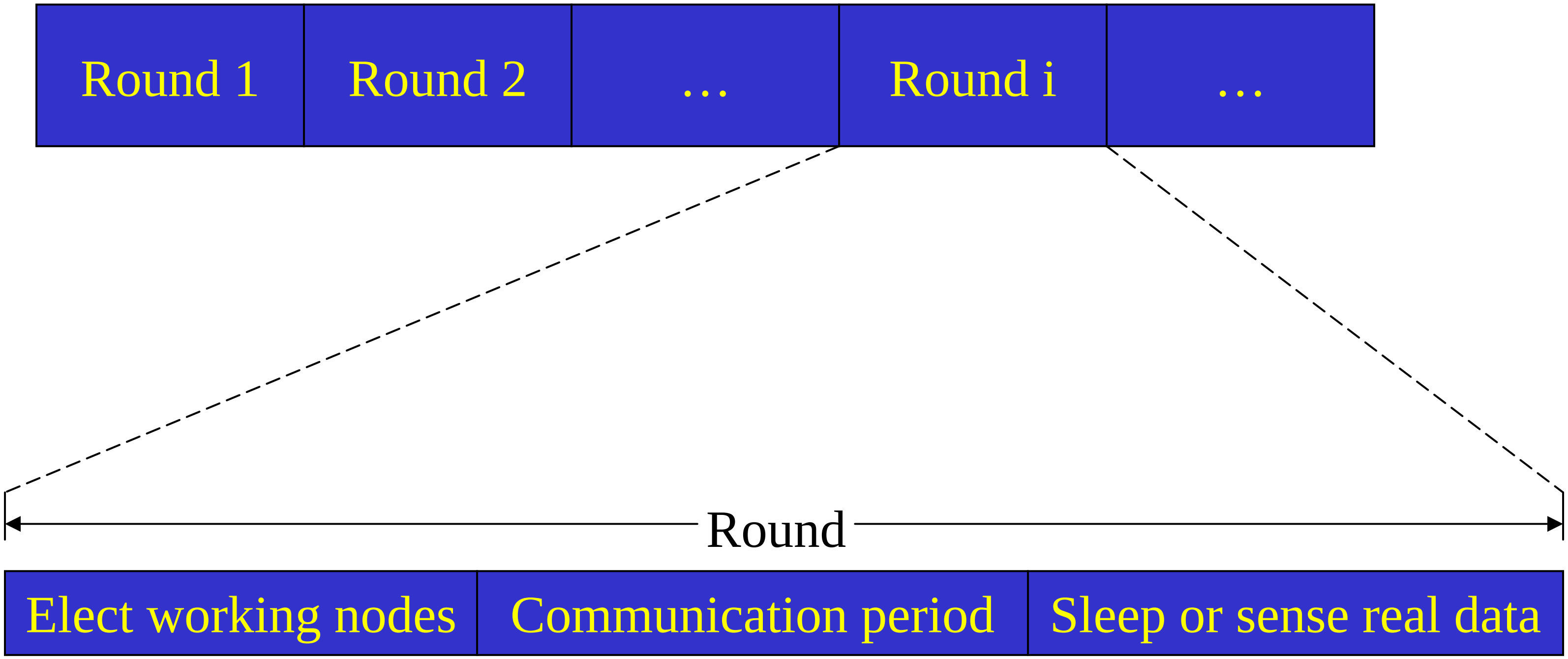
Sensor Energy-efficient Approximate Query Answer

(SEAQA) Problem:

Given a three dimensional area A and a set of N sensors $S = \{s_1, s_2, \ldots, s_N\}$, derive a working scheme ws for S such that:

1. For each sensor s_i, its returned value V_e deviates its real sensing value V as little as possible, that is, $| V_e - V |$ is minimized.
2. The energy consumptions among all the sensors are balanced.
3. The network lifetime is maximized.

Working Scenario



Estimation Models

- Data Estimation using Physical Model (DEPM)
  Physical characteristics of light intensity.

- Data Estimation using Statistical Model (DESM)
  Usually, the collected data has strong temporal and spatial locality property.

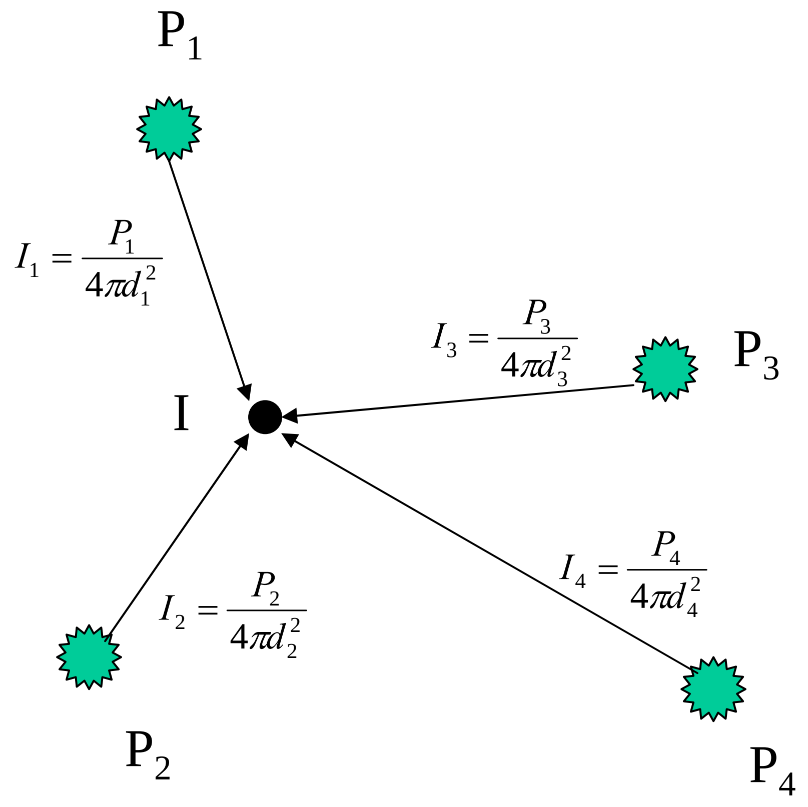
DEPM

Inverse square law:

$$I = \frac{P}{4\pi d^2}$$

principle of linear superposition

$$I = \frac{P_1}{4\pi d_1^2} + \frac{P_2}{4\pi d_2^2} + \frac{P_3}{4\pi d_3^2} + \frac{P_4}{4\pi d_4^2}$$



Light source

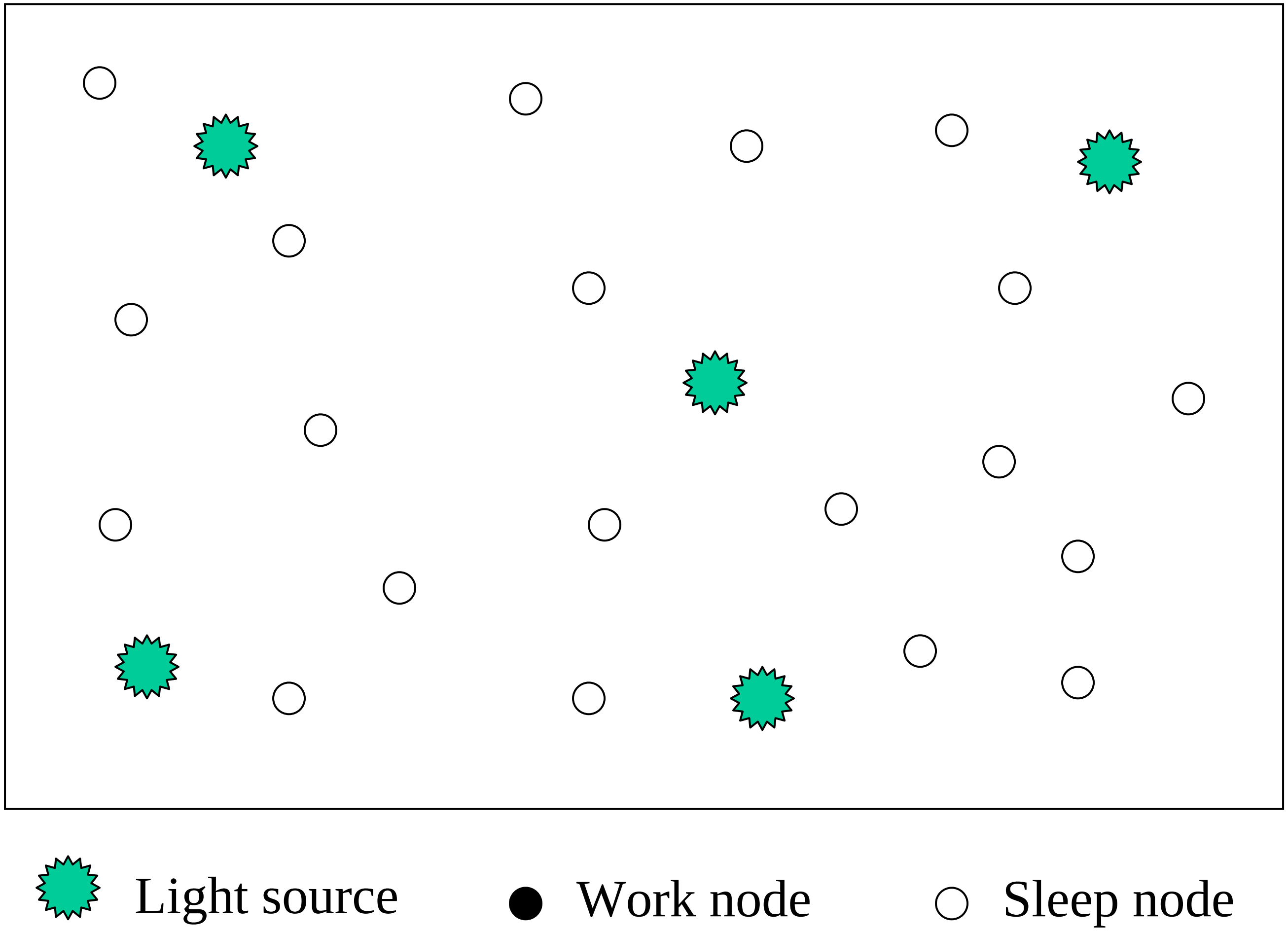
DEPM

$$I_k = \frac{P_j}{4\pi d^2(l_j, s_k)}$$

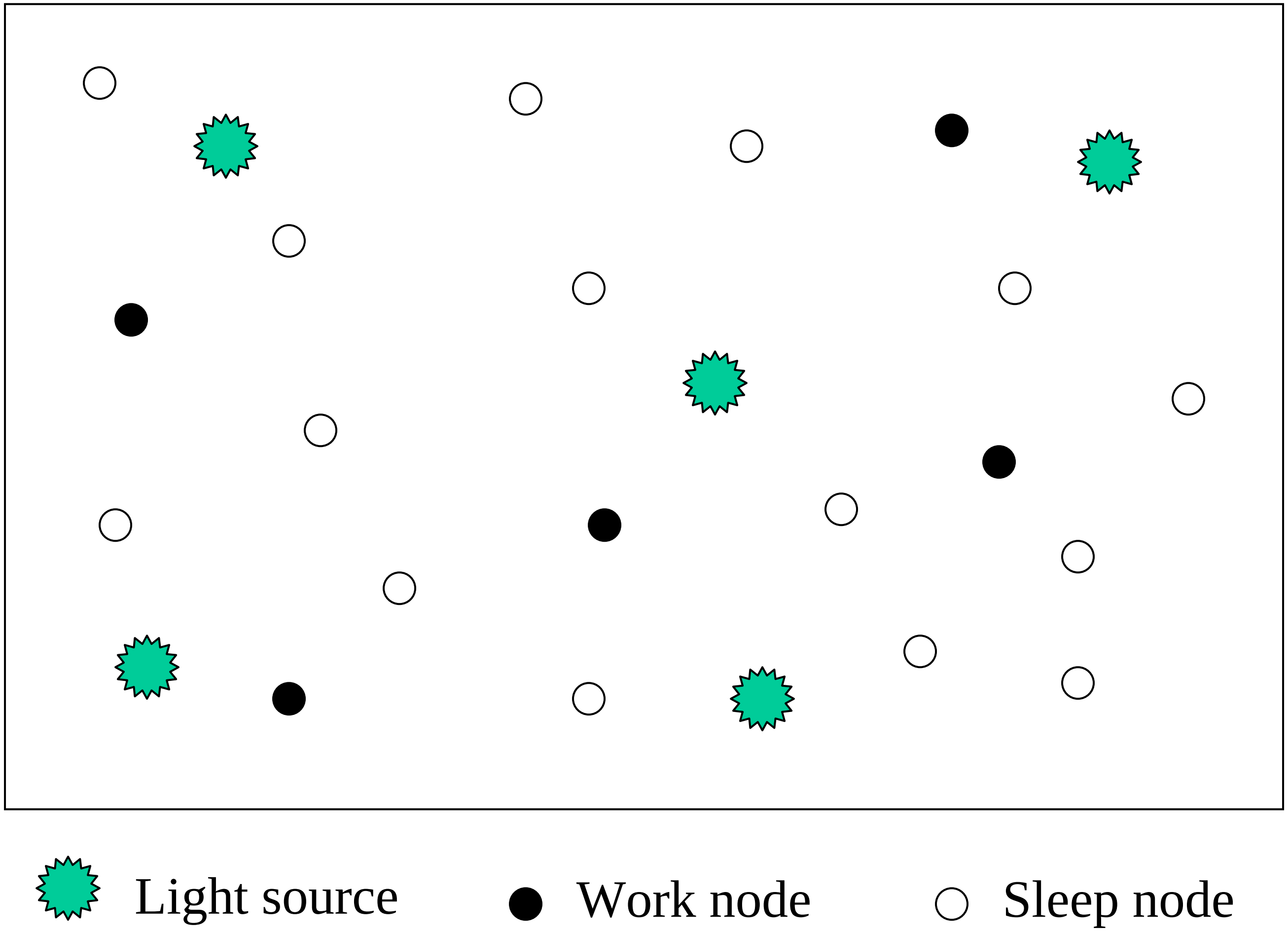
$$I_k = \frac{P_1}{4\pi d^2(l_1, s_k)} + \frac{P_2}{4\pi d^2(l_2, s_k)} + \cdots + \frac{P_M}{4\pi d^2(l_M, s_k)} = \sum_{j=1}^{M} \frac{P_j}{4\pi d^2(l_j . s_k)} = \sum_{j=1}^{M} a_{jk} P_j, a_{jk} = \frac{1}{4\pi d^2(l_j, s_k)}$$

$$\begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,l} & \cdots & a_{1,M} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{l,1} & a_{l,2} & \cdots & a_{l,l} & \cdots & a_{l,M} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{M,1} & a_{M,2} & \cdots & a_{M,l} & \cdots & a_{M,M} \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_l \\ \vdots \\ P_M \end{bmatrix} = \begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_l \\ \vdots \\ I_M \end{bmatrix}$$

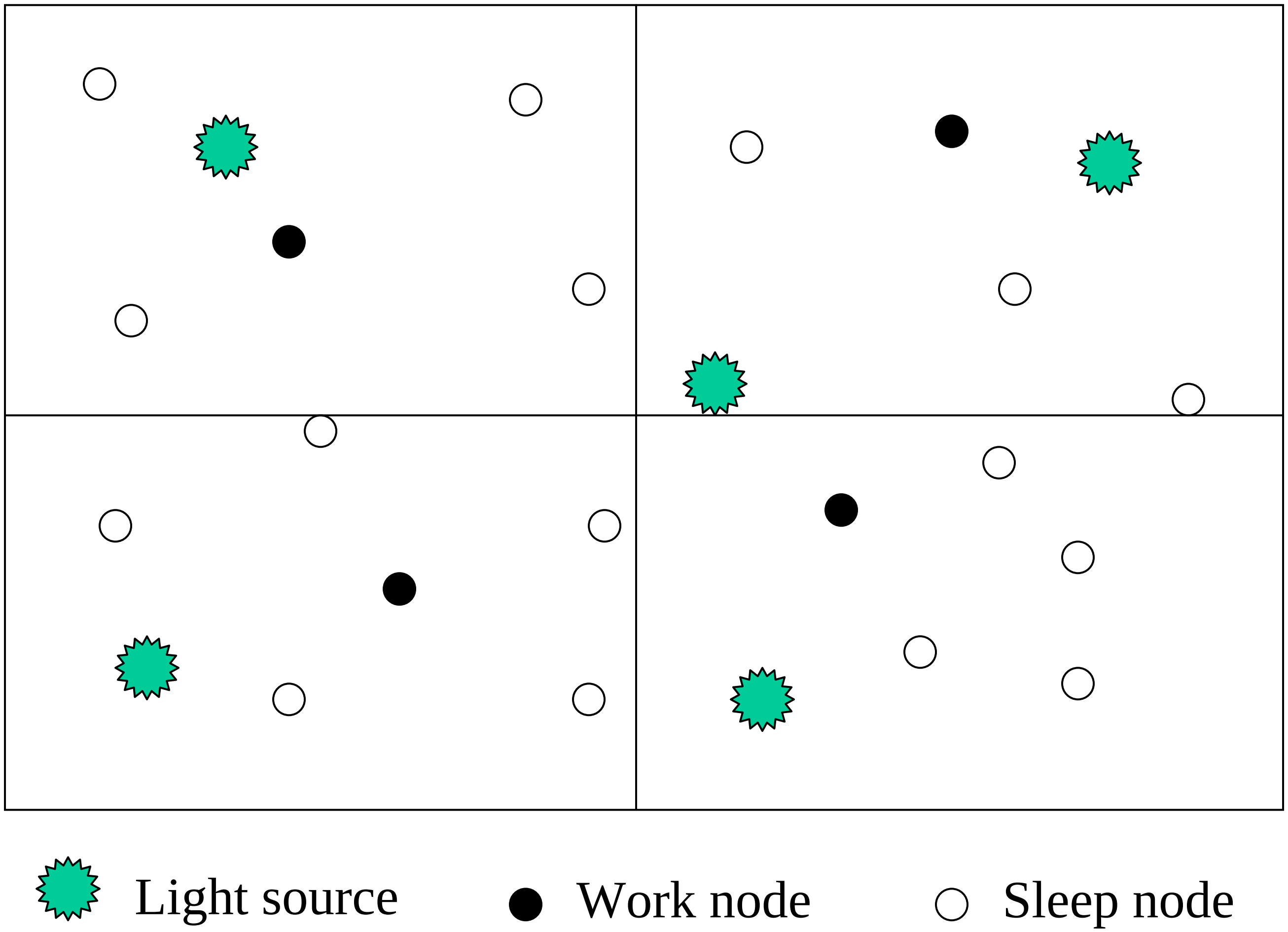
DEPM



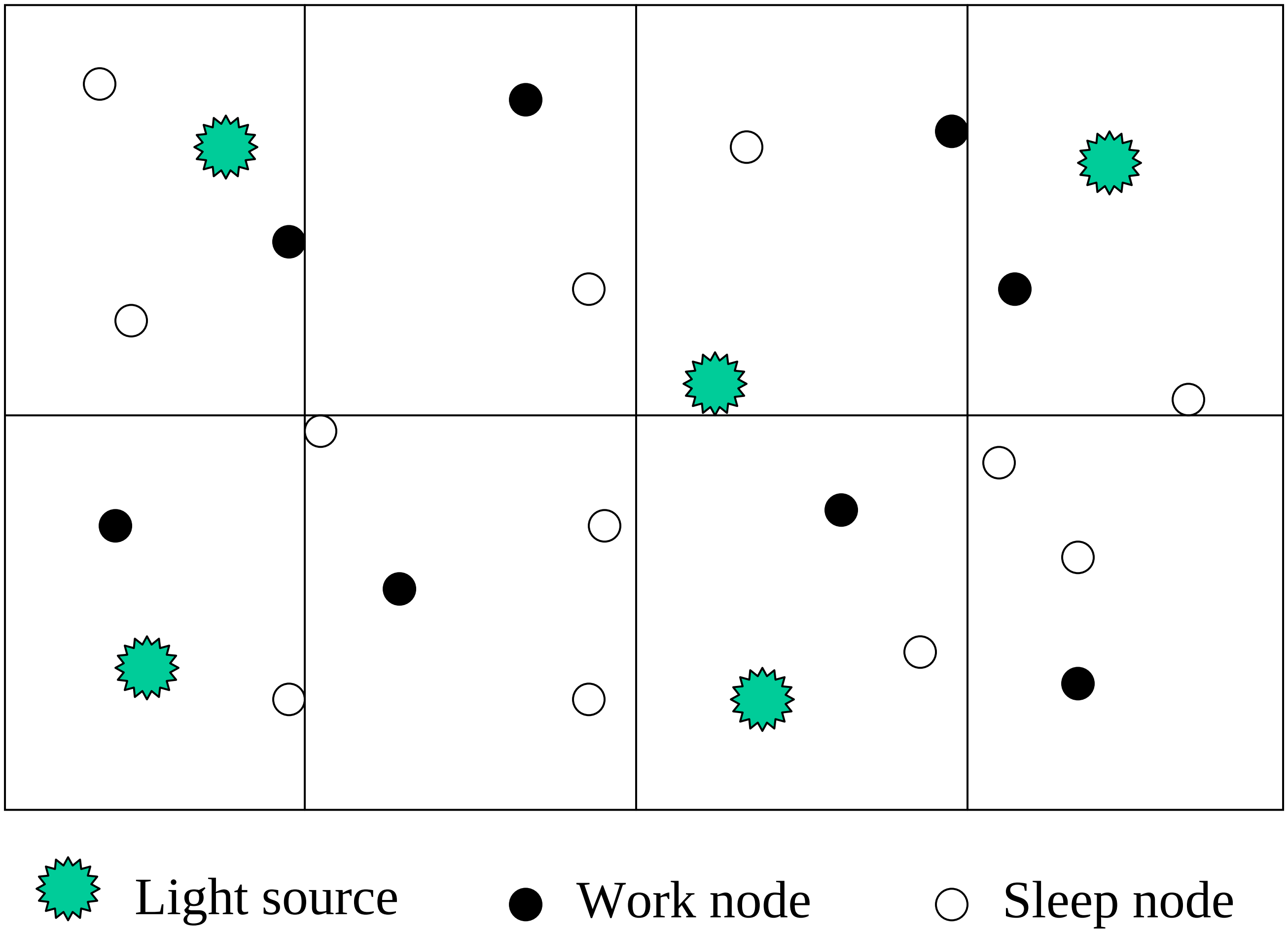
DEPM



DESM



DESM



DESM

Temporal correlation

Spatial correlation

$$\overline{X}_{j(m+1)} = (1-\alpha)\overline{Y} + \alpha\overline{Z}$$

t1:30
t2:31

t1:32
t2:34

t1:31
t2:33

$$\overline{Y} = \overline{X}_{j(m)}$$

$$\overline{Z} = X_{j(m)}\left(1 + \frac{X_{i(m+1)} - X_{i(m)}}{X_{i(m)}}\right)$$

Covariance between X_i and X_j

$$\alpha = \varphi(X_i, X_j) = \frac{Cov(X_i, X_j)}{\sigma_{X_i}\sigma_{X_j}}$$

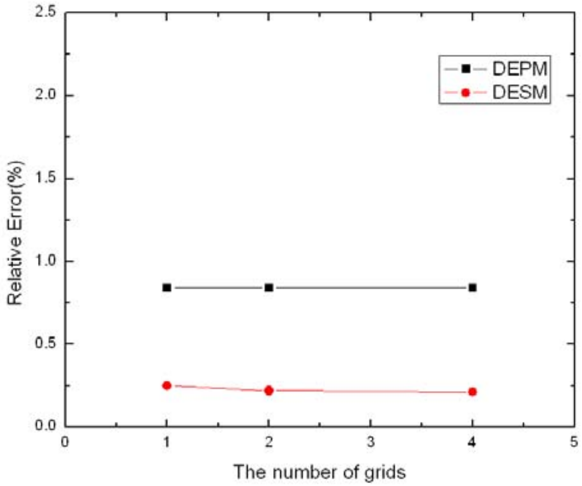
According to a historical data set.

Simulation Setting

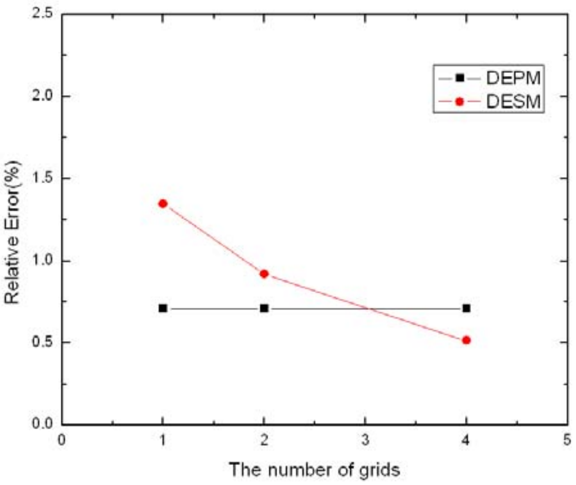
- 20 TelosB sensors
- Employed in a laboratory of size 25’x50’
- Collected 3 datasets
  - Dataset 1: both light sources turned on.
  - Dataset 2: two light sources were turned on and off every 5 minutes.
  - Dataset 3: two light sources turned on and off randomly and frequently.

Simulation Results

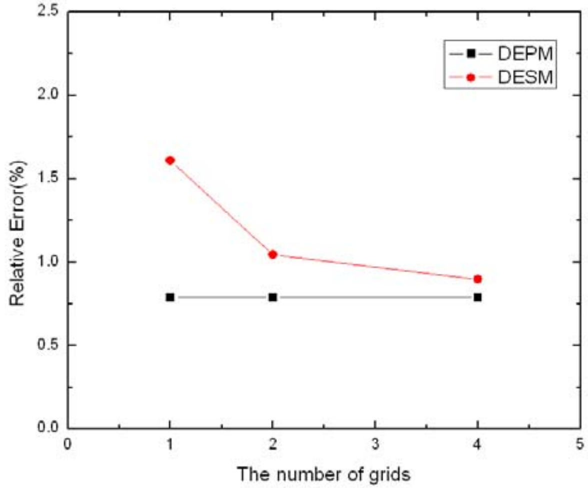
Dataset 1



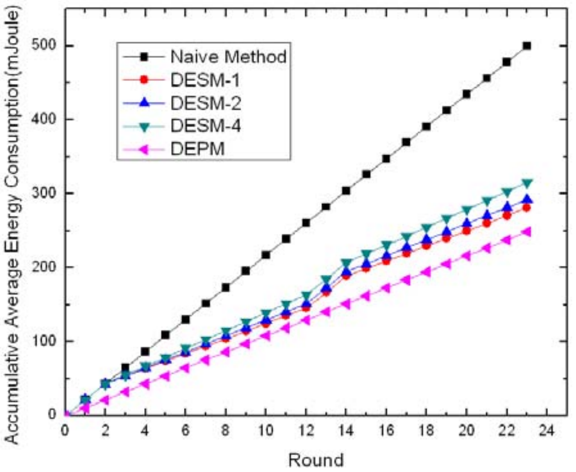
Dataset 2



Dataset 3



Energy Consumption



Conclusion

- A energy-efficient working scheme (sensors work by turns) was proposed to prolong network lifetime.
- DEPM and DESM are energy-efficient and effective.
- Future work: extend DEPM to fit other sensed attributes.