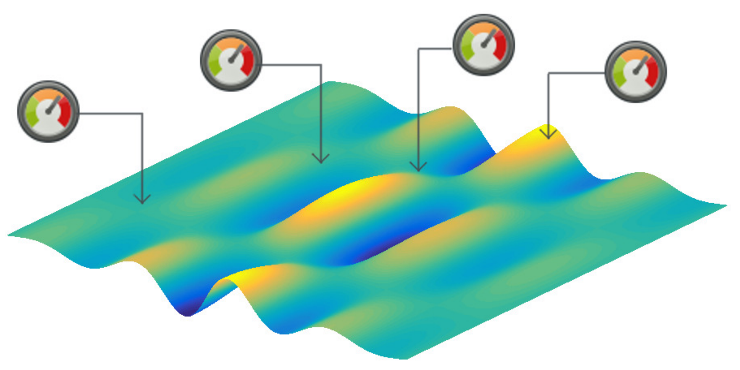
Deep Reinforcement Learning Based Energy Beamforming for Powering Sensor Networks

Ayça Özçelikkale[1], Mehmet Koseoglu[2], Mani Srivastava[3], Anders Ahlén[1]

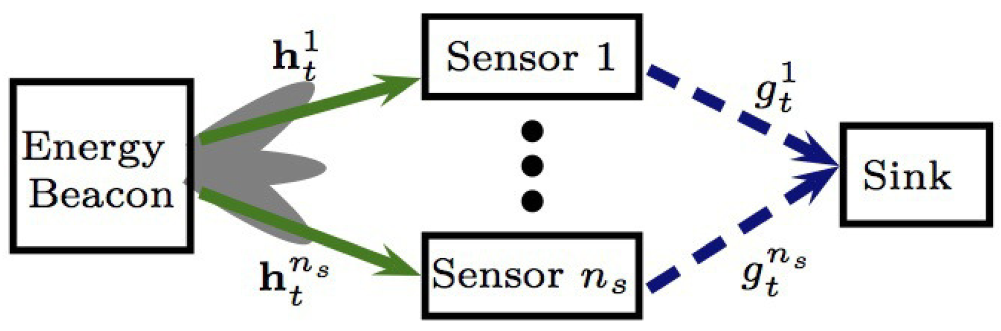
[1]Signals and Systems, Uppsala University, Sweden; [2]Dept. Computer Engineering, Hacettepe University, Turkey; [3]Dept. of Electrical and Computer Engineering, University of California, USA

Sensors measure a spatially correlated unknown field



Application Example: Smart Agriculture

Sensors are powered by a multi-antenna energy beacon and send their measurements to the sink



Main Aim:

Reconstruct the unknown field with as small average distortion as possible under a total power constraint at the energy beacon

Information Transfer

Communications to the Sink

- The signal that is received by the sink from sensor i at time slot t is

$$y_t^i = g_t^i \sqrt{\frac{p_t^i}{\sigma^2_{\bar{s}_t^i}}}\,\bar{s}_t^i + w_t^i,$$

$\bar{s}_t^i$	: samples of the unknown signal at sensor i at time slot t
$\sqrt{p_t^i} \in \mathbb{R}$	: power amplification factor of sensor i at time slot t
$g_t^i \in \mathbb{R}$	: effective channel gain for sensor i at time slot t
$y_t^i \in \mathbb{C}$	: received observations for sensor i at time slot t
$w_t^i \in \mathbb{C}$	: zero-mean proper white noise

Performance Criterion

Mean-square error for reconstruction of the unknown field: $\mathrm{MSE} = \varepsilon_t(\boldsymbol{p}_t) = \mathbb{E}[||\boldsymbol{s}_t - \hat{\boldsymbol{s}}_t||^2]$

$\boldsymbol{s}_t$	: zero-mean complex proper random field denoting measurements from time slot t
$\hat{\boldsymbol{s}}_t$	: the linear minimum mean-square error (LMMSE) estimate
$\boldsymbol{p}_t$	: $[p_t^1;\dots;p_t^{n_s}] \in \mathbb{R}^{n_s \times 1}$, vector of power amplification factors

Wireless Power Transfer

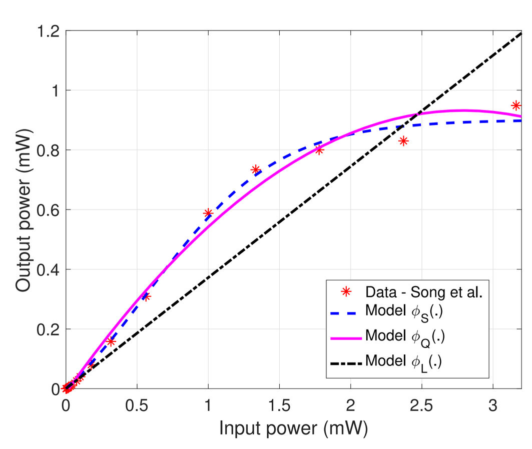
Energy beacon serves n_s sensors using a beamforming strategy $\boldsymbol{K}_{z_t} = \sum_{j=1}^{n_b} \gamma_{t,j} \boldsymbol{e}_j \boldsymbol{e}_j^{\mathrm{H}}$

$$P_{r,t}^i = \mathrm{tr}[h_t^i \boldsymbol{K}_{z_t} (h_t^i)^{\mathrm{H}}]$$
$$E_t^i = \tau_E \phi(P_{r,t}^i),$$

E_t^i	: energy harvested by sensor i during time slot t
$\boldsymbol{K}_{z_t}$	: beamforming strategy at at the energy beacon at time t
h_t^i	: channel for wireless power transfer i at time slot t
$\phi(.)$	: the conversion between power received and power harvested
τ_E	: length of energy harvesting time slot
$\{e_j\}_{j=1}^{n_b}$	: dictionary of beamforming vectors

Practical energy harvesting (EH) efficiency models are considered:

$\phi_L(.)$	: standard linear model
$\phi_Q(.)$	: quadratic model of XuOzcelikkaleMcKelveyViberg˙2017
$\phi_S(.)$	: logistic function model of BoshkovskaNgZlatanovSchober˙2015



Fit of the models to practical EH data

Problem Statement

We jointly design optimal

- **beamforming strategies** K_{z_t} at the energy beacon
- **power amplification factors** p_t^i at the sensors

in order to

- **minimize the MSE** over the time period of $1 \le t \le n_t$

$$\min_{\boldsymbol{p}_t, \boldsymbol{K}_{z_t}} \sum_{t=1}^{n_t} \varepsilon_t(\boldsymbol{p}_t)$$

$$\text{s.t.} \sum_{k=1}^{t} \tau_I p_k^i \le \sum_{k=1}^{t} \tau_E \phi(\mathrm{tr}[h_k^i \boldsymbol{K}_{z_k} (h_k^i)^{\mathrm{H}}]),\ \forall t, \forall i \qquad \text{“energy neutrality constraints @ sensors”}$$

$$\mathrm{tr}[\boldsymbol{K}_{z_t}] \le P_B, \quad \forall t, \qquad \text{“power constraint @ energy beacon”}$$

Two Approaches: Reinforcement Learning vs. Optimization

REINFORCEMENT LEARNING

- ✓ does not rely on prior knowledge
  - no channel state information (CSI)
  - no knowledge on the form of the utility function
- ✓ does not rely on strong assumptions
  - Markovian assumption
  - feedback on the utility function and battery level information from the previous time slot is available
- × does not guarantee convergence
- × takes many iterations to converge
- × ✓ optimize by interacting with the system (or alternatively with a comprehensive simulation environment)

STANDARD OPTIMIZATION

- × typically requires knowledge of system parameters (but robust solutions are also possible)
  - CSI, form of the utility function (error function) and statistics of the unknown field is known
- ✓ may guarantee optimality if the problem is well-behaved (for instance convex)
  - our problem is convex with $\phi_L(.)$ and $\phi_Q(.)$ but not with $\phi_S(.)$
- ✓ may provide convergence guarantees
  - convergence to an optimal solution is guaranteed for $\phi_L(.)$ and $\phi_Q(.)$
- ✓ no online training is required
- × requires a system model

Deep Reinforcement Learning Approach

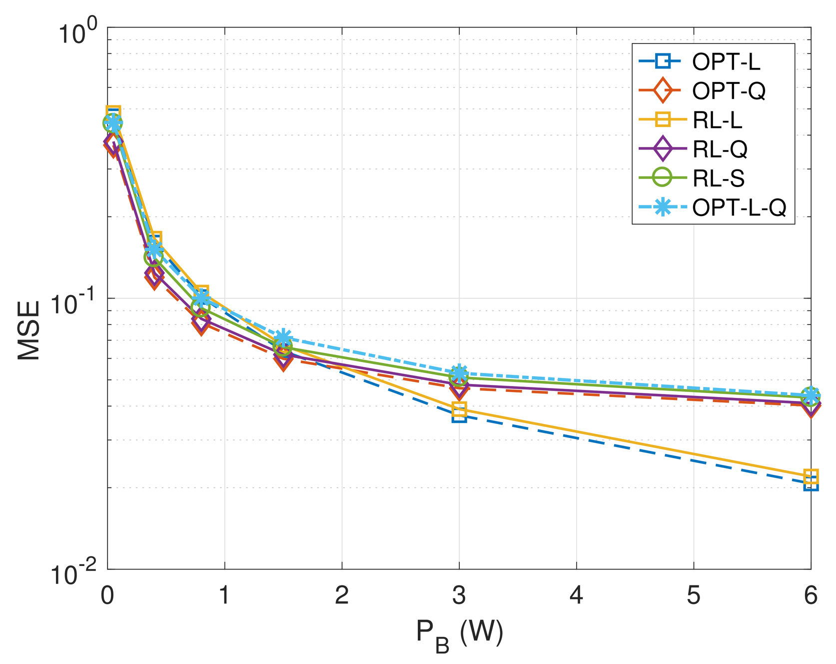
- Method: Proximal Policy Optimization
- Reward: negative of the MSE at each time step
- Decision variables: i) ratio of the energy to be used to the battery level at each sensor; ii) energy allocated to each beamforming dictionary element at each time step at the energy beacon
- The widths of the hidden layers are adapted to the size of the sensor network.

Example: $n_s = 33$:
Value function: NN with 3 hidden layers of size $\{340, 41, 5\}$
Policy: NN with 3 hidden layers of size $\{340, 534, 840\}$

Experiments

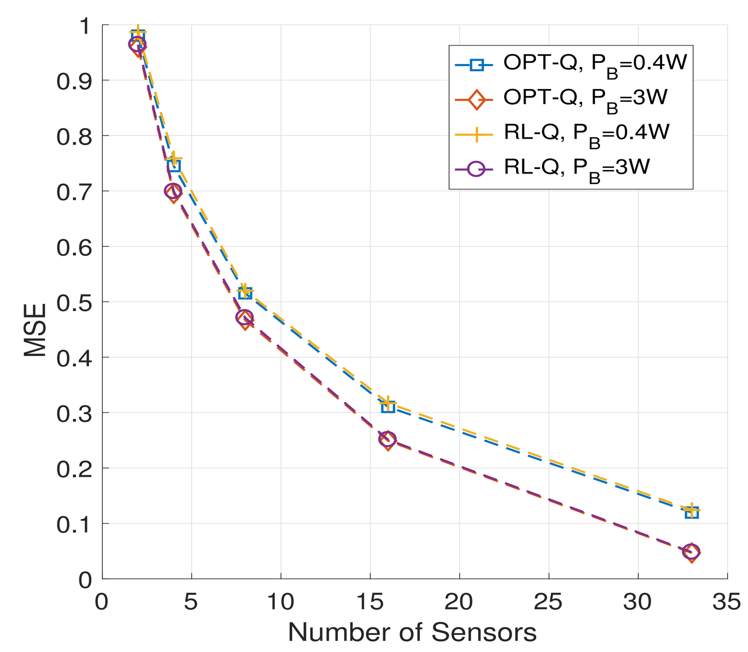
Set-up for the Experiments:
Random Field Model: Gaussian-Schell model (GSM) with time-varying parameters
Sensor Network: Energy beacon at (0,-1), sensors on the line at $y = 0$, sink at $(0, 4)$ (meters)
Aim: Estimate the unknown field values at $n = 33$ positions on the line at $y = 0$

MSE vs. Power Budget



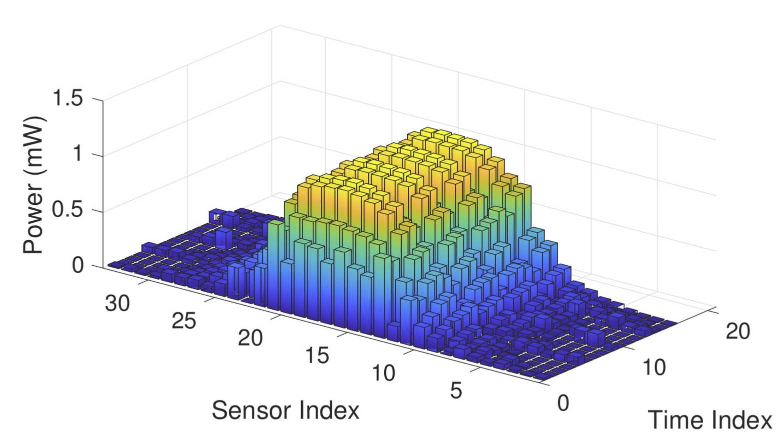
RL approach successfully learns to minimize the MSE without a priori knowledge of system parameters

MSE vs. number of sensors



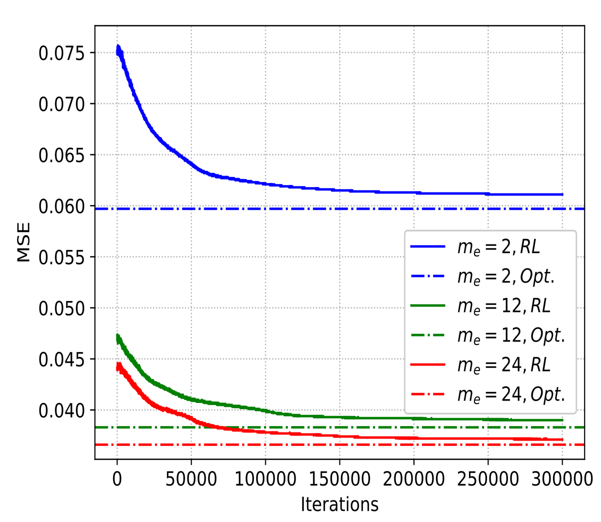
MSE depends significantly on the number of sensors (consistent with the fact that the unknown field becomes uncorrelated periodically)

Power Allocation for Communications to the Sink



Power allocation is time varying (consistent with the time varying nature of the field correlation)

RL Convergence



With four 3.5 GHz cores and a Quadro K620 GPU, direct optimization and RL (10^5 iterations, utilizing GPU) takes 15 and 62 minutes, respectively.

Acknowledgements

A. Özçelikkale acknowledges the support from Swedish Research Council under grant 2015-04011. M. Koseoglu acknowledges the support from Fulbright Program with grant number FY-2017-TR-PD-02.