

# Sustainable Smart City Building-energy Management Based on Reinforcement Learning and Sales of ESS Power

Dae-Kug Lee<sup>1</sup>, Seok-Ho Yoon<sup>2</sup>, Jae-Hyeok Kwak<sup>1</sup>, Choong-Ho Cho<sup>1</sup>, and Dong-Hoon Lee<sup>1\*</sup>

<sup>1</sup> Department of Computer and Information Science, Korea University,  
2511 Sejong-ro, Sejong, 30019, Korea

[e-mail: daekuglee@korea.ac.kr, wlsansdl123@korea.ac.kr, chcho@korea.ac.kr, kinno21@korea.ac.kr]

<sup>2</sup> Department of Artificial Intelligence Big Data, Sehan University,  
33 Sehandae-gil, Sinpyeong-myeon, Dangjin-si, Chungcheongnam-do, Korea  
[e-mail: bluepig@sehan.ac.kr]

\*Corresponding author: Dong-Hoon Lee

*Received November 7, 2022; revised January 2, 2023; accepted March 16, 2023;  
published April 30, 2023*

---

## Abstract

In South Korea, there have been many studies on efficient building-energy management using renewable energy facilities in single zero-energy houses or buildings. However, such management was limited due to spatial and economic problems. To realize a smart zero-energy city, studying efficient energy integration for the entire city, not just for a single house or building, is necessary. Therefore, this study was conducted in the eco-friendly energy town of Chungbuk Innovation City. Chungbuk successfully realized energy independence by converging new and renewable energy facilities for the first time in South Korea. This study analyzes energy data collected from public buildings in that town every minute for a year. We propose a smart city building-energy management model based on the results that combine various renewable energy sources with grid power. Supervised learning can determine when it is best to sell surplus electricity, or unsupervised learning can be used if there is a particular pattern or rule for energy use. However, it is more appropriate to use reinforcement learning to maximize rewards in an environment with numerous variables that change every moment. Therefore, we propose a power distribution algorithm based on reinforcement learning that considers the sales of Energy Storage System power from surplus renewable energy. Finally, we confirm through economic analysis that a 10% saving is possible from this efficiency.

---

**Keywords:** AI, Big Data, Building-energy Management, Energy Storage System, Reinforcement Learning, Renewable Energy, Smart City

---

This work was partly supported by Korea Institute of Energy Technology Evaluation and Planning(KETEP) grant funded by the Korea government (MOTIE) (2019271010015C, Development of Control System for Smart City Energy Consumption Operation Management) and the Korea Agency for Infrastructure Technology Advancement(KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2020-KA157018).

## 1. Introduction

**R**esearch on the use of renewable energy and the efficient use of energy is continuously conducted as a solution to the depletion of fossil fuels and for the reduction of greenhouse gases. Considering the current levels of resource extraction, oil is expected to be depleted in about 53 years, coal in 139 years, and natural gas in 49 years [1]. Therefore, various renewable energy markets, such as solar, tidal, and geothermal power, have begun to attract attention. According to an IRENA report in 2021, the current global renewable energy market in 2020 achieved 261GW of new renewable energy power-generation capacity, an increase of 10.3% compared to the previous year, with a cumulative record of 2,799GW [2].

Solar power is leading the expansion of renewable energy supplies, taking first place among all renewable energy sources with 127GW from new facilities, a 22% increase compared to 2019. Despite the increase in the manufacturing cost of solar modules due to the sharp rise in raw material prices, 160GW were expected from new installations in 2022 (a 17% increase compared to 2021), and a total of 1,100GW is expected from new installations between 2021 and 2026 [3].

Energy efficiency is essential to reducing carbon dioxide emissions and supplying new and renewable energy, and continuous technological development is required. Improving energy efficiency (37%) and the dissemination of renewable energy (32%) are the primary means to reduce carbon emissions below 10GtCO<sub>2</sub> by 2050 [4] [5]. To this end, several countries, including the United States, the United Kingdom, and Germany, are establishing and implementing demand-management policies in various fields, such as industry, construction, transportation, and equipment manufacturing.

In particular, the Energy Storage System (ESS), which can store electricity produced from renewable energy for use whenever necessary, can control the grid connection and is an optimal technology and facility for efficient energy use [6]. Recently, the ESS industry has been dramatically expanding along with the supply of new and renewable energy to respond to the increasing demand for electricity. And smart-ESS research is continuously conducted using various new technologies such as the Internet of Things (IoT), the Building Management System (BMS), and Building Information Modeling (BIM) [7].

Various studies have been conducted to propose efficient energy management systems based on time-of-use (TOU) tariffs in a multi-power environment with a variety of renewable energy sources, such as solar power and geothermal energy, and by using an ESS along with grid power [8] [9] [10].

The law has been amended to enable two-way electricity sales instead of the one-way sales structure in which consumers buy electricity from power generation companies. It is possible to create a two-way supply-and-demand system where energy can be bought and sold through the existing supply-and-demand system, which unilaterally supplies energy at a uniform unit price from Korea Electric Power Corporation (KEPCO).

This study was conducted because it was determined that more efficient energy use would be possible through the sale of electricity between users from the existing energy uses management using the electricity supply price only.

The main contents of this paper are as follows:

Section 2 analyzes energy consumption for five public buildings in Korea's first renewable-energy-convergence, eco-friendly energy town in Chungcheongbuk-do in the Republic of Korea. In this section, we analyze building energy usage patterns from data collected every minute for a year from five buildings: An Integrated Control and Management Center (ICMC), a public health center, a nursery, a library, and a high school.

Section 3 proposes a building-energy management model for a smart city in a multi-power environment where renewable energy and grid power are mixed. Using TOU and system marginal price (SMP) information, we design a system that can manage energy efficiently by considering power distribution to and from distributed power sources, from charging and discharging an ESS, and from sales of surplus power.

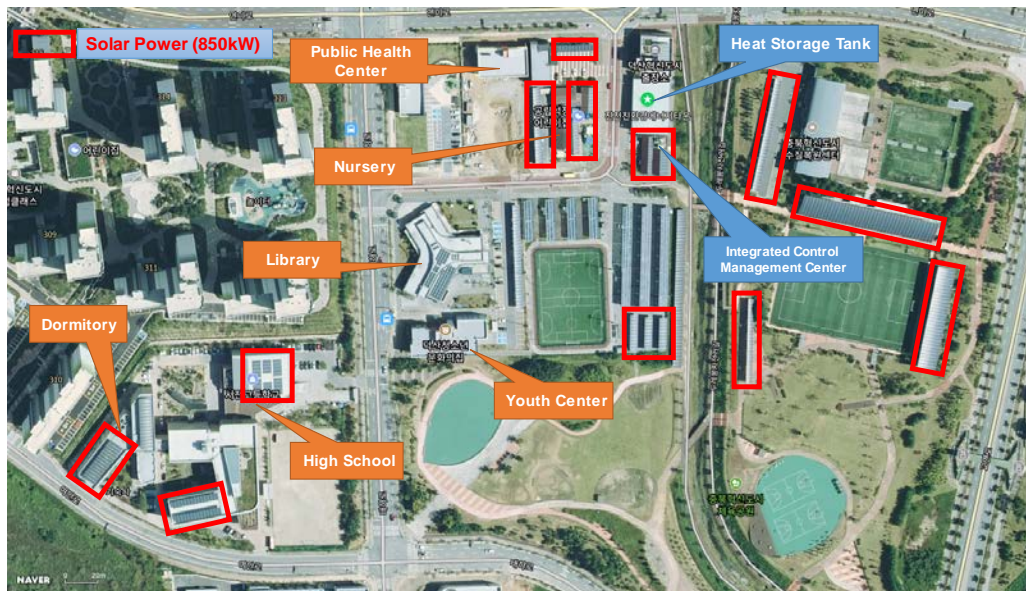
Section 4 proposes an AI reinforcement learning model to determine demand and the optimal cost of energy production. In this section, we design a smart city building-energy management platform and a reinforcement learning, smart city building-energy-management algorithm. Through a reinforcement learning model, the electricity use action of the building in the smart city is divided into four actions (Power use, sales, charging, and standby), and a cost function for the energy use environment is derived. Finally, economic feasibility is analyzed by applying the proposed SMP and reinforcement learning power management system to smart city energy data.

## 2. Analysis of Power Consumption in a Zero-energy Town

### 2.1. Eco-friendly energy in Korea

Chungbuk in Chungcheongbuk-do in the Republic of Korea was Korea's first new and renewable-energy convergent, energy-independent, eco-friendly project. Empirical research was conducted by introducing a small-scale centralized heat energy-supply system (block heating) concurrently using various new and renewable energy facilities.

Energy supply to the unit area through the convergence of new and renewable energy facilities had never been attempted in earnest in Korea. **Fig. 1** shows the green energy town of Chungbuk Innovation City from the sky.



**Fig. 1.** An eco-friendly town built in the innovative city of Chungbuk, Republic of Korea

As shown in **Table 1**, the eco-friendly town is a  $72,000m^2$  area composed of six public buildings; a plurality of new and renewable energy facilities, such as solar heat and photovoltaic power generation systems, are installed in the area. Seasonal thermal energy

storage and night thermal energy storage bridge the time gap between thermal energy supply and demand; a network supplying thermal energy, and a control and monitoring system are installed.

**Table 1.** Overview of six public buildings in Chungbuk Innovation City

Building	ICMC	Library	High School	Nursery	Health Center	Youth Center
Total Floor Area ( $m^2$ )	386.98	1,986	10,432	916	248	735.90
Building Area ( $m^2$ )	318.80	1,565	3,872	916	248	728.35
Air-Conditioned Area ( $m^2$ )	154.64	1,601	5,068	489	217	728.35
Stories	1	1	4	1	1	1
Watt-hour meters (ea.)	1	5	24	4	2	-
Sensors (ea.)	1	4	20	4	4	-
Calorimeters (ea.)	-	2	1	2	2	-

To design a system for efficient energy use in multiple buildings in a heterogeneous, distributed power environment within a smart city (rather than for a single building), the energy consumption of a zero-energy town was analyzed using data collected from sensors.

## 2.2. Building-energy usage analysis

We analyzed building energy usage patterns from data collected every minute over a full year for five buildings in the town (the ICMC, a public health center, a nursery, a library, and a high school). The youth center was excluded from the analysis due to the delayed completion of the building.

### 2.2.1. Average energy use by month, per hour, and per weekday

We compared average energy consumption per building by year, month, day, hour, and minute. There was no significant difference in average energy use by year, day, and minute.

However, analyzing the average energy consumption by month, hour, and day of the week found significant differences depending on the nature and characteristics of each public building.

Comparing the monthly energy consumption of these five buildings, [Fig. 2](#) confirms that the energy use patterns differed depending on the purpose of the building, who used it, and when it was used.

Looking at the average monthly energy consumption of the ICMC, there is no significant difference in monthly energy consumption because it operates 24 hours a day, 365 days a year. The public health center and the library had similarities in that they used less energy in summer and more energy in winter. It can be interpreted as much heat being used for patients visiting the public health center and for patrons using the library. The nursery showed energy consumption above a certain level, except in October and November, suggesting that heating and cooling were continuously used to maintain constant temperatures throughout the four seasons because children are vulnerable to heat and cold and the changing seasons. In high school, energy consumption was high due to the intensive use of air conditioners in August (the hottest month), but other energy consumption was similar.

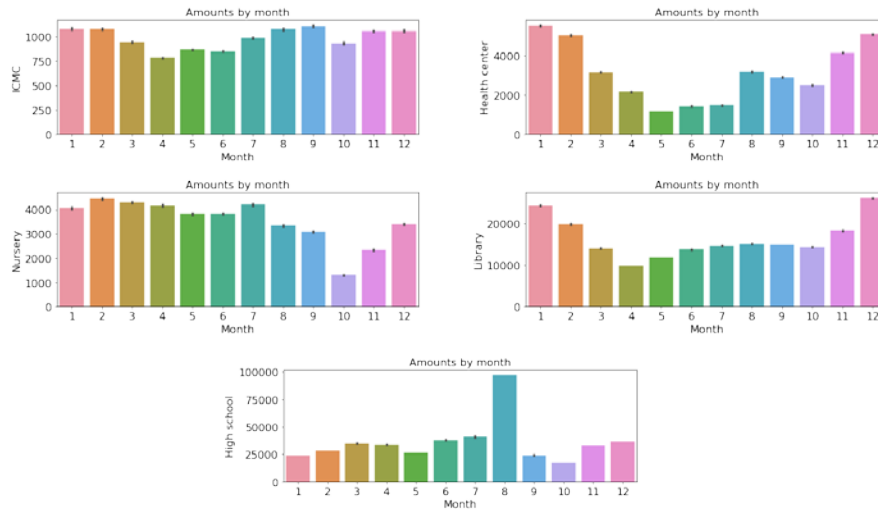


Fig. 2. Average monthly energy use for the five buildings

As seen in Fig. 3, energy consumption increased from 8 a.m. to 9 a.m. (when people go to work or school), briefly decreased at noon (lunchtime), continued until people left work or school, and decreased after that. In the library and high school, energy consumption continued until 9 p.m. because patrons used the library after work, and high school students studied after school.

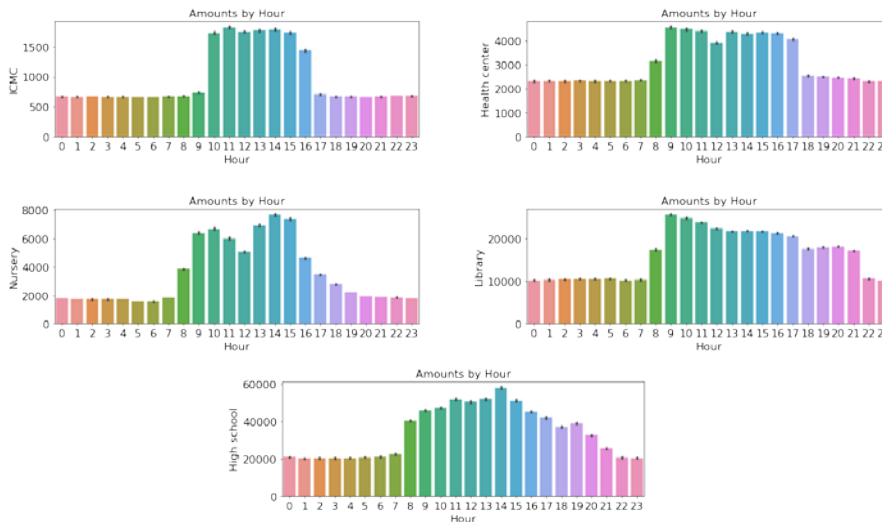


Fig. 3. Average hourly energy use for the five buildings

Looking at the average energy consumption by days of the week, as shown in Fig. 4, we can see that, except for the library, energy consumption significantly decreased on weekends because there was no commuting to/from work or going to/from school. However, the library had as many visitors on weekdays as it did on weekends.

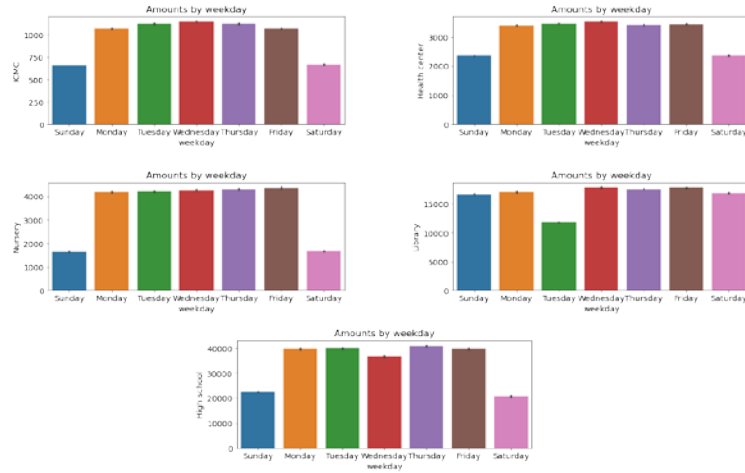


Fig. 4. Average energy use for the entire town by day of the week

### 2.2.2. Changes in building energy consumption by time period

To understand the town’s overall energy usage patterns, the trends in hourly changes according to the season, month, and day of the week were analyzed.

First, total hourly energy consumption was similar to the results of the previous analysis. As seen in Fig. 5, energy consumption increased sharply after 7 a.m. when it is the time to go to work or school and slightly decreased at noon when it is lunchtime. Consumption peaked at 2 p.m. and then gradually decreased until it was time to leave work or school.

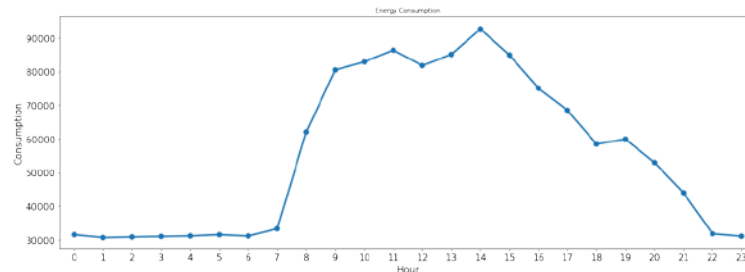


Fig. 5. Total hourly energy consumption of the town

Looking at total hourly energy consumption shows the characteristics of Korea’s four seasons very well. Fig. 6 shows that energy consumption was highest at 2 p.m. in the summer, and heating was used all day long in winter, as confirmed by energy consumption above a certain level.

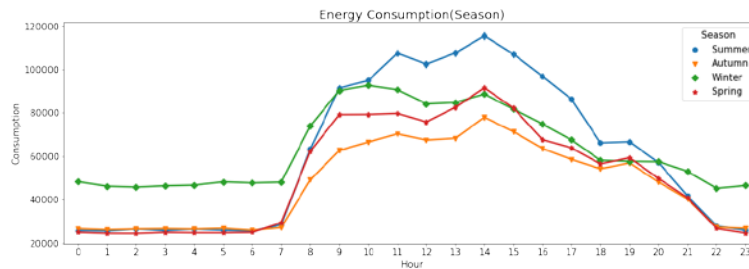
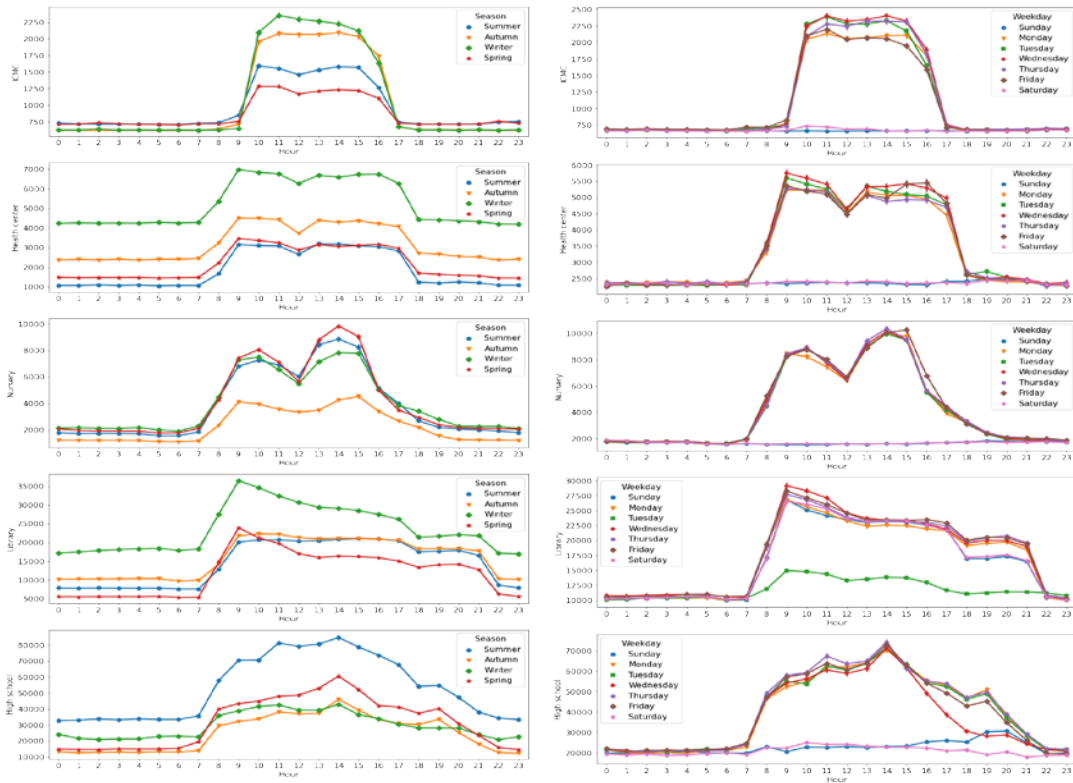


Fig. 6. Total hourly energy consumption in the town according to the seasons

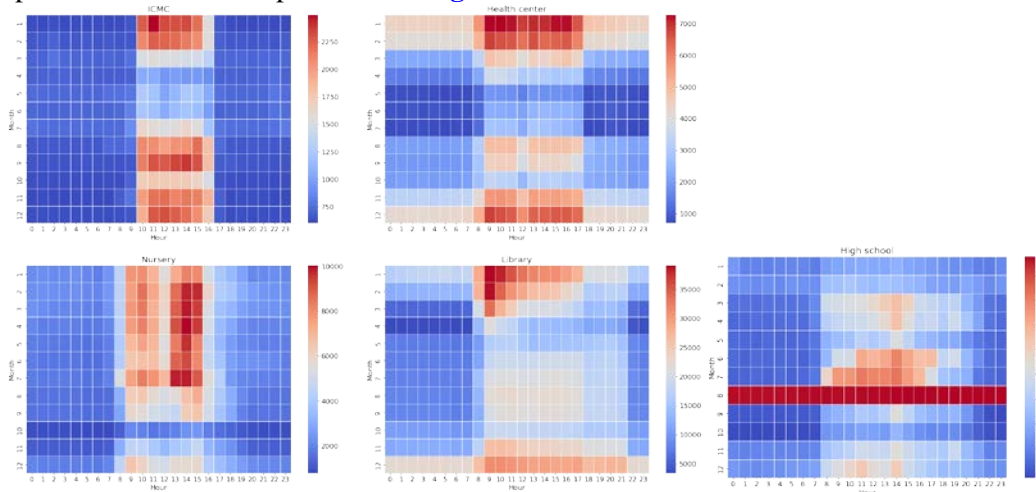
**Fig. 7** shows that each building has a different 24-hour energy usage pattern by season and day of the week.



**Fig. 7.** Average daily energy consumption for the five buildings per season and day of the week

### 2.2.3. Correlation analysis

To understand the correlations of energy consumption by month for each building, they are expressed in the heat maps shown in **Fig. 8**.



**Fig. 8.** Energy usage heat maps by month for the five buildings

Looking at the monthly energy consumption for each time zone, we can see that the results are similar to seasonal energy consumption. For example, the ICMC used the most energy in January at 11 a.m.; the public health center during January working hours; the nursery at 2 p.m. from February to April; the library at 9 a.m. in January, and the high school throughout August.

#### 2.2.4. Power production analysis

Changes in electricity production according to the season were analyzed, as shown in Fig. 9. Looking at the electricity production graph, it is intuitive that electricity can be produced while the sun is up and cannot be produced from sunset to sunrise when the sun is down.

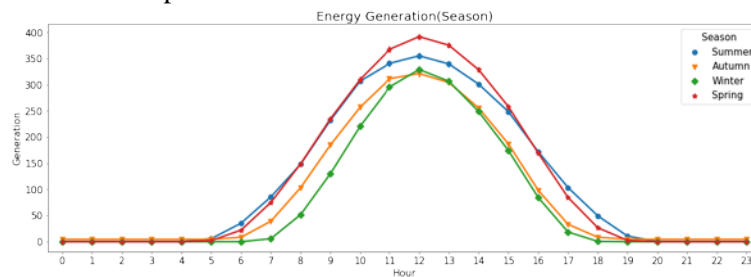


Fig. 9. Power production by the time of day in each season

#### 2.2.5. Justification for using reinforcement learning

In a smart city where renewable energies like solar power, geothermal energy, and fuel cells can be used together with grid power, it is necessary to design an efficient and economical building energy management system based on a heterogeneous distributed power environment.

Therefore, in various ways, we analyzed the energy consumption of five public buildings in an eco-friendly energy town (Chungbuk Innovation City, South Korea). As a result of the analysis, it was found that it was not easy to derive uniform rules for efficient energy use control because each building had different energy use patterns and irregular energy consumption.

Renewable energy sources do not pollute the natural environment and offer unlimited and permanent use. However, energy density is generally low and intermittent, making it challenging to produce electricity constantly. In particular, the amount of electricity generated from sunlight is greatly affected by weather conditions such as fog and rain, so production and charging cannot be constant and are impossible to predict. It is challenging to rely 100% on renewable energy alone, so using grid power is inevitable.

In addition, short-term power storage facilities such as an ESS that can store surplus power are required, along with grid power and facilities for long-term power such as converting and storing hydrogen.

The remaining energy after consumption (i.e., surplus power) can be sold at a profit, but the price is not constant daily. It is difficult to predict when it will be profitable to sell surplus power, when to use it, or when to use grid power instead.

Supervised learning that solves classification or regression problems from data with input values (problems) and output values (correct answers), or unsupervised learning that extracts features of the given input data to find structures or features and to select rules, is considered too difficult for solving this problem. We intend to solve it by applying reinforcement learning, which chooses a behavior that maximizes the reward from among the selectable behaviors in a constantly changing environment.



### 3. An Energy Management Plan for Excessive Energy Situations

#### 3.1. The proposed zero-energy town energy management model

In this section, we propose an efficient and optimal building energy management system by applying the power distribution method in a heterogeneous distributed power environment to a smart city and by considering the power supply unit price and power sales price based on time-of-use and system marginal price information [11].

The proposed system connects various renewable energy sources, such as solar power, solar heat, and geothermal energy, with an energy storage system. In an environment where multiple buildings in a smart city are managed in an integrated way, more efficient and economical power distribution is determined through sales of excess power based on the SMP and the power supply unit price [12]. Using reinforcement learning, we propose a method to reduce energy consumption costs through the design of power consumption, charging, standby, and sales models.

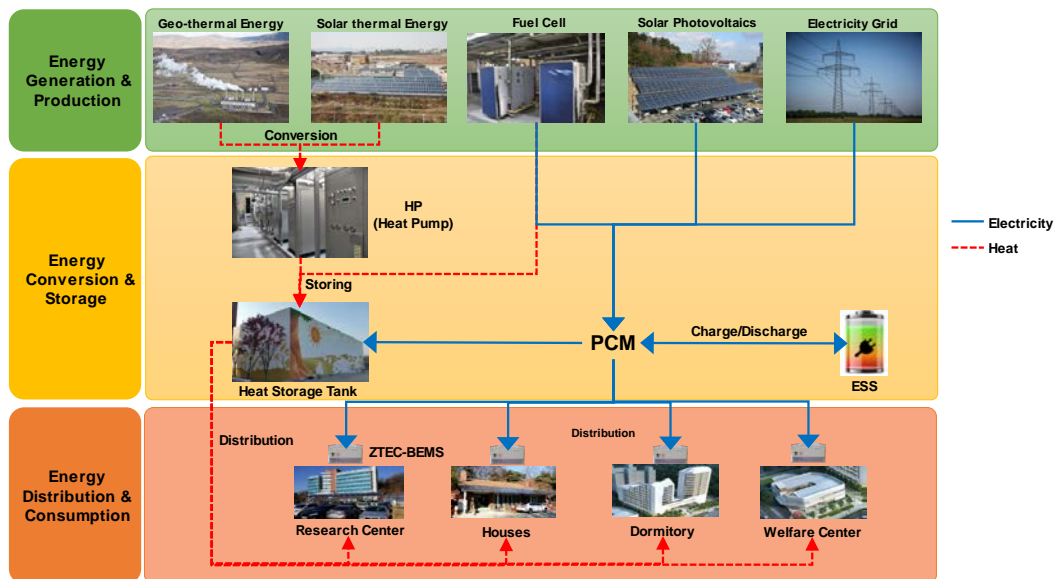


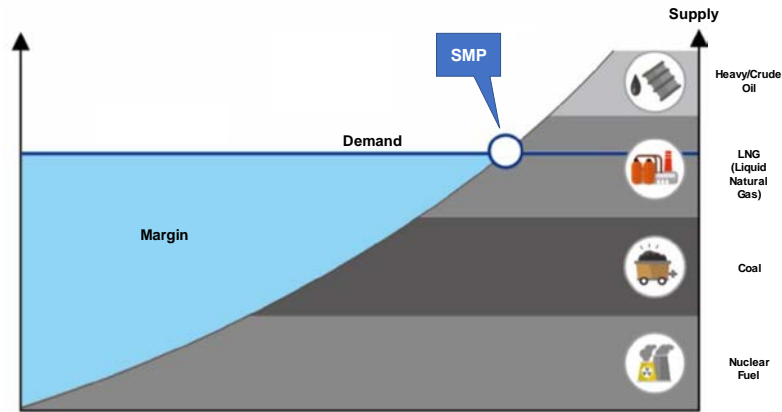
Fig. 10. The proposed smart city building-energy management model

Fig. 10 shows a diagram of a smart city for the application of the proposed building-energy management system in a heterogeneous power environment [13].

In this paper, a building energy management system using solar power is limited to using only the production and storage of electricity by using an ESS. The energy produced through solar power and the existing grid power is distributed to the buildings in the smart city by determining whether to use power, charge the ESS, or store power according to the integrated management system.

#### 3.2. Building energy management using SMP and TOU

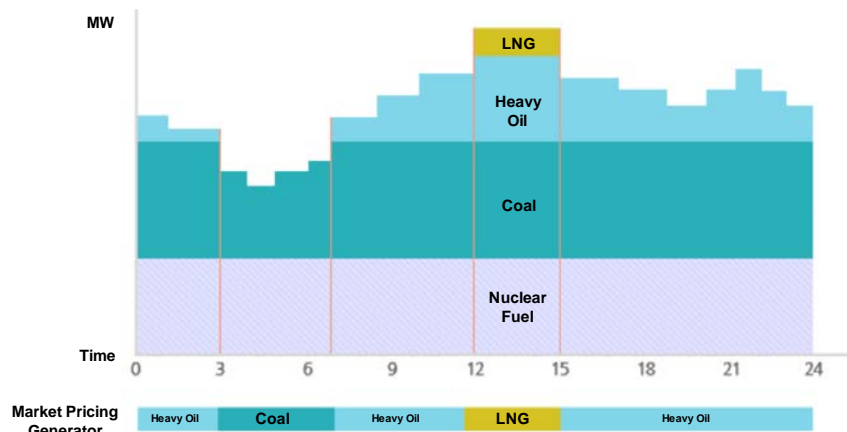
SMP refers to the price at which generated electricity is sold to the Korea Electric Power Corporation (the grid) through the Korea Power Exchange (KPX).



**Fig. 11.** Determination of Korea's electricity market price (the SMP) [14]

The electricity market price in the Republic of Korea is determined in hourly units on the day before the transaction. It is decided at the point where the demand predicted the day before meets the generator's power generation supply bid. The transaction price is determined at a point where the price desired by the buyer meets the price offered by the seller.

Vendors who sell electricity incur different costs when generating electrical energy from different sources. These costs vary depending on the type of power source (oil, LNG, coal, nuclear) and the cost of power generation. Therefore, the SMP is the price set for smooth sales and supply by setting an appropriate limit based on the average cost of power generation.



**Fig. 12.** Hourly electricity demand in the Republic of Korea determining the SMP [15]

The SMP offers the advantage of being able to plan for energy use because the price is provided for 24 hours on the day before power is needed. On the other hand, since the cost of each power generation source is considered first based on the SMP formulation principle, it is most affected by the global oil price and the government's energy policy, so the disadvantage is volatility.

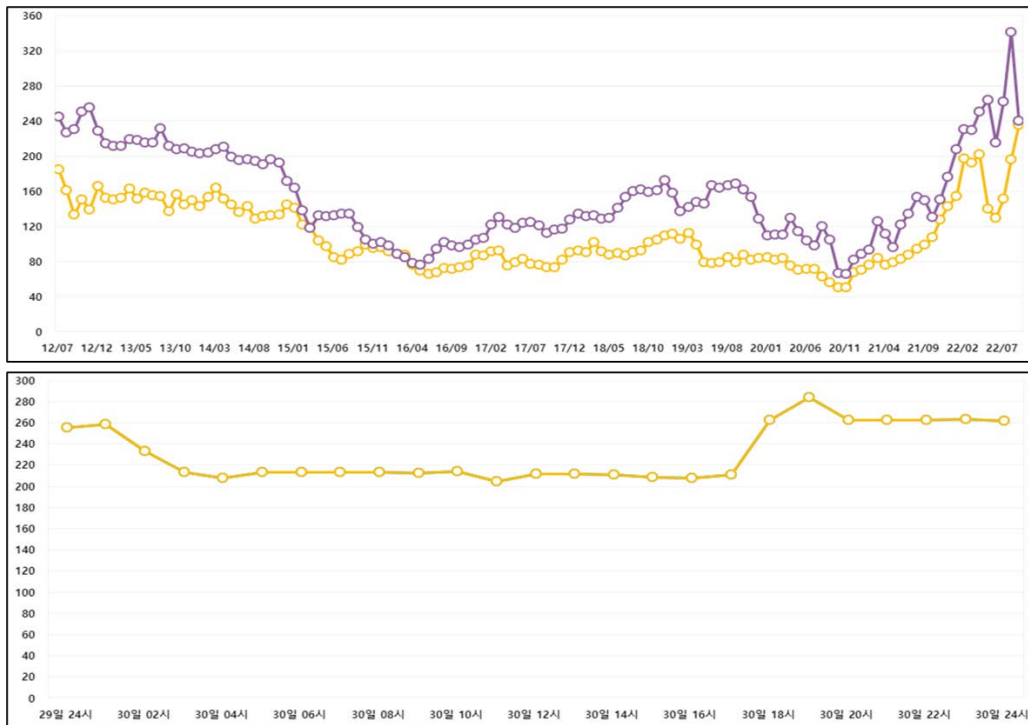


Fig. 13. Monthly SMP for 10 years, and the SMP over a 24-hour period [16]

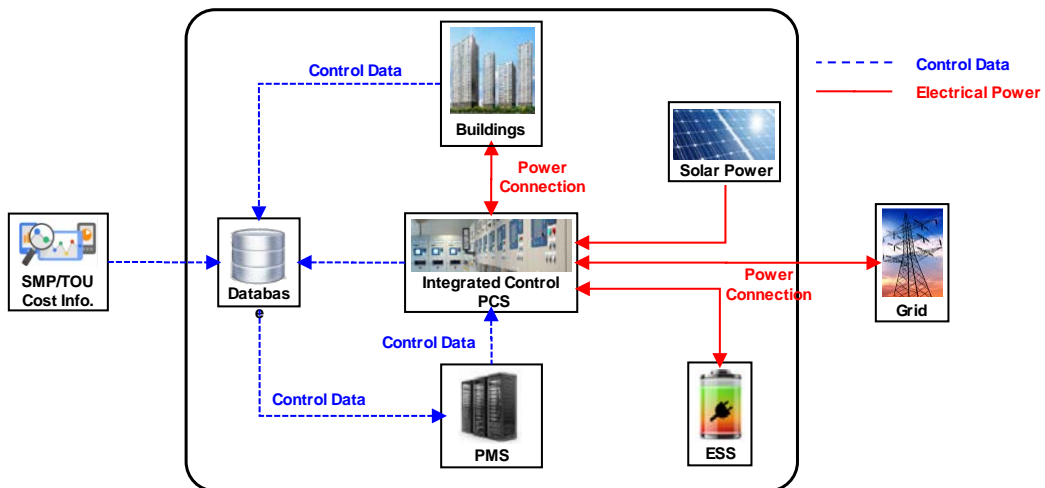


Fig. 14. Building-energy management system using SMP and TOU information in a heterogeneous distributed power environment.

Fig. 14 is a conceptual diagram of a building-energy management system for a smart city by using SMP and TOU information. The proposed building-energy management system monitors the distributed power, including the existing grid power, renewable energy such as solar power, and an ESS through an integrated Power Conversion System (PCS). A Power Management Server (PMS) provides the PCS with the optimal power distribution method based on SMP and TOU information. Through this, the use, charging, and sale of each distributed power source are determined, and power distribution is performed in a way that

maximizes energy efficiency.

### 3.3. Power distribution algorithm using SMP and TOU

For efficient power distribution in a heterogeneous distributed power environment, we propose an algorithm that applies power sales by considering SMP information as well as renewable energy, grid power, and the ESS.

The proposed power distribution algorithm's flow chart, shown in Fig. 15, is designed to determine the use or sale of energy, or charging the ESS, by comparing the current TOU information (the power supply price) and the SMP information (electricity sale price). It is intended to improve efficiency by adding sales of power to the existing power distribution method by using the TOU-information-based power supply standard time period.

In this paper, electricity sales are considered by using SMP information in the existing simple electricity distribution method, which was classified by time period based on demand. We propose an efficient and economical algorithm using reinforcement learning to determine the power usage pattern, and we compare and verify the performance against existing studies.

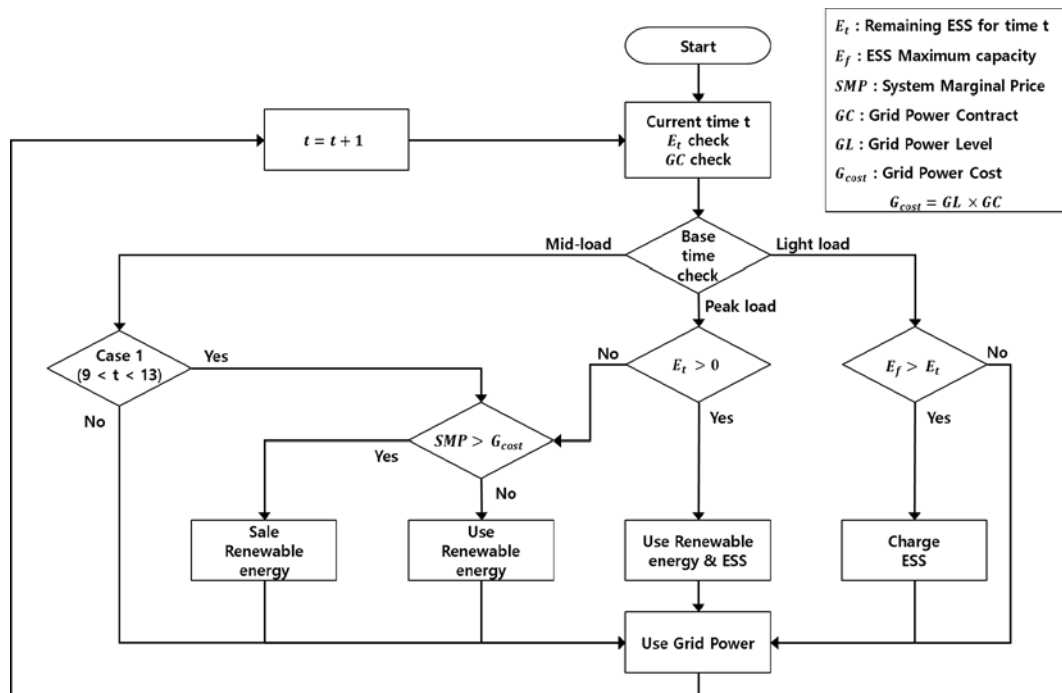


Fig. 15. Power distribution method comprising heterogeneous distributed power sources using SMP and TOU information [11]

Because the SMP is set based on the cost of each power generation source, it is inevitably affected by changing international oil prices and the government's energy policy. Since the SMP does not have characteristics of time series data, like the existing energy usage data, there is a limit to applying previously used prediction models.

Therefore, in this study, it was judged efficient and effective to apply a reinforcement learning model that generates the optimal policy by reflecting the experience of judgment to the problem [17] [18] [19].

**Table 2.** Comparison of a TOU-based algorithm and the proposed SMP power-distribution method

Power load	TOU-based Algorithm	TOU/SMP-based Algorithm
Light	Charge the ESS	
Mid-peak	<ul style="list-style-type: none"> <li>• 09:00~13:00: Use Renewable Energy</li> <li>• 18:00~23:00: Use Grid Power</li> </ul>	<ul style="list-style-type: none"> <li>• TOU&gt;SMP: Use Renewable Energy and ESS</li> <li>• TOU&lt;SMP: Sell Renewable Energy and ESS</li> </ul>
Peak	Use Renewable Energy and ESS	
ESS and Solar Power Unavailable	Use Grid Power	

## 4. Proposed AI reinforcement learning model for the optimal cost of energy production and demand

### 4.1. Reinforcement learning-based smart city building-energy management

Reinforcement learning is a type of machine learning. There is an agent and an environment, and learning is carried out in a way that maximizes the reward given according to the state of the environment as changed by the actor acting on the environment [20].

Model-based reinforcement learning is used when all information about a given environment, such as state, state transition probability, reward, behavior, and depreciation rate, is known. On the other hand, model-free reinforcement learning is used when only partial information is known [21].

If all information about the environment is known (Model Based), the value of the state value function of MDP (Markov Decision Process) can be calculated as follows, considering all actions and all states according to all fixed policies.

$$v_{\pi}(s) = \sum_{a \in A} \pi(a|s) R_s^a + \gamma \sum_{a \in A} \pi(a|s) \sum_{s' \in S} P_{ss'}^a v_{\pi}(s') \quad (1)$$

The Monte-Carlo Method is one of the most frequently used methods when information about the environment is insufficient (Model Free), such as when the Reward Function and State Transition Probability are not known, especially when the next state is not known. The state value function of MC (Monte-Carlo Method) uses the average of the returned values as the state value function.

First, the agent is operated on until the episode ends. After an episode ends, the count N is incremented by one. All return values collected during the episode are stored in variable S (cumulative return value). Unlike MDP, it does not consider all actions and states, only the actions performed during an episode and the states the agent has passed through. Finally, by dividing the cumulative return value by the cumulative count and taking the average, we can calculate the state value function. A fixed policy can be evaluated with the value of the state value function calculated in this way.

$$\begin{aligned}
v_{\pi}(s) &= V(s) \text{ when } N(s) \rightarrow \infty \\
\text{Cumulative Count: } N(s) &\leftarrow N(s) + 1 \text{ (perform one episode)} \\
\text{Cumulative Return: } S(s) &\leftarrow S(s) + G_t \\
\text{Average Return: } V(s) &\leftarrow S(s)/N(s)
\end{aligned} \tag{2}$$

Representative examples of model-free reinforcement learning include Q-learning and Policy Optimization techniques. In Q-learning, a value function is obtained to derive a policy in value-based reinforcement learning. Then, it finds the Q value to create a policy that selects the action that produces the largest Q value. However, this method is unsuitable for optimizing energy use because it always finds only the maximum value [22]. Therefore, this study used the policy optimization technique, a policy-based reinforcement learning that learns policies directly [23] [24].

The policy neural network objective function  $J(\theta)$  of Policy Gradient is a function that can calculate the value that can be obtained from a selected action through the policy function  $\pi_{\theta}(s, a)$  composed of the same variable  $\theta$ . Since it means the value of the policy, the maximum value is found using the gradient ascent method.

Considering only a one-time step in the MDP value function, it can be expressed as the following value function.

$$v_{\pi}(s) = \sum_{a \in A} \pi(a|s) R_s^a \tag{3}$$

Expressed as a policy objective function using the value function of One Step MDP, it is as follows.

$$J(\theta) = \sum_{a \in A} \pi_{\theta}(a|s) R_s^a \tag{4}$$

Policy Gradient uses the gradient ascent method to find  $\theta$  that can make the result of the policy objective function the largest.

$$\nabla_{\theta} J(\theta) = \sum_{a \in A} \nabla_{\theta} \pi_{\theta}(a|s) R_s^a \tag{5}$$

The policy objective function can be expressed as follows using the concept of the likelihood ratio.

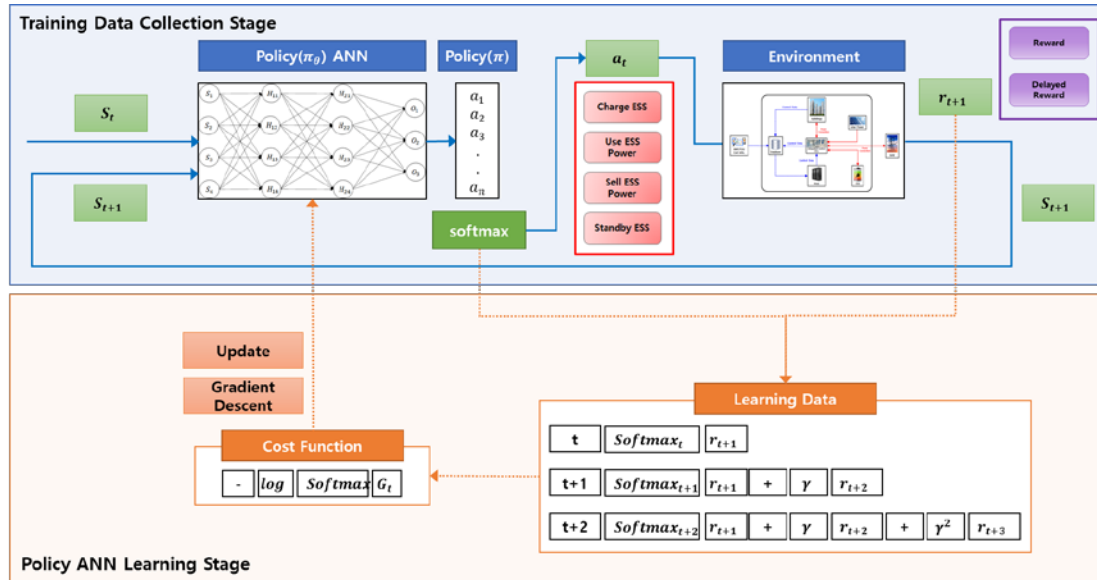
$$\nabla_{\theta} J(\theta) = \sum_{a \in A} \pi_{\theta}(a|s) \nabla_{\theta} \log \pi_{\theta}(a|s) R_s^a \tag{6}$$

Also, by changing the first part of the above formula to the expected value, it can be expressed as:

$$\nabla_{\theta} J(\theta) = E_{\pi_{\theta}}[\nabla_{\theta} \log \pi_{\theta}(a|s) R_s^a] \tag{7}$$

Using stochastic gradient descent (SGD) using sampling, the following policy gradient objective function can be obtained.

$$\nabla_{\theta} J(\theta) \approx \nabla_{\theta} \log \pi_{\theta}(a|s)r \quad (8)$$



**Fig. 16.** The proposed reinforcement learning-based building-energy Management platform

**Fig. 16** shows the proposed building energy management platform in which the Monte Carlo policy gradient algorithm, one of the policy gradient algorithms, was used.

As a component of this platform, there is a Policy Artificial Neural Network (Policy ANN) that expresses policy, and there is a policy that is the output value of Policy ANN. This policy consists of the result of the Softmax function. There is an environment where the action ( $a_t$ ) obtained through the policy is executed, and a learning data area stores the reward (including delayed reward) obtained through the policy and environment. Finally, it consists of a cost function for policy ANN training.

The Monte Carlo policy gradient algorithm consists of two steps. One is to collect training data while running the agent until the episode ends, and the other is to learn the policy ANN using the collected training data.

First, the state ( $S_t$ ) is entered into the Policy ANN to determine the agent's action. In the state, TOU cost information and SMP cost information data, as well as energy usage and renewable energy generation information, are transmitted to the system. Then, the Policy ANN executes the Softmax function and returns a policy. After the episode ends, the result value of the policy ANN is stored in the learning data area to be used for learning the policy ANN.

Then, the action with the largest weight ( $a_t$ ) is selected so that the agent can operate in the environment. The agent uses the policy of the Policy ANN to calculate the gain for actions such as energy use, charging, waiting, and selling and decides whether to perform it.

The environment returns a reward ( $r_{t+1}$ ) and a state ( $S_{t+1}$ ) as a result of the action performed by the agent. Rewards are also stored in the learning data area for policy ANN learning. The previous process is repeated until the episode ends by entering the following state ( $S_{t+1}$ ) into the policy ANN.

After the episode, the policy ANN is learned using the data stored in the learning data area. During the episode, the Softmax function result value and the return value for the reward (cumulative reward discounted by the depreciation rate) must have accumulated as much as the time step the agent operated during the episode.

Then, a pair of Softmax function results and the return value are taken from the training data area and inputted into the cost function to calculate the value. Finally, by adding a minus sign in front of the cost function, we find the maximum value of the policy function represented by the policy ANN through gradient descent. Then, while learning in the direction of minimizing the cost function, it learns as a policy ANN that expresses a more efficient policy.

The proposed energy management system determines power distribution for heterogeneous distributed power sources and performs efficient energy management by considering excess-power sales based on the SMP. In addition, the cost function can be derived through reinforcement learning, and the energy-saving effect can be predicted.

#### **4.2. The building-energy management algorithm**

In this paper, an efficient model for ESS use is proposed and implemented in a reinforcement-learning-based heterogeneous distributed power environment. The model divides power consumption behavior into four operations (power use, sales, charging, and standby), and a cost function is created considering the SMP and TOU systems.

By applying power usage behavior and ESS power status to the actions and rewards from reinforcement learning, the power usage action is taken in actual building energy use, and the cost function for the energy usage environment is derived. The model was designed to economically judge the result of the cost function according to behavior and learn it to increase profits.

In this paper, cost gains and losses are calculated from the decision on whether to use the ESS based on what was learned, on the actual ESS power consumption, on all the power used, on the current and past TOU prices, and on the current SMP price. Immediate/delayed rewards for actions performed in the present/past are determined according to the cost gain and loss, and they are reflected in the neural network.

Rewards are either immediate or delayed. An immediate reward is based on a result that can be confirmed immediately, and a delayed reward is collected when results that could not be confirmed at the time are confirmed later. Determination of the delayed reward is made with a separate reward criterion. The delayed reward is governed by a certain threshold and is determined when the gain or loss obtained through the learning process exceeds the threshold.

Immediate rewards are reflected in the policy neural network and learning proceeds. In the case of a delayed reward, the environment states, and the actions performed until receiving the delayed reward, are generated as batch training data, and the neural network is updated by applying the generated batch training data.

After learning, the weights of the neural network are updated, and the result of the updated neural network is reflected in the subsequent execution process. If the standard for a delayed reward is set low, training is likely undertaken multiple times with a small batch of data. When the standard for the delayed reward is set high, batch learning is performed a small number of times using a large amount of batch learning data.



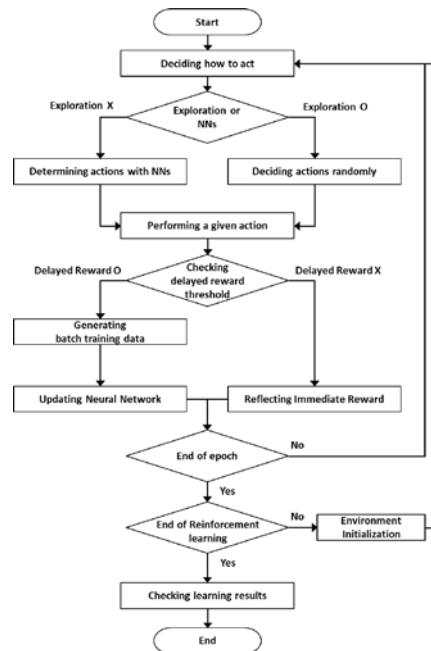


Fig. 17. Flowchart of the reinforcement learning process [18]

### 4.3. The smart city building-energy management module

#### Nomenclature

$t$	Time in the current state
$U$	A variable containing the action of the ESS as determined by policy <ul style="list-style-type: none"> <li>- Use: Amount of ESS power consumption</li> <li>- Charging: Amount of ESS power charge</li> <li>- Sales: Amount of ESS power sales</li> </ul>
$Tou_t$	TOU base price at time $t$ (kWh/KRW)
$Smp_t$	SMP reference price at time $t$ (kWh/KRW)
$EssTou_t$	ESS power consumption TOU price at time $t$ (kWh/KRW)
$E_t$	Total amount of energy used at time $t$
$ActionResult_t$	Cost as a result of the action performed at time $t$ (KRW)
$EssC_t$	Cumulative cost of ESS power charging at time $t$ (KRW)
$EssV_t$	Amount of ESS power at time $t$ (kWh)
$EssR_t$	Rate of ESS power retention at time $t$
$PV_t$	ESS usage gain at time $t$ (KRW)
$BPV_t$	ESS usage gain in a past time (KRW)
$PVR_t$	Proportion of gain by using ESS
$BC_t$	TOU cost used at time $t$ $E_t \times Tou_t$
$BCC_t$	TOU accumulated cost used up to time $t$ $BCC_t = BCC_{t-1} + BC_t$
$ACC_t$	Accumulated cost of power consumption action up to time $t$
$RewardRatio$	Weight of fluctuation (ex. -5% or +5%)
$ActionRatio_t$	The fluctuation rate of the current ESS value compared to the standard ESS value

### 4.3.1. Action definition and implementation details

An agent performs a direct action on the environment, receives data defined in the environment, acts according to a set policy, and updates the policy while receiving immediate rewards and delayed rewards through the cost function.

The actions performed by the agent are classified based on using the power of the ESS, and are divided into four categories: use, sale, charging, and standby.

**Table 3.** Actions according to the type of ESS power use

Action	Building power usage status	ESS status	Grid power usage status
Use	ESS	ESS discharge (Use)	Not in Use
	ESS + Grid Power	ESS discharge (Use)	Use
Sale	Grid Power	ESS discharge (Sales)	Use
Charging	Grid Power	ESS charging	Use
Standby	Grid Power	ESS standby	Use

The agent takes action after receiving from the neural network a decision on ESS energy consumption based on the action to be performed. In addition, it is necessary to determine if the corresponding action can be performed; if it is possible, it is performed. If it cannot be performed, the instruction given by the policy neural network is rejected, and ESS reverts to standby.

Determination of how much power from the ESS is used is reflected in the confidence value, which is a probability value indicating the reliability of the action taken according to the exploration or selection by the neural network, and it is necessary to determine the percentage of power to use (expressed as U) from between the minimum and maximum usage values.

Actions to use ESS power refer to states in which the building either uses the ESS for the required electricity or it uses the ESS and grid power at the same time. At this time, the ESS is in the discharge state (Use). After checking the amount of power remaining in the ESS, it can be used to cover some or all of the required power. This action is mainly performed under mid-peak and peak load conditions.

$$ActionResult_t = ((E_t - U) \times Tou_t) + (U \times EssTou_t) \quad (9)$$

The action for ESS power sales is where the building uses grid power for the required electricity, and it sells energy stored in the ESS to make a profit. At this time, the ESS is in a discharge state (Sales). This action is taken when the SMP cost is higher than the TOU cost and while power remains in the ESS.

$$ActionResult_t = (E_t \times Tou_t) - (U \times (Smp_t - EssTou_t)) \quad (10)$$

Charging the ESS is where the required electricity in the building is covered by grid power, and the ESS is charged with grid power for future use. At this time, the ESS is in the charging state. This action is taken when the ESS is not fully charged, and the TOU cost is lower than

the SMP.

$$ActionResult_t = (E_t \times Tou_t) + (U \times Tou_t) \quad (11)$$

The ESS in standby uses grid power in the building to cover the required power, and this state is for situations excluding power use, sale, and charging.

$$ActionResult_t = E_t \times Tou_t \quad (12)$$

#### 4.3.2. States and rewards

The agent's status is defined by the ESS's current power retention ratio and the energy use gain ratio.

$$EssR_t = \frac{EssV_t}{\left(\frac{EssC_t}{Tou_t}\right)} \quad (13)$$

Equation 13 shows the power retention ratio of the ESS: ESS power consumption at time  $t$  [over] ESS power consumption rate at time  $t$  [divided by] TOU standard price at time  $t$ . It is calculated to obtain the ESS power retention ratio.

$$PVR_t = \frac{PV_t}{BPV_t} \quad (14)$$

Equation 14 shows the gain ratio according to use of the ESS: (Gain of ESS use at time  $t$  / Gain of ESS use at time  $t-i$ ).

Rewards and delayed rewards for the action performed are determined using the gain from ESS use at time  $t$  and the gain from use of the ESS before time  $t$ , respectively. The states in which an immediate benefit can be obtained from ESS power use are the use of ESS power and the sale of ESS power. The rewards obtained from ESS power charging and standby cannot be collected immediately, but both must be paid as a delayed reward depending on the subsequent situation. To pay a reward and a delayed reward, the power use gain of the current time is defined.

$$PV_t = (BCC_t + EssC_t) - ACC_t \quad (15)$$

To derive the power use gain due to the time  $t$  action, calculate [(accumulated TOU cost used up to time  $t$  + the accumulated amount of power charging charges of the ESS remaining after performing the action time  $t$ ) – the power usage action used up to time  $t$  cumulative cost].

By reflecting the cost to charge the ESS compared to the fee obtained by using only the TOU rate system, the benefits obtained are confirmed.

$$ActionRatio_t = \frac{(PV_t - PV_{t-i})}{PV_{t-i}} \quad (16)$$

The fluctuation rate of the current ESS value relative to the ESS value based on time  $t$  was calculated as [(the gain of power use of action at time  $t$  – the gain of power use of action at

time  $t-i$ ) / the gain of power use of action at time  $t-i$ ].

The gain from power use from action performed at time  $i$  is updated whenever there is a decision to pay the delayed reward.

$$Reward = \begin{cases} 1, & \text{if } ActionRatio_t \geq 0 \\ -1, & \text{otherwise} \end{cases} \quad (17)$$

For an immediate reward, 1 is paid if  $ActionRatio_t$  is greater than or equal to 0, and -1 is paid in other cases.

$$DelayReward = \begin{cases} 1, & ActionRatio_t > RewardRatio \\ -1, & ActionRatio_t < RewardRatio \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

For a delayed reward, the defined  $ActionRatio_t$  is paid by comparing it with  $RewardRatio$ , which is the range of compensation fluctuations defined by reinforcement learning. If  $ActionRatio_t$  is greater than  $RewardRatio$ , 1 is paid; if it is less than  $RewardRatio$ , -1 is paid; if equal, 0 is paid.

#### 4.4. Economic analysis of the proposed model

To verify an efficient building energy management system in a reinforcement-learning-based heterogeneous distributed power environment, the energy usage fee for the ESS power usage derived from the algorithm based on the time period defined in the TOU policy and the proposed reinforcement-learning-based model was compared.

Economic feasibility analysis was conducted using energy data collected from five buildings in operation in a smart city (zero-energy town): Chungbuk Innovation City, Chungcheongbuk-do, Republic of Korea. Using the cost function derived using reinforcement learning, the results of the efficiency analysis for total energy use of the buildings in the smart city for one year are as follows.

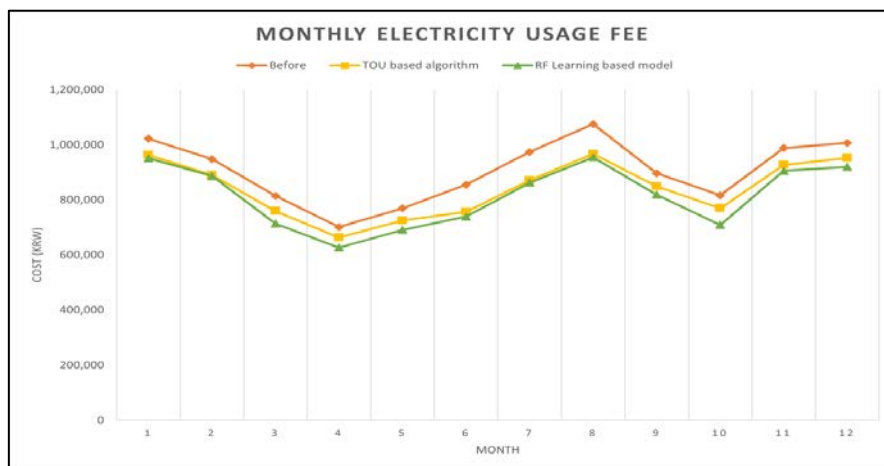


Fig. 18. Monthly electricity usage fees according to the ESS operation method

Fig. 18 shows a graph of monthly electricity usage rates according to the ESS operation method. As seen in the graph, both the TOU-based algorithm and the reinforcement learning

model show similar fluctuations.

Overall, we can see that reinforcement learning models have similar or less cost compared to the TOU-based algorithm in most cases. There is no significant difference between the two graphs in winter (January, February) and summer (June, July, and August) when electricity consumption is high due to heating and cooling, respectively.

However, we confirmed that the reinforcement learning model can obtain higher gains in spring (March, April, and May) and autumn (September, October, and November).

As seen from the comparison table of monthly electricity usage charges and the final result of the economic analysis, the ESS operation method based on reinforcement learning achieved 10.05% higher efficiency compared to the state without ESS operation, and 1,090,629 KRW could be saved. In addition, we confirmed 2.9% greater efficiency and that 315,940 KRW can be saved compared to TOU-based ESS operations.

#### **4.5. Analyzing the availability and complementation of the proposed model**

The result shows economic benefits from the proposed reinforcement learning-based energy management system model. In addition, reinforcement learning can solve even if information about the given environment is insufficient and the correct answer to the problem is unknown.

Each of the five buildings had different energy usage patterns, it was challenging to predict SMP (power supply unit cost) in advance, and it was not easy to find which ESS action was the best at what time and under what conditions.

This study shows the possibility of optimizing energy use through reinforcement learning by converging information from various sensors in a more complex energy environment in the future.

However, since the policy is derived only from data obtained through interaction with the environment, more data and countless learning attempts are required for the efficiency and stability of the model.

In addition, it is necessary to cope with the uncertainty of the model or to secure stability during parameter identification, as the risk burden is high to perform energy management control by applying it to an actual physical system.

The pre-designed reward function for reinforcement learning is a single scalar function, which makes it challenging to reflect complex systems. Therefore, finding an appropriate reward function according to various policies is necessary.

## **5. Conclusion**

In this paper, we analyzed building energy usage patterns by using data collected every minute for a year for five public buildings in Korea's first renewable energy eco-friendly energy town. Based on analysis results, an energy management model for a smart city was proposed.

An efficient and optimal smart city building-energy management system was proposed in consideration of the power supply unit price and sales price based on TOU and SMP information in a heterogeneous distributed power environment. Using reinforcement learning, ESS power usage, charging, standby, and sales models were designed, and a plan to reduce energy consumption costs was proposed.

That efficient energy management is possible (based on energy consumption, ESS power consumption, TOU price, SMP price, and the selection of an ESS operation method suitable for the time) was demonstrated using energy data collected from five buildings in operation in the smart city.

As interest in the use of new and renewable energy increases as a countermeasure against the depletion of fossil fuels and climate change, it is thought that various studies using an ESS as well as energy-saving measures must continue.

It is common sense to minimize electricity waste in existing power plants or from excess electricity generated. Efficient building-energy management in smart cities is expected to be possible through research on ESS utilization for power trading and sharing by various users in a heterogeneous power environment including ESS charging and discharging, by using electric vehicles, and through surplus power trading between users.

A sustainable smart city inevitably becomes a heterogeneous multi-power environment with various renewable energies, such as solar power, geothermal energy, and wind power, as well as from grid power. Building-energy management through electricity supply and electricity sales using an ESS between various types of buildings and various users is quite complex, and there are many variables to consider. For efficient and economical optimal energy management, the study of ESS charging/discharging models using reinforcement learning is likely to be utilized as a major technology for optimal ESS operations in the future.

### Acknowledgement

This work was partly supported by Korea Institute of Energy Technology Evaluation and Planning (KETEP) grant funded by the Korea government (MOTIE) (2019271010015C, Development of Control System for Smart City Energy Consumption Operation Management) and the Korea Agency for Infrastructure Technology Advancement(KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2020-KA157018).

### References

- [1] S. Dale, "BP Statistical Review of World Energy 2022," BP p.l.c. London, UK. [Online] Available: <https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2022-full-report.pdf>
- [2] IRENA, "Renewable Energy Statistics 2021," The International Renewable Energy Agency. Abu Dhabi, United Arab Emirates. [Online] Available: [https://mc-cd8320d4-36a1-40ac-83cc-3389-cdn-endpoint.azureedge.net/-/media/Files/IRENA/Agency/Publication/2021/Aug/IRENA\\_Renewable\\_Energy\\_Statistics\\_2021.pdf](https://mc-cd8320d4-36a1-40ac-83cc-3389-cdn-endpoint.azureedge.net/-/media/Files/IRENA/Agency/Publication/2021/Aug/IRENA_Renewable_Energy_Statistics_2021.pdf)
- [3] IEA, "Renewables 2021-Analysis and forecast to 2026," International Energy Agency. Paris, France. [Online] Available: <https://iea.blob.core.windows.net/assets/5ae32253-7409-4f9a-a91d-1493ffb9777a/Renewables2021-Analysisandforecastto2026.pdf>
- [4] S. Lee, "2022 KEA Energy Handbook," Korea Exchange Agency. Ulsan, Korea. [Online] Available: [https://www.energy.or.kr/web/kem\\_home\\_new/info/data/open/kem\\_view.asp?sch\\_key=&sch\\_value=&sch\\_cat=&c=306&q=23544](https://www.energy.or.kr/web/kem_home_new/info/data/open/kem_view.asp?sch_key=&sch_value=&sch_cat=&c=306&q=23544)
- [5] IEA, "World Energy Outlook 2019," International Energy Agency. Paris, France. [Online] Available: <https://iea.blob.core.windows.net/assets/98909c1b-aabc-4797-9926-35307b418cdb/WEO2019-free.pdf>
- [6] L. Al-Ghussain, R. Samu, O. Taylan, M. Fahrioglu, "Sizing renewable energy systems with energy storage systems in microgrids for maximum cost-efficient utilization of renewable energy resources," *Sustainable Cities and Society*, vol.55, pp.102059, 2020. [Article \(CrossRef Link\)](#)

- [7] F. Nasiri, R. Ooka, F. Haghghat, N. Shirzadi, M. Dotoli, R. Carli, P. Scarabaggio, A. Behzadi, S. Rahnama, A. Afshari, F. Kuznik, E. Fabrizio, R. Choudhary, S. Sadrizadeh, "Data Analytics and Information Technologies for Smart Energy Storage Systems: A State-of-the-Art Review," *Sustainable Cities and Society*, vol.84, pp.104004, 2022. [Article \(CrossRef Link\)](#)
- [8] S. Yoon, S. Kim, G. Park, Y. Kim, C. Cho, B. Park, "Multiple power-based building energy management system for efficient management of building energy," *Sustainable Cities and Society*, vol.42, pp.462-470, 2018. [Article \(CrossRef Link\)](#)
- [9] S. Zamanloo, H. A. Abyaneh, H. Nafisi, M. Azizi, "Optimal Two-Level Active and Reactive Energy Management of Residential Appliances in Smart Homes," *Sustainable Cities and Society*, vol.71, pp.102972, 2021. [Article \(CrossRef Link\)](#)
- [10] J. Song, "A study on Economic Analysis of ESS Charging/Discharging System for Building Energy Management," M.S. thesis, Dept. Comput. Inf. Sci., Korea Univ., Seoul, Republic of Korea, 2017.
- [11] J. Kwak, S. Yoon, M. Kim, D. Lee, C. Cho, "Power distribution method between distributed power sources considering SMP and TOU," in *Proc. of the 10<sup>th</sup> Int. Conf. Convergence Technology*, Jeju, Republic of Korea, Jul. 8-10, 2020.
- [12] [12] M. Kim, S. Yoon, C. Cho, "A Study on the Energy Optimal Control System Model in Smart City Considering SMP for Efficient Power Energy Management," in *Proc. of the Fall Conf. KISM*, Gwangju, Republic of Korea, Nov. 20-21, 2020.
- [13] S. Yoon, "A study of energy bigdata-based intelligent building energy management system," Ph.D. dissertation, Dept. Comput. Inf. Sci., Korea Univ., Seoul, Republic of Korea, 2019.
- [14] J. Ko, "Electricity Market Structure and Electricity Market Price (SMP) Determination," *KEPRI News*, vol.278, pp.32-37, May. 2019. [Article \(CrossRef Link\)](#)
- [15] GSOLAR, 2018. [Online]. Available: <https://tistorykan.tistory.com>
- [16] KPX, 2022. [Online]. Available: <https://epsis.kpx.or.kr>
- [17] S. Lim, Y. Son, S. Yon, "Reinforcement Learning-based ESS Scheduling for Cost Minimization," in *Proc. of the Fall Conf. Smart Grid Res. Soc.*, Gwangju, Republic of Korea, Nov. 7, 2019.
- [18] J. Kwak, "A study of efficient ESS charging/discharging model in a reinforcement learning-based heterogeneous distributed power environment," M.S. thesis, Dept. Comput. Inf. Sci., Korea Univ., Seoul, Republic of Korea, 2020.
- [19] D. Lee, Y. Kim, M. Jeon, C. Cho, "Efficient energy use management based on reinforcement learning considering power supply and sales," in *Proc. of the Summer Conf. ITFE*, Jeongseon, Republic of Korea, Aug. 24-26, 2022.
- [20] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2<sup>nd</sup> ed., Cambridge, MA, USA: MIT Press, 2018.
- [21] K. Mason and S. Grijalva, "A review of reinforcement learning for autonomous building energy management," *Computers & Electrical Engineering*, vol. 78, pp. 300-312, 2019. [Article \(CrossRef Link\)](#)
- [22] S. Kim, H. Lim, "Reinforcement Learning Based Energy Management Algorithm for Smart Energy Buildings," *Energies*, vol. 11(8), p. 2010, 2018. [Article \(CrossRef Link\)](#)
- [23] L. Yu, W. Xie, D. Xie, Y. Zou, D. Zhang, Z. Sun, L. Zhang, Y. Zhang, T. Jiang, "Deep Reinforcement Learning for Smart Home Energy Management," *IEEE Internet of Things*, vol. 7, no. 4, pp. 2751-2762, April 2020, [Article \(CrossRef Link\)](#)
- [24] L. Yu, S. Qin, M. Zhang, C. Shen, T. Jiang and X. Guan, "A Review of Deep Reinforcement Learning for Smart Building Energy Management," *IEEE Internet of Things*, vol. 8, no. 15, pp. 12046-12063, Aug, 2021, [Article \(CrossRef Link\)](#)



**Dae-Kug Lee** received his B.S. in Computer Science in 2004 from UCL, University of London, U.K. He received an M.S. in Computer Information Systems in 2018 from Korea University, Korea, where he is working toward a Ph.D. in the ICT Convergence Lab. His research interests include Self-Driving cars, Energy Management Systems, Smart Cities, Smart Housing, Blockchain, Deep Learning, and ICT convergence.



**Seok-Ho Yoon** received his Ph.D. in Computer Science in 2018 from Korea University, Korea. He is a professor at Sehan University in the Department of Artificial Intelligence Bigdata since 2021. His research interests include Smart city platforms, building energy management systems, ICT convergence, Future mobility, Deep Learning.



**Jae-Hyeok Kwak** received his B.S. in Computer Science in 2018 from Korea University, Korea. In 2020, he received his M.S. from the Graduate School of the same university. His interests are energy management systems, deep learning, and Smart Home.



**Choong-Ho Cho** received his Ph.D. in Computer Science in 1989 from INSA de Lyon University, France. He is a professor at Korea University in the Department of Computer Convergence Software since 1994. He has conducted various research projects in 4G/5G mobile networks and urban and building energy management. Currently, his main areas of interest are artificial intelligence-based self-driving mobility, smart city platforms, and building energy efficiency.



**Dong-Hoon Lee** works as a software and embedded hardware advanced developer. He is in a combined master's and doctoral course in the Computer and Information Science Department at Korea University in Korea. He was recently involved in various projects, such as greenhouse gas and energy data analysis, algorithms, and system development.