Journal of
*Low Power Electronics and Applications*

MDPI

*Article*

# A Q-Learning-Based Hierarchical Power Delivery Architecture for the Efficient Management of Heterogeneous Loads

**Andreas Tsiougkos *, Georgia Amanatiadou and Vasilis F. Pavlidis**

Electronics and Computer Engineering Department, School of Electrical and Computer Engineering, Faculty of Engineering, Aristotle University of Thessaloniki, Egnatia Odos, 54124 Thessaloniki, Greece; georaman@ece.auth.gr (G.A.); vpavlid@ece.auth.gr (V.F.P.)
* Correspondence: atsiougk@ece.auth.gr

**Abstract**

A new approach to end-to-end power delivery for increasingly sought-after hierarchical power delivery units (PDUs) is presented, improving the power efficiency of portable systems. The benefits of the technique are demonstrated through a PDU comprising multiple DC–DC converters, such as low-dropout regulators (LDOs), and the support of heterogeneous loads. A properly tailored Q-algorithm is combined with power gating to manage the power supplied by a multi-level PDU. The effectiveness of the proposed method is evaluated via a realistic PDU for different combinations of loads. The learning-based technique yields up to 13% higher total end-to-end power efficiency in the case of similar loads by utilizing four available LDOs compared to the case of a single LDO, which supports the same span of loads. Moreover, the proposed method improves power efficiency by up to 5% in the case of heterogeneous loads when compared to other autonomous state-of-the-art power management units.

**Keywords:** power delivery unit; Q-algorithm; heterogeneous loads; multi-level PDU; hierarchical PDU; power gating; machine learning; learning-based power management; PMU; portable systems

## 1. Introduction

The rapid market growth of Internet of Things (IoT) devices during the last decade has propelled research in power management units (PMUs) [1,2]. Challenges such as efficient power management, effective power management, and high end-to-end power efficiency for portable systems are amplified due to the wider range of loads (heterogeneous loads) that need to be supported by the PMUs.

Methods that adjust either the input impedance of power delivery units (PDUs) onto the output impedance of the power source [3,4] or the output impedance of the PDU onto the input impedance of the loads in order to maximize total power efficiency [5] have been developed to address these challenges. Typically, these methods are applied to PDUs that consist of a single DC–DC converter. The main limitations of these PDUs are the inherent low efficiency for small loads and, more importantly, the inability to supply dynamic and heterogeneous loads with appropriately high power efficiency.

Towards this direction, hierarchical power management [6] and distributed power management schemes [7] have been developed as a result of the increasing number of voltage domains needed for heterogeneous loads. Furthermore, multi-level hierarchical PDUs enable power gating [8] by powering off specific voltage domains, thereby yielding

Check for updates

Academic Editor: Maurizio Palesi

Received: 22 November 2025
Revised: 21 January 2026
Accepted: 25 January 2026
Published: 28 January 2026

**Copyright:** © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

lower power consumption and higher power efficiency [9]. However, the use in these hierarchical PDUs of multiple DC–DC converters, e.g., low-dropout regulators (LDOs), towards higher power efficiency and the support of heterogeneous loads has yet to be properly addressed.

In addition, in the case of similar loads, the currently known methods display moderate power efficiency due to the inappropriate criteria used to guide PDU policy (e.g., the power efficiency achieved in [5] is only 43%). In several methods, the minimum quiescent current of the LDOs is used to update the policy of the PDU, but this approach does not consider the load connected to the LDO. The power losses of the on-chip power grid or PCB traces are also significant factors that contribute to the overall power efficiency. Therefore, the consideration of the quiescent current alone yields moderate end-to-end power efficiency [5]. This situation highlights the need for novel approaches to improve energy efficiency, particularly in light of hierarchical PDUs. These approaches are developed at the circuit level and are compared vis a vis with the state of the art. This comparison is less relevant to the specific system-level implementation of these approaches, which yield different overheads. Consequently, this work emphasizes and demonstrates the need for new power management methodologies that can benefit from learning-based techniques.

Moreover, there are efforts relating to on-chip power management for homogeneous loads [10,11], which, however, cannot be used for power management in a multi-level system due to the different loads contained in these systems. A multi-level PDU is typically intended for heterogeneous loads; hence, the aforementioned methods need to be revisited. The most important challenges, when these methods are applied to multi-level PDUs, are the increase in both the form factor and overall power [11]. These limitations are due to the additional circuits required for measurements and calculations, since power should be measured at every intermediate level of the PDU for effective power management.

Alternatively, there have been techniques based on reinforcement learning (RL) that aim at high power efficiency over a wider span of loads [12–23], but again, only on-chip power management is considered. These methods do not aim at high-level implementation but rather at how the energy efficiency at the circuit level can be improved.

Li et al. propose an RL-based chip-specific power co-management scheme [24], which supports heterogeneous cores but is limited to on-chip power management. Adaptive power management (APM) for on-chip power management through adaptive clustering has also been proposed [25]. However, this method cannot be used for dynamic power management (DPM) in portable (battery-powered) systems, e.g., IoT nodes, due to the high runtime of the method [26]. Furthermore, there are power management approaches based on a discrete-time Markov Decision Process (MDP) with multiple power states [18,19]. However, the formulation of a linear optimization problem, which has to be solved in real time, makes those methods unsuitable for portable systems due to the related high energy overhead.

Towards this direction, a table-based Q-learning approach has been extensively utilized as one of the most popular RL algorithms because of its simplicity and strong convergence properties [23,27]. A table-based Q-learning approach has been used for power management in System on Chips (SoCs), either for maximizing CPU performance under different constraints [28,29] or minimizing CPU power consumption [30]. However, these approaches are either designed at the application level [28,29] or optimized for systems that are not supplied by batteries or harvesters and do not minimize energy overhead from computation and learning updates. These limitations are more evident in recent works exploring data-driven approaches [31,32]. In most cases, these approaches require significant computational resources and extensive offline training (e.g., thousands of epochs), making them unsuitable for resource-constrained systems [33–35]. Hence, these methods

are not directly applicable to energy-constrained, battery-powered systems with a strict power budget. These methods lower the consumed power, but they do not investigate or address how efficiently the energy is supplied to each of the loads (e.g., the cores of the processor) and, thus, address a different problem from this work. Specifically, this work aims to improve end-to-end power efficiency, which, similar to the state of the art, can also benefit from Q-learning, as shown in the results section.

This new technique is based on a broader view of the PDU, since this approach leads to higher benefits for the overall power management system, as presented in the results section. In addition, to the best of the authors' knowledge, the proposed work is the first attempt that considers a multi-level PDU for portable systems supplying heterogeneous loads.

The remainder of this paper is structured as follows. The related work and analytic formulation of end-to-end power efficiency are presented in Sections 2 and 3, respectively. The proposed technique for power delivery is described in Section 4, starting with the utilization of multiple DC–DC converters for homogeneous loads and followed by the general case of heterogeneous loads. Six real-life scenarios are investigated with and without the application of the proposed method, and experimental results are provided in Section 5, highlighting the usefulness of the method. Finally, some conclusions are drawn in Section 6.

## 2. Related Work

Although the terms power delivery and power management are often used interchangeably, they refer to distinct approaches to improving power efficiency [2–30,36]. Power management focuses on configuring loads in a fixed system to improve efficiency—i.e., maximizing performance with the available energy or minimizing power consumption through techniques, such as dynamic voltage scaling (DVS) and dynamic voltage and frequency scaling (DVFS) [15–17,30,37]. In contrast, power delivery defines how the system should be configured for given loads to optimize energy transfer—i.e., providing maximum usable energy to the load using methods such as power gating [8]. In other words, the PMU ensures that the delivered energy is best exploited, while the PDU ensures that the maximum energy is delivered.

Power management methodologies and related PDUs have been extensively investigated to support portable systems powered by batteries or energy harvesters, where long operational lifetimes are critical [3,8]. Some PMUs adapt to the intrinsic properties of the power source, such as output impedance, to maximize conversion efficiency [3]. Other methods optimize PDU efficiency by disconnecting inactive loads or regulating low-dropout regulators (LDOs) based on quiescent current for autonomous operation [4,5]. These approaches perform well for single or narrowly defined load ranges; however, the increasing demand to support heterogeneous and fast-varying loads cannot be met by present PMUs [10,11,36], leading to sub-optimal efficiency in emerging IoT and portable systems.

Early studies on dynamic power management (DPM) adopted stochastic and analytical optimization frameworks to minimize system power under performance constraints. Benini and De Micheli [12] and Paleologo et al. [13] introduced finite-state machine and Markov decision process (MDP)-based models for deriving optimal DPM policies. These were extended in [14] to formulate policy optimization problems solvable in polynomial time but dependent on accurate workload statistics. Despite their efficiency, such model-driven methods lack scalability and adaptability in heterogeneous and highly dynamic environments where workload characteristics vary rapidly.

Qinru and Pedram [18] proposed an MDP-based DPM scheme. However, solving real-time linear optimization problems renders it impractical for low-power systems due to excessive computational overhead. Dhiman et al. utilized multiple DPM policies selected by

a machine learning algorithm [19]. Debizet et al. presented a Q-learning-based PMU with eight power states, which highly restricts the exploration area [22]. Gupta et al. introduced a deep Q-learning-based power management approach for heterogeneous processors, which remains unsuitable for battery-powered platforms because of the overhead in power [23].

However, the set of policies or the number of power states is optimized for specific loads, yielding sub-optimum policies for dynamic loads. Alternatively, an infinite or a practically large number of sets or power states is formally required to improve power efficiency for heterogeneous loads. Moreover, the computational cost of more complex algorithms, which reaches up to 18,268 sec for adaptive clustering on a multi-core system of four Intel(R) Core (TM) i3-2120 CPU @ 3.3 GHz processors [25], is prohibitive for battery supply and, generally, low-power systems.

Furthermore, in order to measure total end-to-end power efficiency $\eta_{tot}$ in a multi-level system, as shown in Figure 1, the input power and the output power at every load should be measured. However, this approach increases both the form factor and overall power, as an additional circuit is required for measurements and calculations. On the other hand, once the efficiency of each power conversion level is characterized after fabrication, the total efficiency $\eta_{tot}$ can be estimated from load-side current measurements, provided that an accurate formula for $\eta_{tot}$ is derived based on the PDU hierarchy.

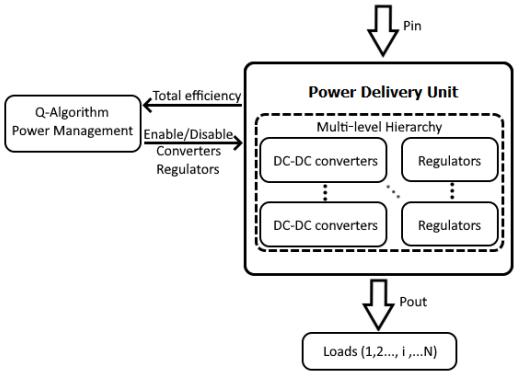



**Figure 1.** Block diagram of a system with a hierarchical PDU, managed by a Q-learning algorithm.

Recent efforts have applied RL to enable adaptive, model-free DPM. Debizet et al. [22] implemented a Q-learning-based adaptive power manager in hardware for IoT SoCs, achieving energy savings during suspend states. Kwon et al. [38,39] extended RL-based control to mobile multiprocessor SoCs, demonstrating significant power and latency improvements. Giardino et al. [40] proposed a software-level RL-based manager (2QoSM) that enhanced power efficiency and application QoS in embedded systems. Nevertheless, these RL-based approaches typically operate at a single system layer and manage homogeneous or limited-state components, lacking hierarchical coordination across power domains [22,39,40].

Consequently, a novel technique that supports adaptive PDUs with high power efficiency over heterogeneous loads is presented to address the aforementioned limitations through the following contributions:

- Real-time optimization framework: We introduce a properly tailored Q-learning algorithm that adapts the system in real time to any combination of heterogeneous loads, determining an operating point that yields higher end-to-end power efficiency than current state-of-the-art methods.
- Adaptive hierarchical PDU support: By effectively coordinating disparate DC–DC converters and power gating, the proposed method achieves significantly higher total efficiency across different supply voltages compared to single-converter systems.

- Dynamic load adaptation: Our online training methodology allows the system to maintain peak performance under changing load conditions, such as the addition of new modules or fluctuations in the transmitter power of radio modules.
- Low-overhead implementation for ULP systems: By combining a search table approach with efficiency estimation, the methodology eliminates the need for complex measurement circuitry, reducing the computational and power overhead to levels suitable for battery-powered IoT nodes.
- Low-overhead estimation for IoT: We utilize a search table and efficiency estimation approach that eliminates the need for redundant measurement circuitry, maintaining a computational runtime below 60 ms—ideal for battery-powered IoT nodes.

The proposed Q-learning-based hierarchical power delivery architecture introduces a multi-level control framework coordinating heterogeneous loads across hardware and system layers. Employing interacting RL agents at different levels enables distributed, load-aware decision-making that scales with system complexity. This unified architecture bridges local and global optimization, providing an adaptive solution for heterogeneous IoT platforms operating under diverse and dynamically varying workloads.

## 3. Background and Theoretical Formulation

Some basic theory of Q-learning, as well as the key parameters that determine the performance of the algorithm, are presented in this section. In addition, the definition of the terms and metrics used throughout the paper is provided.

### *3.1. Q-Algorithm*

Reinforcement learning is a subfield of machine learning that utilizes intelligent agents that operate in a finite environment to perform actions that maximize a specific reward [27]. Q-learning is a model-free RL algorithm in the sense that no specific model of the environment is needed. Q-learning is used to learn the value of an action $\alpha_t$ in a particular state $s_t$ of the system.

For any MDP, the algorithm finds an optimal policy, which maximizes the expected value of the long-term reward when the final state is reached [41]. This goal is achieved by a function that evaluates the quality of each $(s_t, \alpha_t)$ pair according to

$$Q'(s_t, \alpha_t) = Q(s_t, \alpha_t) + \alpha \cdot [r_t + \gamma \cdot max_{\alpha} Q(s_{t+1}, \alpha_t) - Q(s_t, \alpha_t)], \quad (1)$$

where $t$ stands for each iteration within an epoch, $\alpha$ is the learning rate, $r_t$ is the reward function, $\gamma$ is the discount factor, and $Q$ and $Q'$ are the current and the next value of the Q-matrix, respectively. The next value of $Q'(s_t, \alpha_t)$ is determined by the best action $\alpha_t$ taken that maximizes $Q(s_{t+1}, \alpha_t)$ according to (1). The total number of iterations needed by the agent in order to reach a final state is called an *epoch*. In general, the number of iterations is different for every epoch. The Q-matrix stores the reward function $Q$ and is of size $S \times A$, where $S$ is the number of states, and $A$ is the number of actions. There are several options regarding the initialization of the matrix to zero or random values [42] or pre-populating the Q-Matrix for faster convergence [43].

The learning rate $\alpha \in [0, 1]$ is also referred to as the step size and weighs prior knowledge over newly acquired knowledge and *vice versa*. As $\alpha$ decreases, the importance of current knowledge decreases as well, with $\alpha = 0$ yielding no learning at all. The factor $\gamma \in [0, +\infty)$ evaluates the significance of future rewards. In the case of $\gamma = 0$, the agent is "myopic" because only the current rewards are considered. A high long-term reward is achieved as $\gamma \to 1$. However, $\gamma \geq 1$ may lead to infinite loops due to the non-convergence of the Q-function. The reward function $r_t$ is critical for the performance of

the Q-algorithm [44]. In this work, the reward $r_t$ function is defined as the power efficiency of the converter used each time $\eta_{Nextnode}$ and is described in detail in Section 4.2. The formal problem statement is summarized as follows:

- Hypothesis: A model-free reinforcement learning (RL) agent can autonomously learn to maximize the end-to-end efficiency of a multi-level PDU without requiring prior knowledge of converter topologies or load profiles;
- Input parameters: The system state space is defined by (1) real-time load current demands ($I_{Load}$), (2) available voltage domains, and (3) battery status ($V_{bat}$);
- Expected outcomes: An optimal connectivity policy that dynamically delivers power through the most efficient combination of converters, thereby maximizing the total system efficiency $\eta_{tot}$ compared to static or heuristic control baselines.

The nomenclature used throughout this paper is defined in Table 1. The power efficiency metric used for the evaluation of the system is described in the following subsection.

**Table 1.** Definition of terms.

| Term | Description |
|---|---|
| Cascaded | Describe systems with components connected in series, thus forming a chain. The input port of the first and the output port of the last component are the input and output ports of the complete system, respectively. |
| Multi-level PDU | Describes a hierarchical PDU with multiple levels, including different cascaded DC–DC converters. |
| System | Refers to the multi-level PDU, along with the loads connected to the PDU and the control block, which implements the Q-algorithm, as shown in Figure 1. |
| $\eta_i$ | The power efficiency $\eta_i = \frac{P_{out_i}}{P_{in_i}}$ for every level $Level_i$ or load $I_i$ connected to the system shown in Figure 1. |
| (Total) end-to-end power efficiency | The overall power efficiency $\eta_{tot} = \frac{\sum_{i=1}^{L} P_{out_i}}{P_{in_1}}$ for $L$ loads $I_i$ connected to the system. |
| Quiescent current | The current drawn by an LDO when no load is connected. |
| Homogeneous loads | Refers to loads that exhibit similar power profiles over time, and their peak currents are close to each other. |
| Heterogeneous loads | Describes loads that draw considerably different currents, e.g., 10×. |
| Autonomous PDU | Refers to a PDU that operates without user intervention by utilizing a predefined power management scheme in order to achieve high end-to-end power efficiency [5]. |
| Power gating | Describes a technique that disables idle converters in order to reduce the overall power consumed by a PDU [8]. |
| Resonance frequency ($f_o$) | Used to describe the frequency of an electrical system in which the input impedance of the system is minimum or, equivalently, the amplitude of the output signal is maximum [45]. |
| $I_{Load}$ | The aggregate current drawn by the specific set of heterogeneous IoT peripherals (DHT11, Servo, and LoRa). Loads include sensors (DHT11, PIR, and HC-SR04), actuators (DC Motor and Servo), and communication modules (BLE, WiFi, and LoRa) representing diverse power profiles. These loads have been selected to represent extreme dynamic range variance (from 1 µA sleep currents to 360 mA motor stall currents), creating the non-linear disturbance the Q-learning agent must manage. |

### 3.2. Power Efficiency Decomposition

Modern integrated systems comprise different DC–DC converters, including buck converters [46,47], LDOs [48], or boost converters [49], that support heterogeneous loads [50] in different voltage domains and with different power demands [51]. Recent advance-

ments in LDO circuit design focus on optimizing the internal topology for stability and transient response [52–54]. These circuit-level optimizations are complementary to the system-level hierarchical management proposed in this work, which treats the converter as a functional block within a larger learning framework. Without loss of generality, LDOs are considered as the DC–DC converters in the following analysis, as LDOs are often used in cascaded PMUs [55]. In this work, LDOs are utilized as a testbench to validate the proposed framework. However, the Q-learning algorithm treats the power converter as a "black box" characterized solely by its load current and efficiency. Consequently, the optimization methodology is topology-agnostic and applies equally to other converters (e.g., Buck and Boost), as the Q-agent simply learns the unique efficiency profile of the connected hardware without requiring *a priori* knowledge of its internal operating characteristics. Each one of those LDOs exhibits a different power efficiency for the same load, as shown in Figure 2, which means there is an optimum combination of the converter–load pair that maximizes overall PDU efficiency. The theoretical efficiency curves for the depicted LDOs, each with a different maximum load current and $I_{qsc}$, also demonstrate degradation at light loads, when a regulator operates with a load much smaller than $I_{qsc}$. Based on this observation, a hierarchical PDU with disparate DC–DC converters can provide a means to constantly deliver the highest end-to-end power efficiency, as described in the following sections.

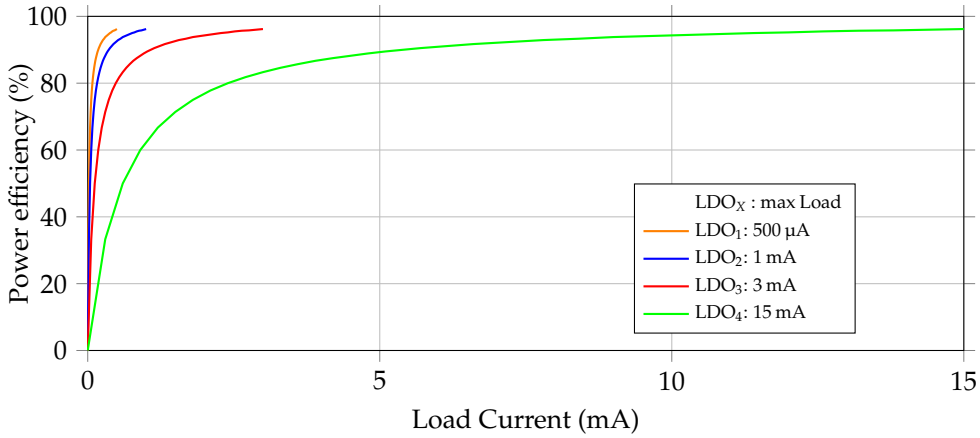



**Figure 2.** Power efficiency for four different LDOs.

Suppose, without loss of generality, that the LDOs of the system are sorted in ascending order of the maximum load current they can provide. The efficiency of the *i-th* LDO can be described by primary factors, which are, respectively, the load $I_{Load}$ and quiescent current $I_{qsc_i}$, as

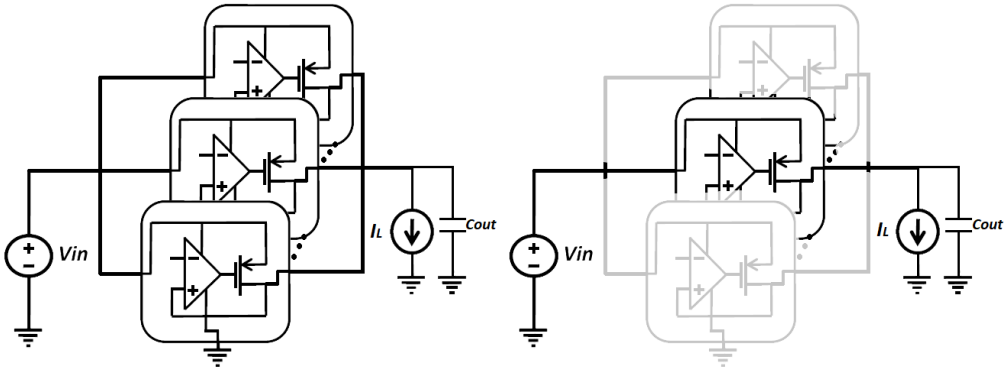
$$\begin{aligned} \eta_i &= \frac{P_{out_i}}{P_{in_i}} \\ &= \frac{V_{out_i} I_{out_i}}{V_{in_i} I_{in_i}} \\ &= \frac{V_{out_i} I_{Load_i}}{\lambda V_{out_i} (I_{Load_i} + I_{qsc_i})} \\ &= \frac{I_{Load_i}}{\lambda (I_{Load_i} + I_{qsc_i})}, \end{aligned} \tag{2}$$

where $\lambda$ is the ratio $\frac{V_{in}}{V_{out}}$, and, in case of high efficiency, LDO is, by definition, close to one [56]. Although LDOs are widely used for general voltage step-down tasks, in energy-critical systems aiming for maximum efficiency, they are primarily utilized as post-regulators following a switching converter. In this specific configuration, the input voltage is pre-regulated to slightly above the output voltage (e.g., creating a 3.3 V rail from

a 3.4 V intermediate rail). Under these design constraints, the voltage ratio $\lambda$ is maintained close to unity, minimizing the power dissipation across the pass element. Moreover, $\lambda$ is constant for all the available LDOs connected in parallel since they operate with the same input voltage $V_{in}$ and provide the same output voltage $V_{out}$ to the load. If the available LDOs, depicted in Figure 3a, are sorted by ascending order in terms of maximum load current they can provide, the overall efficiency of $N$ available LDOs is given by

$$\begin{aligned} \eta_{N_{LDOs}} &= f_{LDO}(I_{Load}, \sum_{i=1}^{N} I_{qsc_i}) \\ &= \frac{I_{Load}}{\lambda(I_{Load} + \sum_{i=1}^{N} I_{qsc_i})}. \end{aligned} \tag{3}$$

Therefore, the overall efficiency is highly dependent upon the total number $N$ of LDOs, although only one LDO is enabled at each point in time, as shown in Figure 3b. This behavior is expected since the quiescent current of each LDO is added to the power consumed by the PDU, yielding lower overall efficiency for the same load as $N$ increases. This situation means that beyond a certain number of LDOs, the power efficiency is worse than using a single large LDO, which provides the same maximum load current. Moreover, the form factor of the system or the on-chip area increases along with the increasing number of LDOs.




(**a**) (**b**)

**Figure 3.** Parallel LDOs for homogeneous loads. (**a**) Multiple LDOs available for a single load. (**b**) Only one LDO is enabled each time.

However, a greater number of LDOs ($N$) means that the load span is fragmented into more sub-regions where, in each of these regions, a specific LDO operates at the highest efficiency. This approach yields the maximum power efficiency for the given set of LDOs and the same span of loads. Therefore, there is a trade-off between $N$ and the total power efficiency of the system. This trade-off implies that the number $N$ should be chosen based on the following:

- The power efficiency of the available LDOs;
- The maximum total current $I_{\text{tot}} = \sum I_{Load_i}$ that must be supplied by the system.

In general, a mixture of on-chip and off-chip DC–DC converters typically satisfies the load current requirements [48]. Moreover, multiple DC–DC converters are needed to improve overall power efficiency, as mentioned previously. These converters cannot be flat but rather must be hierarchical due to the diverse form factor of the converters [6,9].

### *3.3. Power Efficiency of Hierarchical PMUs*

In a hierarchical PDU, the converters of the $i$-th level are connected to the converters on the $(i-1)$-th and $(i+1)$-th level, respectively, forming a power tree $T$. An example of a three-level PDU is depicted in Figure 4.

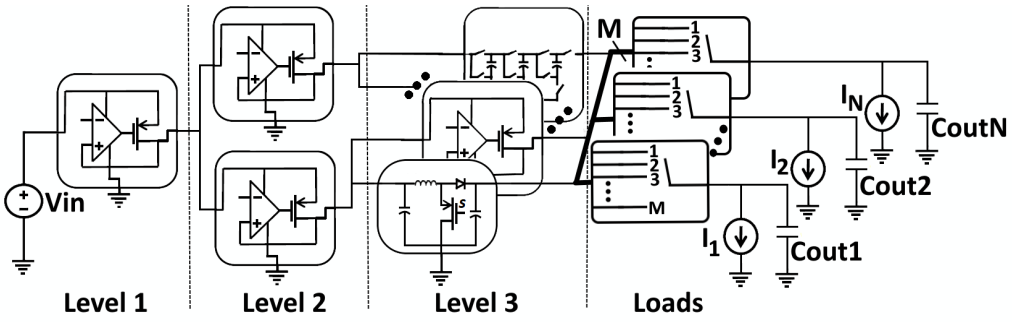



**Figure 4.** Three-level PDU with disparate DC–DC converters for heterogeneous loads.

In addition, every level of the PDU can be modeled separately, and hence, the PDU can be seen as a series of cascaded Voltage Regulator Modules (VRMs). For the cascaded VRMs illustrated in Figure 5, the total end-to-end power efficiency $\eta_{tot}$ is given by

$$\eta_{tot} = \frac{P_{out}}{P_{in}} = \frac{V_{out} I_{out}}{V_{in} I_{in}} = \prod_{i=1}^{M} \eta_i, \tag{4}$$

where $M$ is the number of power levels from input (e.g., the board) to output (i.e., the loads) and $\eta_i$ is the power efficiency of the $i$-th level of the PDU, as illustrated in Figure 5, which consists of parallel LDOs powering a single voltage domain. All LDOs share the same voltage ratio $\frac{V_{in}}{V_{out}}$ but differ in maximum current capability $I_{max}$ and quiescent current $I_{qsc}$. In addition, every level includes several converters, e.g., LDOs, which exhibit a different power efficiency for the same load, as shown in Figure 6.

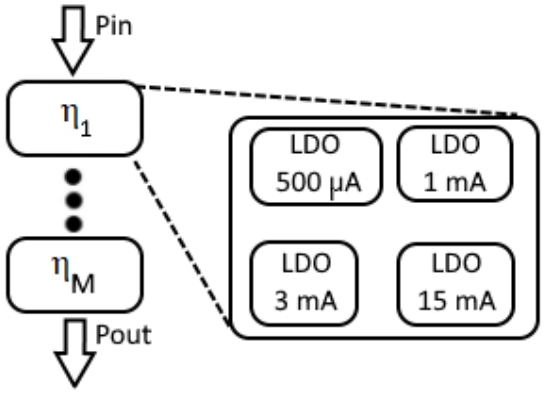



**Figure 5.** A chain of $M$ voltage levels, which consist of diverse VRMS, such as *LDOs*.

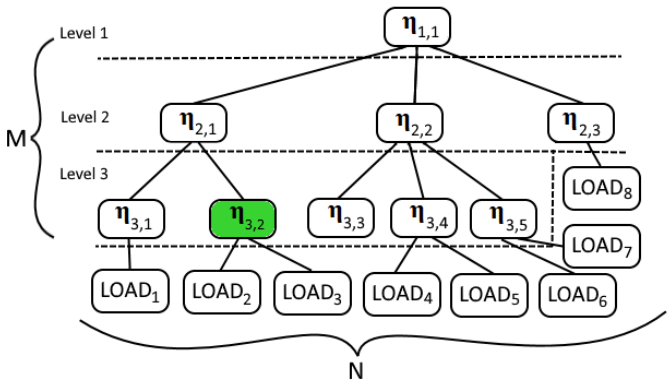



**Figure 6.** A three-level hierarchical PDU with eight *Loads*.

Let us consider two successive levels of a multi-level PDU, as plotted in Figure 7a. The overall power efficiency for this two-level sub-system is

$$\eta_{sub} = \frac{\sum_{j=1}^{K} V_{out(i,j)} I_{out(i,j)}}{V_{in(i-1)} I_{in(i-1)}}, \tag{5}$$

where $K$ is the number of converters in the $i$-th level. If we model the converter of the $(i-1)$-th level as $K$ parallel identical copies of itself while keeping the same overall consumption, the system obtains the form of Figure 7b. Therefore, the power efficiency can be written as

$$\begin{aligned} \eta_{sub} =& \frac{\eta_{i-1,1}\eta_{i,1} V_{in(i-1,1)} I_{in(i-1,1)} + \cdots}{V_{in(i-1,1)} I_{in(i-1,1)} + \cdots} \\ & \frac{\cdots + \eta_{i-1,K}\eta_{i,K} V_{in(i-1,K)} I_{in(i-1,K)}}{\cdots + V_{in(i-1,K)} I_{in(i-1,K)}} \\ =& \frac{\sum_{j=1}^{K} \eta_{i-1,j}\eta_{i,j} V_{in(i-1,j)} I_{in(i-1,j)}}{\sum_{j=1}^{K} V_{in(i-1,j)} I_{in(i-1,j)}}. \end{aligned} \tag{6}$$

As described in (6), the power efficiency $\eta_{sub}$ is a weighted mean of the power efficiency $\eta_{i-1,j} \cdot \eta_{i,j}$ of each $j$ path according to Figure 7b, where $j \in [1, K]$.

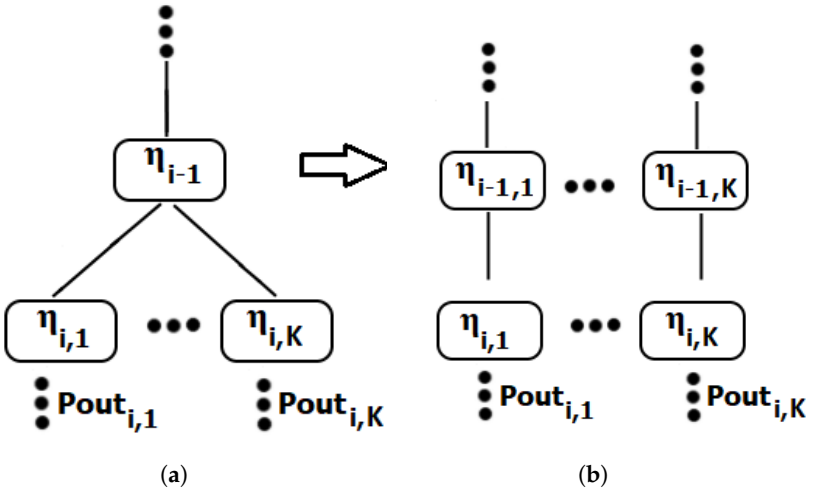



**Figure 7.** A two-level sub-system that consists of $K$ converters at the $i$-th level. (**a**) Sub-hierarchy instantiation. (**b**) Decomposition of the $(i-1)$-th level.

Moreover, each "parent" converter at the $i$-th level is decomposed based on the "children" converters or loads connected to the $(i+1)$-th level. As an example, the decomposition of converter $\eta_{3,2}$ at *Level 3* of the PDU shown in Figure 6 results in two parallel converters due to the loads $Load_2$ and $Load_3$ connected to $\eta_{3,2}$, as depicted in Figure 8. Therefore, a bottom-to-top decomposition of the power tree results in the top-most level consisting of at least $N$ converters in total because there are $N$ loads connected to the bottom-most level of the PDU.

In the same way, the decomposition of *Level 3* of the PDU shown in Figure 6 results in eight parallel converters because there are seven loads connected to *Level 3* and the unused converter $\eta_{3,3}$. Consequently, the decomposition of *Level 2* results in nine parallel converters since $\eta_{2,1}$ is modelled as three parallel converters, $\eta_{2,2}$ is modelled as five equivalent converters, and $\eta_{2,3,}$ is not decomposed. As a result, the converter at *Level 1* is modelled as nine parallel converters due to eight loads and the unused converter $\eta_{3,3}$, as depicted in Figure 8.

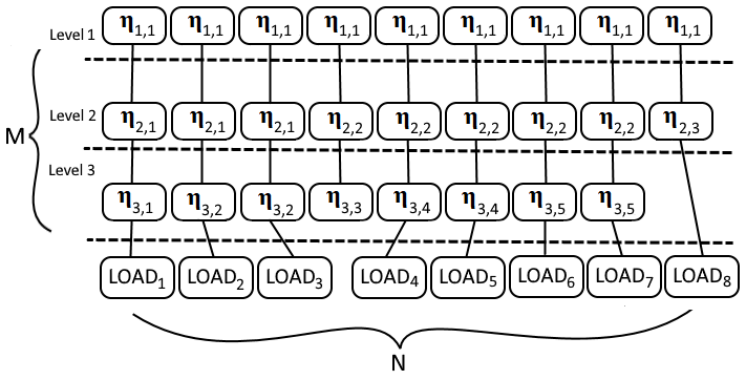



**Figure 8.** Equivalent model of a three-level hierarchical PDU with eight *Loads* for estimating $\eta_{tot}$.

However, only paths with connected loads contribute to the consumed power; hence, exactly $N$ paths, which end with the same number of loads, are used to estimate overall power efficiency. Therefore, expanding (6) for all levels $M$ of the PDU results in the end-to-end power efficiency of the system, given by

$$\eta_{tot} = \frac{\sum_{j=1}^{N} \prod_{i=1}^{M(j)} \eta_{i,j} V_{in(1,j)} I_{in(1,j)}}{\sum_{j=1}^{N} V_{in(1,j)} I_{in(1,j)}}, \tag{7}$$

where *M(j)* is the level of the hierarchy along the path to each leaf converter, as illustrated in Figure 6. Generally, the number of converters connected to the leaf load may be of a different "depth" from the root of the power delivery tree. In addition, every $\eta_{i,j}$ requires only the total current $I_{LOAD}$, which loads the converter, as the power efficiency $\eta_{i,j} = f_{LDO}(I_{LOAD}, I_{qsc_{i,j}})$ and the quiescent current $I_{qsc_{i,j}}$ of every converter are known. Moreover, the mean input power of all paths is given by

$$\overline{Pin} = \frac{\sum_{j=1}^{N} V_{in(1,j)} I_{in(1,j)}}{N}. \tag{8}$$

If the mean input power is used to estimate every $V_{in(1,j)} I_{in(1,j)}$ term, (7) can be written as

$$\begin{aligned} \eta_{tot} \approx \eta_{est} &= \frac{\sum_{j=1}^{N} \prod_{i=1}^{M(j)} \eta_{i,j} \overline{Pin}}{\sum_{j=1}^{N} V_{in(1,j)} I_{in(1,j)}} \\ &= \frac{\overline{Pin} \sum_{j=1}^{N} \prod_{i=1}^{M(j)} \eta_{i,j}}{N \overline{Pin}} \\ &= \frac{1}{N} \sum_{j=1}^{N} \prod_{i=1}^{M(j)} \eta_{i,j}. \end{aligned} \tag{9}$$

Therefore, the load currents alone suffice to determine the overall power efficiency, as when following an upstream traversal of the PDU tree from the leaves to the root, the load condition of every level $M(j)$ depends only on the loads of the lower level connected to level $M(j)$.

This behavior allows for the utilization of a search table to determine the most efficient converter depending on the output load. The power efficiency of each converter is measured once for the whole span of loads supported by the converter. These measurements are then stored in a table such that the most efficient converter for any connected load is always selected. The reward function of the proposed Q-algorithm uses this approach to determine the power efficiency $\eta_{tot}$ of the PDU, as discussed in Section 5.

### 3.4. Approximation Error Analysis

To quantify the accuracy of the efficiency estimation used in (9), the error term relative to the exact efficiency defined in (7) has to be derived. Let $\eta_{path,j} = \prod_{i=1}^{M(j)} \eta_{i,j}$ denote the cumulative efficiency of the $j$-th path, and let $P_{in,j} = V_{in(1,j)} I_{in(1,j)}$ denote the input power of that path. Consequently, the exact total efficiency $\eta_{tot}$ corresponds to the power-weighted mean of the path efficiencies:

$$\eta_{tot} = \frac{\sum_{j=1}^{N} P_{in,j} \cdot \eta_{path,j}}{\sum_{j=1}^{N} P_{in,j}} \tag{10}$$

The estimated efficiency $\eta_{est}$ utilized by the proposed algorithm is the unweighted arithmetic mean:

$$\eta_{est} = \frac{1}{N} \sum_{j=1}^{N} \eta_{path,j} = \overline{\eta_{path}} \tag{11}$$

By invoking the statistical definition of covariance, $\mathrm{Cov}(X, Y) = \frac{1}{N} \sum X_i Y_i - \bar{X}\bar{Y}$, we can express the summation of the product (the numerator of the exact efficiency) as

$$\frac{1}{N} \sum_{j=1}^{N} P_{in,j} \eta_{path,j} = \mathrm{Cov}(P_{in}, \eta_{path}) + \overline{P_{in}} \cdot \overline{\eta_{path}} \tag{12}$$

By substituting this identity back into the expression for $\eta_{tot}$, where the total input power is $N\overline{P_{in}}$, we have

$$\begin{aligned} \eta_{tot} &= \frac{N[\mathrm{Cov}(P_{in}, \eta_{path}) + \overline{P_{in}} \cdot \overline{\eta_{path}}]}{N\overline{P_{in}}} \\ &= \frac{\mathrm{Cov}(P_{in}, \eta_{path})}{\overline{P_{in}}} + \eta_{est} \end{aligned} \tag{13}$$

Consequently, the approximation error $\epsilon_{est}$ is formally derived as the covariance between the path input power and path efficiency, normalized by the mean input power:

$$\epsilon_{est} = \eta_{tot} - \eta_{est} = \frac{\mathrm{Cov}(P_{in}, \eta_{path})}{\overline{P_{in}}} \tag{14}$$

This derivation demonstrates that the estimation error approaches zero when the path efficiency is uncorrelated with the input power magnitude. In the context of the proposed PDU, where the Q-algorithm maintains high efficiency across all active paths regardless of load magnitude, the covariance term remains minimal, resulting in a negligible estimation error (<2%). For example, if one load consumes 10 mW and another consumes 1 mW (high power variance), but due to the Q-algorithm, their supply converters operate at $\approx 80\%$ efficiency, then this means that the efficiency is "uncorrelated" with power magnitude. Therefore, the weighted mean $\eta_{tot}$ and the arithmetic mean $\eta_{est}$ are almost identical.

### 3.5. Practical Advantages of the Method

The efficient update of the policy of a PDU requires the measurement or the estimation of the overall end-to-end power efficiency $\eta_{tot}$, as well as the power efficiency of each converter of the PDU, as discussed in Section 3.3. To measure the efficiency of every converter, multiple circuits for current and voltage measurements are required, yielding a higher form factor [3,7,9]. Moreover, these measurements are repeated every time $\eta_{tot}$ needs to be calculated, thereby increasing the power and the runtime of the method when

managing the system's energy [5]. The power lost due to these measurements is especially critical in portable systems due to the limited power budget [10,11].

The combination of the appropriate estimation of $\eta_{tot}$ and a search table with the measured efficiency over different loads for every converter of the PDU addresses these limitations. In this new approach, the most efficient converter is selected based on the already measured power efficiency for the connected load, as described in Section 3.3, reducing the hardware requirements of the method. Consequently, the power consumed to manage the system energy decreases, since the converter efficiency is not calculated every time. Therefore, the proposed method, based on the estimation of power efficiency described in Section 3.3, is particularly suitable for portable systems.

## 4. Power Delivery Method

By combining power gating and Q-learning, the new technique takes as an input a hierarchy of connected DC–DC converters and the load currents, i.e., $I_{Load}$, and returns the optimum policy with the maximum available end-to-end power efficiency $\eta_{tot}$. The losses due to the *RLC* characteristics of the interconnects between the converters are considered during the measurement and storage of the efficiency for each individual converter. In case loads are homogeneous, the Q-algorithm can be disabled. For both types of loads, power gating is utilized, where the PDU can enable, if required, only one of the available DC–DC converters, yielding the highest power efficiency under all conditions and across all available converters. The stability of the PDU, where power gating is utilized, is presented in Section 4.1. In addition, the DC–DC converters are dynamically assigned to heterogeneous loads by the Q-algorithm in order to increase total end-to-end power efficiency as much as possible according to (3). The tailored Q-algorithm, along with the reward function and clarification examples, is described in Section 4.2.

### *4.1. Stability of the PDU*

For homogeneous loads, one DC–DC converter across the PDU typically suffices to supply the loads [5]. Therefore, the proposed power delivery method has been designed so that there is always an LDO fully supplying the connected loads. This practice assumes that the loads are known. The interchange of the available LDOs is controlled by the PDU before high load transitions, since loads are already scheduled, e.g., data transmission via BLE or switching to a different power mode. Therefore, the proposed PDU scheme can be designed to be *a priori* stable for homogeneous loads. This point arises from the fact that only one converter is enabled each time to supply the loads, such that the resonance frequency is determined by the specific converter [45]. After selecting the LDO regulator that is the most efficient among the available LDO regulators, the PDU is modeled as a circuit with two poles, as depicted in Figure 9. A dominant pole is required for the error amplifier (EA), and a higher-order pole is required for the power transistor with the output capacitor $C_{out}$ [57].

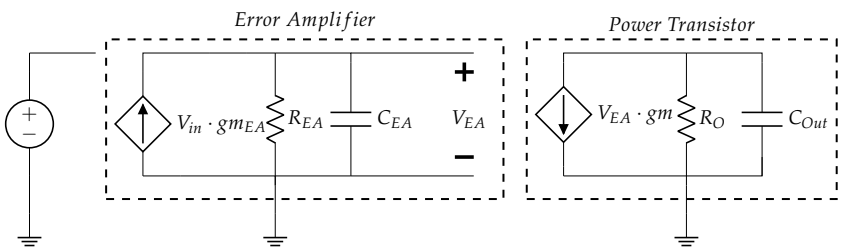



**Figure 9.** Two-pole model for the LDO [58].

The open-loop transfer function $G$ of the enabled LDO, which is shown in Figure 3b, is given by

$$\begin{aligned} G(s) &= \frac{V_{OUT}}{V_{IN}}(s) \\ &= \frac{A_{OL}}{(1+sR_{EA}C_{EA})(1+sR_oC_{out})} \\ &= \frac{g_m g_{m_{EA}} R_{EA} R_o}{(1+sR_{EA}C_{EA})(1+sR_oC_{out})}, \end{aligned} \tag{15}$$

$$A_{OL} = g_m R_o A_{EA}, \tag{16}$$

$$A_{EA} = g_{m_{EA}} R_{EA}, \tag{17}$$

where $A_{OL}$ is the open-loop gain of the LDO, $A_{EA}$ is the open-loop gain of the EA, $R_o$ is the output impedance of the LDO, and $R_{EA}$ is the output impedance of the EA. The resonance frequency $f_o$ and the Q factor of the LDO are, respectively, given by

$$f_o = \frac{1}{2\pi\sqrt{R_{EA}R_oC_{out}C_{EA}}}, \tag{18}$$

$$Q = \frac{\sqrt{R_{EA}R_oC_{out}C_{EA}}}{(R_{EA}C_{EA}+R_oC_{out})}. \tag{19}$$

In addition, stability is ensured only if $Q < \frac{1}{2}$, which leads to

$$\begin{aligned} \frac{\sqrt{R_{EA}R_oC_{out}C_{EA}}}{(R_{EA}C_{EA}+R_oC_{out})} < \frac{1}{2} \Rightarrow \\ (R_{EA}C_{EA} - R_oC_{out})^2 > 0 \Rightarrow \\ R_{EA}C_{EA} \neq R_oC_{out}. \end{aligned} \tag{20}$$

Therefore, $R_{EA}$, $R_o$, $C_{out}$, and $C_{EA}$ can be chosen appropriately to provide a stable LDO. Consequently, the overall PDU is stable for every state of the available LDO regulators since each LDO is designed to separately satisfy (20). Simulation results for a step load from 112 µA to 50 mA are depicted in Figure 10. Initially, the LDO supplying 1 mA as a maximum is enabled until a 50 mA load is requested. For greater loads, the LDO that supplies a maximum of 200 mA is enabled. With a settling time of less than 20 µs and an undershoot of 195 mV, the overall system is robust and supports diverse loads with low-power modes that have wake-up times of a few tens of µs [48,59,60].

On the other hand, guaranteeing stability in the case of heterogeneous loads is not straightforward because many of the LDOs, which are shown in Figure 4, should operate simultaneously, thereby jeopardising overall stability [58]. The coexistence of multiple LDOs, especially for light loads (<1 mA), leads to an oscillatory response due to the interaction with the off-chip parasitic resistances and the input impedance among LDOs [45]. Specifically, as the number of LDOs increases, the equivalent resistance decreases due to the parallel connection of the parasitic resistances and the input impedance of the LDOs. The resonance frequency also shifts, and consequently, phase shifts as well, thereby yielding an unstable PDU. The number of LDOs, below which the stability of the system is guaranteed, is not fixed because this condition strongly depends upon the following:

- The characteristics of each LDO comprising the PDU;
- The load conditions;
- The output capacitance *Cout* of each LDO.

While a formal control-theoretic derivation of multi-loop stability margins is a distinct research topic [45,58], the practical stability of the experimental setup is ensured through two mechanisms. First, passive ballast resistances (≈50 mΩ) are placed in series, with each

regulator mitigating circulating currents. Moreover, the proposed Q-algorithm is designed to activate parallel configurations only under heavy-load conditions. Consequently, the heavy-load condition required for parallel operation naturally coincides with the region of highest loop stability [46,47]. In addition, the output pole of an LDO shifts to higher frequencies as load current increases $I_{Load}$, which inherently improves the phase margin [61].

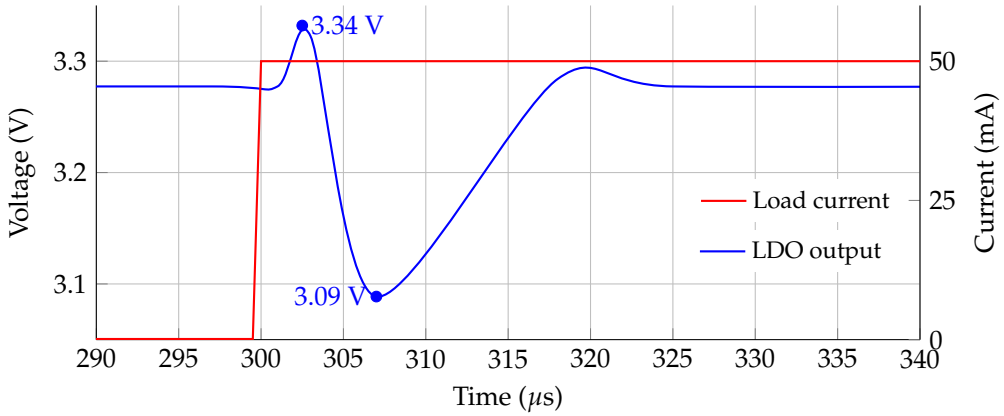



**Figure 10.** Simulation results of step response for load transitioning from 112 μA to 50 mA.

Moreover, a practical limit of 15 parallel LDOs ensures stability against oscillatory responses caused by parasitic interactions [45]. However, the proposed Q-algorithm framework can scale beyond this limit through hierarchical sub-clustering. In such a scenario, the PDU architecture can be organized into nested voltage domains, where each sub-cluster of regulators is isolated by an intermediate power stage. This structural decomposition allows the Q-algorithm to continue optimizing global end-to-end efficiency by treating these sub-clusters as modular branches within the power tree $T$.

### *4.2. Q-Algorithm-Based Power Delivery*

In the case of a large number of converters, which is often required, as analyzed in Section 3.2, the existing methods cannot optimally match these converters with the loads without sacrificing optimality [4,5] or without large overheads [18,25]. In order to address both limitations, Q-learning is utilized.

A variant of the fundamental Q-algorithm [41] properly tailored to power efficiency metrics is introduced to optimize the PDU for heterogeneous loads. The PDU is hierarchical, comprising $M$ levels of DC–DC converters and $N$ different loads, as illustrated in Figure 4. Since Q-learning operates on graphs [6], the PDU is modeled through the tree $T(V, E)$, as drawn in Figure 6, where each node $V$ is a DC–DC converter or the load of the system, and each unweighted edge $E$ implies a connection between two converters at different levels of the PDU or between a converter and a load. The power supply of the PDU is connected to the converters at level 1 of the PDU; thus, the converters at level 1 are referred to as *input nodes*.

A state $s_t$ is defined as an instantiation of the system after the assignment of one load to a converter. The number of states $S$ for the PDU is equal to the number of converters, incremented by one for each load used in every epoch. In addition, the actions $A$ correspond to all possible transitions between states. Therefore, the exploration "plane" for the Q-algorithm is a discrete 2-D $S \times S$ space where each point (x,y) represents the action of transition from state $x$ to state $y$. This selection covers every possible combination of load-to-regulator assignment compared to the limited number of eight states in [22]. The exploration of the $S \times S$ space, which results in higher end-to-end power efficiency $\eta_{tot}$ compared to the scheme with eight states, is possible in real time, as Q-learning algorithms are suitable for the fast exploration of large spaces [43,62]. Furthermore, the learning rate

$\alpha$ is set to be 0.8, favoring prior over new knowledge [27] because the current knowledge may correspond to a sub-optimum state, leading to slow learning or even a false agent.

Moreover, the discount factor $\gamma$ is set at 0.5 to achieve a high long-term reward whilst keeping the number of iterations sufficiently low [41] to remedy the overhead in power, which is important for portable systems. In addition, the matrix Q, where *Q(x,y)* represents the weight of the path between nodes $x$ and $y$, is initialized to zero, except for the primary diagonal, which is set to a large negative value; here, $-10$. Indeed, a path from a node to itself is meaningless, representing an impractical action that could lead to an infinite loop. Furthermore, the reward function $r_t$ is defined as follows:

$$r_t = \eta_{Nextnode} = f_{LDO}(\sum_{i=1}^{N} I_{Load_i}, I_{qsc_{Nextnode}}) \quad (21)$$

where $\eta_{Nextnode}$ is the power efficiency of the next node (i.e., converter) for the overall loads connected ($\sum_{i=1}^{N} I_{Load,i}$) and the quiescent current of the next node ($I_{qsc}$), as shown in Equation (2). The characterization of each converter allows for the calculation of $r_t$ only by measuring the total load current $\sum_{i=1}^{N} I_{Load_i}$, as described in Section 3.5. Furthermore, the total power efficiency $\eta_{tot}$ is calculated by (9). Note that while the algorithm does not explicitly query the quiescent current $I_{qsc}$ of the regulators, this current is intrinsically considered within the reward function. Since the reward is defined as the power efficiency $\eta_{tot}$ and includes the quiescent power consumption, any degradation in performance due to high $I_{qsc}$ at light loads directly reduces the reward. Consequently, the Q-agent naturally learns to avoid selecting high-loss regulators for low-current loads, optimizing the trade-off without requiring any modeling of the converters.

The pseudo-code of the Q-algorithm is listed in Algorithm 1 and proceeds as follows. In order to explore the entire $S \times S$ space, each load in tree $T$ is initially assigned to every last-level DC–DC converter. The set $D(j)$ is used to store the decision path of the $j$-th load, enabling the detection and avoidance of cyclic transitions and invalid loops during exploration. There are some assignments that are not feasible due to higher load demands than converter supplies (e.g., a load current of 25 mA from a 500 µA converter), where these assignments are pruned during the execution of the Q-algorithm. An example of this preliminary assignment is given below. For each load $Load_i$ connected to the system, the algorithm starts from each *input node*, connected to the power supply, and selects the path that maximizes the end-to-end power efficiency according to (1).

As long as the current state *node* is not the leaf $Load_i$, the most "appropriate" downstream nodes are found, as described in line 13. These nodes exhibit the maximum *Q(s,a)*, where $s$ is the present state *node*, and $a$ is the action to be taken, which leads to the state *Nextnode*. If there are multiple candidate actions to be taken, the action for state *Nextnode* is chosen randomly [62,63]; otherwise, the action with the maximum *Q(s,a)* is taken (lines 17–23). If the maximum *Q(s,a)* is negative, then all the paths from the current *node* to neighbour *Nextnodes* are penalized according to (1) in order to be excluded from future iterations (line 32). The *Nextnode* is checked for whether it can support $Load_i$ in terms of both voltage supply level and maximum output current demand. If the *Nextnode* cannot support $Load_i$, then the path between *node* and *Nextnode* is excluded from future iterations (line 43) in order to avoid infinite loops of the Q-algorithm. Otherwise, given that the *Nextnode* can supply $Load_i$, *Nextnode* is checked against $Load_i$. If *Nextnode* is $Load_i$, then the reward is set to the ideal power efficiency, equal to 1 (line 35), because this is the final state *Goal*, else the reward is set to the power efficiency between *node* and *Nextnode*. The power efficiency between *node* and *Nextnode* is determined based on the $I_{tot}$ loading of *node*, where $I_{tot}$ is the sum of all currents provided by *node* incremented by $Load_i$.

**Algorithm 1:** Q-algorithm for optimum PM policy

```
Result: Tree T of PDU
T % tree of PDU with loads;
Initial assignment of each load to all the "bottom" converters;
% Q-algorithm learning rate settings;
alpha = 0.8, gamma = 0.5;
for j ← 1 step 1 to num_of_Loads do
    Goal ← T[j];
    D(j) ← ∅ ;                                   // D(j): Decision path for load j
    for i ← 1 step 1 to max_epochs do
        node ← find_root(Q) ;
        add node to D(j);
        while node ≠ Goal do
            prvNode=node;
            find potential next nodes with max(Q);
            find unvisited nodes;
            if there are unvisited nodes then
                if more than one unvisited then
                    Nextnode is chosen randomly;
                else
                    Nextnode is the last unvisited;
                end
            else
                if more than one with max(Q) then
                    Nextnode is chosen randomly;
                else
                    Nextnode with max(Q);
                end
            end
            if Nextnode supports Load then
                node ← Nextnode;
                if max(Q) < 0 then
                    reward ← -1;
                    Loop penalty  prvNode ← node;
                else
                    if node = Goal then
                        reward ← 1;
                    else
                        reward ← η_Nextnode ;
                    end
                end
                Update(Q,reward,node,prvNode);
                add node to D(j);
            else
                Q(node,Nextnode)=-10;
                Q(Nextnode,node)=-10;
                Invalid path
            end
        end
    end
end
Function Bellman Update(Q, reward, node, prvNode):
    Q(node, prvNode) = Q(node, prvNode) + alpha*(reward+gamma*max(Q(prvNode, node)) -
      Q(node, prvNode));
    return Q
End Function
```

To practically demonstrate this process, let us consider a case where *node* has an input voltage of 1.8 V, an input current of 2.8 mA, an output voltage of 1.2 V, and a maximum output current of 4 mA. Let this converter node be loaded (from previous iterations) with

3 mA, and let $Load_i$ be 0.5 mA. The efficiency is $\frac{(1.2)\times(3+0.5)}{(1.8)\times(2.8)} = 83.3\%$, which is also the reward for updating matrix Q. For each $Load_i$, the algorithm iterates for a user-defined number of epochs such that $Q$ converges to an optimal path from the power source to $Load_i$ in terms of total end-to-end power efficiency, as described in Algorithm 1.

## 5. Experimental Results

In this section, experimental results for six real-world scenarios are presented. The benchmark scenarios are depicted in Table 2. Homogeneous as well as heterogeneous loads are explored, as the power drawn by individual loads varies considerably between operating states. The learning process was repeated for 10 epochs. In the implemented Q-algorithm, an epoch also denotes the re-initialization of the exploration probability while preserving the Q-table states. This "repeated annealing" strategy prevents the agent from stalling in local optima during early learning. Empirical tests indicate that policy stability is consistently achieved by the fourth epoch. Therefore, 10 epochs were chosen to guarantee robust convergence with a safety margin.

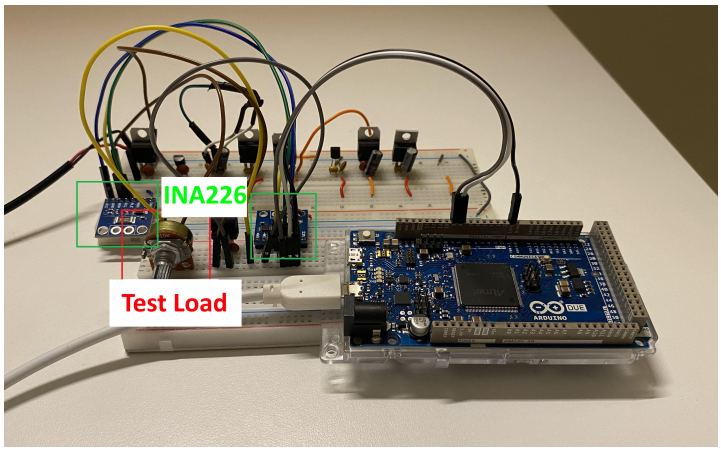
Comparative analysis was conducted for a three-level PDU before (Static) and after (Q-algorithm) the implementation of the proposed method, based on the system configuration shown in Figure 11a,b for homogeneous and heterogeneous loads, respectively. The circuit diagram of the heterogeneous setup depicted in Figure 11b is illustrated in Figure 12. The variables of the Q-algorithm were directly mapped to hardware control signals as follows:

- **State ($s_t$):** Physically represents the total power path of tree $T$ from *Battery* to *Loads* based on the enabled regulators and the selection bits of analog multiplexers (AMUX).
- **Action ($a_t$):** Corresponds to the digital control signals *Select Supply* and *Enable Regulator* sent by the Q-agent (e.g., Arduino Due) to the PMOS transistors and AMUXs. These signals physically reconfigure the circuit topology by enabling/disabling a specific regulator or connecting/disconnecting a *Load* to/from a regulator.
- **Reward ($r_t$):** This is defined as the real-time efficiency $\eta_{Nextnode}$ calculated by Equation (21) using data gathered by current sensors (e.g., INA226 [64], as depicted in Figure 12).

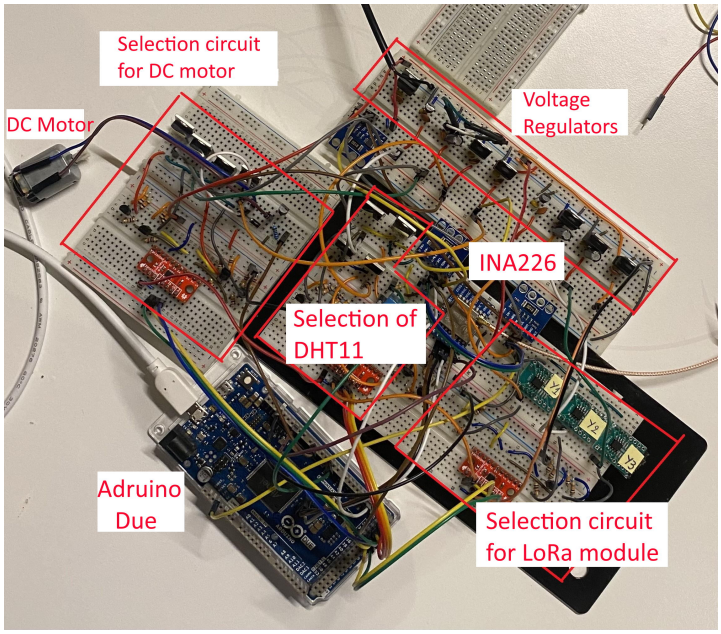
For homogeneous loads, the PDU is considered to utilize either four or three 5 V regulators, respectively. For the configuration with four regulators, L78L05, LM2937, L7805ABV, and L78S05CV are used for the 5 V rail [65–68]. However, load fragmentation into four sub-regions increases computational complexity and power consumption while adding no significant efficiency gain compared to three sub-regions, as shown by the measurements. Therefore, the configuration with only three regulators (L78L05, L7805ABV, and L78S05CV) is preferred for the 5 V rail. The improvement in power efficiency is described for each case in Sections 5.1 and 5.2, respectively.

The performance of the proposed technique in terms of overall end-to-end power efficiency was compared with the power efficiency of existing PDUs. The comparison results are listed in Table 3.

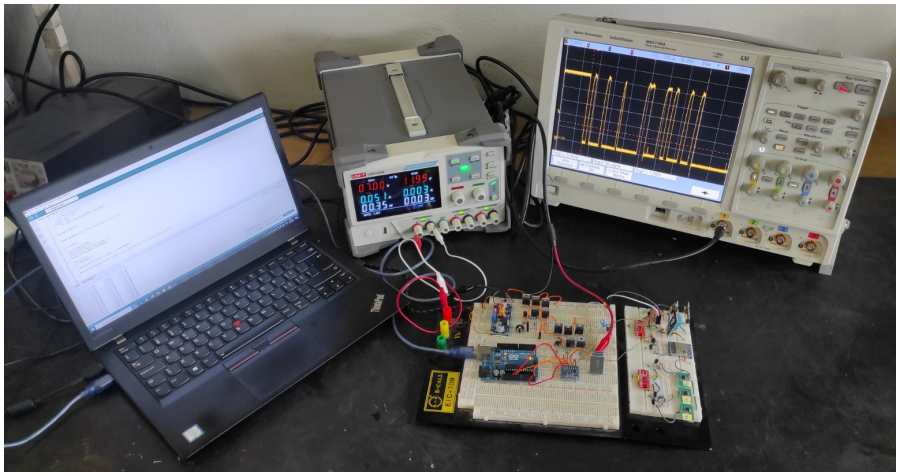
TPS659037 was included solely as a reference point for state-of-the-art commercial efficiency, not as a direct architectural comparison. While TPS659037 utilizes a different, highly integrated architecture (Buck+LDO), it represents the target performance standard for modern power management systems. Therefore, the table should not be interpreted quantitatively, but rather qualitatively, in terms of performance. The efficiency $\eta_{tot}$ of the proposed system is comparable with the datasheet values of TPS659037. The results indicate that the proposed Q-learning framework enables the validation setup to achieve comparable efficiency levels, effectively bridging the gap between flexible software-defined power management and dedicated hardware solutions.




(a)




(b)



(c)

**Figure 11.** Overview of evaluation scenarios and measurement setup. (**a**) Homogeneous setup. (**b**) Heterogeneous setup. (**c**) Measurement setup used for scenario evaluation.

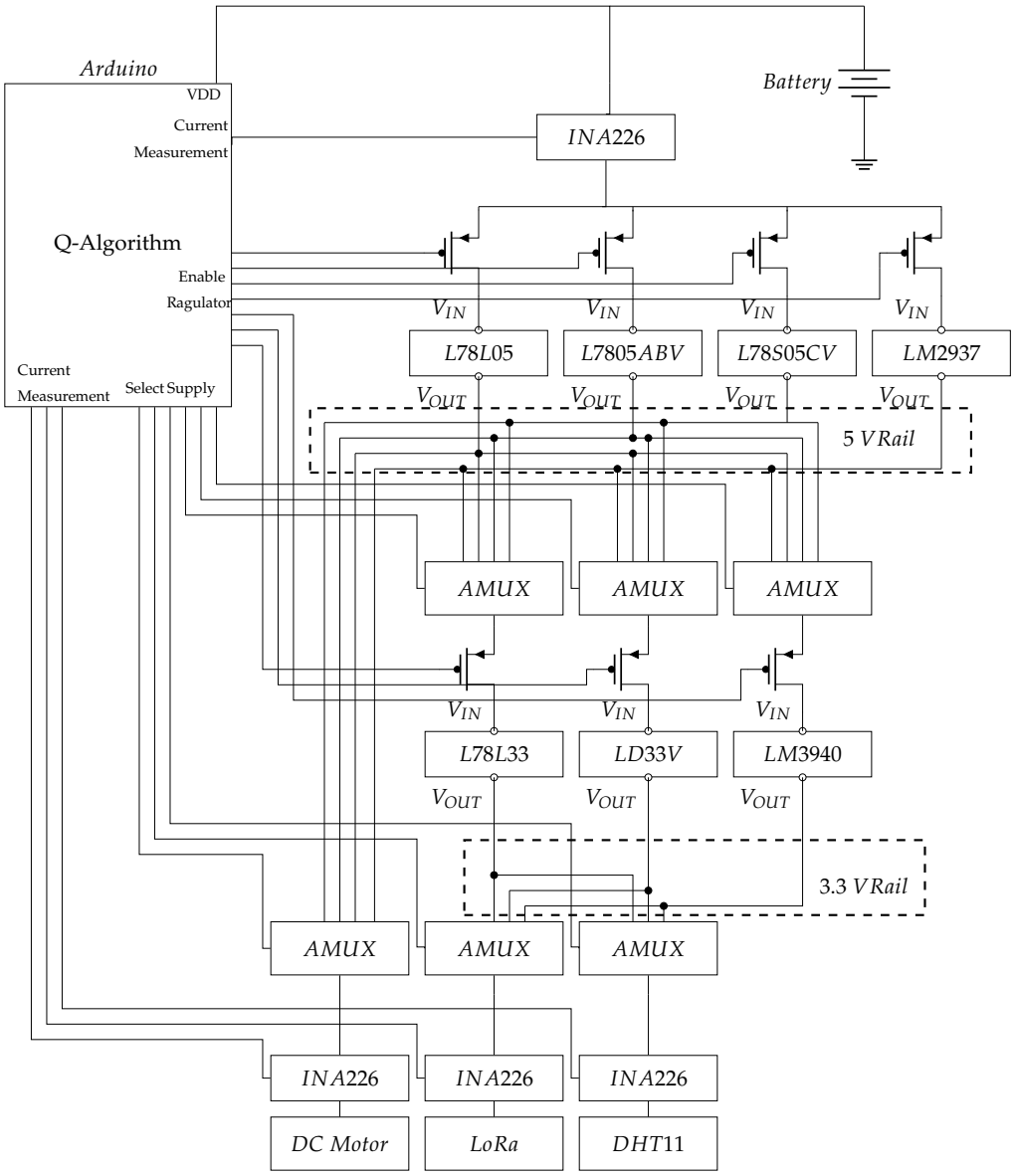



**Figure 12.** Schematic of the heterogeneous setup depicted in Figure 11b.

**Table 2.** IoT scenarios, active loads, and min/max currents (mA) per load. A tick (✓) indicates the load is used in that scenario.

| Scenario | Application | DHT11 | PIR | HC-SR04 | DC Motor | SG90 Servo | HM10 BLE | ESP8266 WiFi | LoRa SX1278 |
|---|---|---|---|---|---|---|---|---|---|
| | Min (mA) | 0.1 | 0.105 | 2.5 | 10 | 15 | 0.004 | 0.01 | 0.001 |
| | Max (mA) | 0.5 | 65 | 15 | 256 | 360 | 106 | 170 | 87 |
| A | Automotive: Parking assistance and pedestrian detection | ✓ | ✓ | ✓ | | | ✓ | | |
| B | Smart farming: Environmental monitoring and auto-watering | ✓ | | | ✓ | | | | ✓ |
| C | Home automation: Person detection, lights | ✓ | ✓ | | | ✓ | | ✓ | |
| D | PV sun tracking: WiFi/LoRa for efficiency | | | | | ✓ | | ✓ | ✓ |
| E | Autonomous vehicle: Self-exploration/RC control | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| F | Window shutter control over Wi-Fi for temperature | ✓ | | | ✓ | | | ✓ | |

**Table 3.** Performance comparison.

| | **2019 [3]** | **2016 [5]** | **TPS659037 [48]** | **This Work** |
|---|---|---|---|---|
| Input Voltage | N/A | 250 mV–1.1 V | 1.75 V–5.25 V | 5.5 V |
| Output Voltage | 4 V | 1.8 V–2 V | 0.9 V–3.3 V | 1.2–5 V |
| Maximum Load Current | 2.5 mA | >1 mA | 1 A | >600 mA |
| Heterogeneous Loads | No | Yes | No | Yes |
| Power Efficiency * @ 66 µA (%) | 77% | 26% | 23% | 83% |
| Maximum Power Efficiency | 95% | 57% | >90% | >80% |
| Autonomous | No | Yes | No | Yes |

* Overall power efficiency estimated by $\eta_{tot}$(%).

### *5.1. Homogeneous Loads Case Study*

For the deployment of the proposed methodology, LDOs were used. Nevertheless, the methodology is equally applicable to other types or a mixture of DC–DC converters. Note that the specific LDO regulators discussed in this work serve as a representative case study for validating the proposed management methodology. The Q-learning algorithm itself is model-free and converter-agnostic to the underlying converter topology. It is equally applicable to systems comprising Buck, Boost, or hybrid converters. Furthermore, the algorithm inherently considers hardware-specific characteristics, such as quiescent current $I_{qsc}$, by treating the power converter as a "black box" where the optimization objective is end-to-end efficiency. The block diagram of the system used for the experiments is illustrated in Figure 1. In this framework, the Q-algorithm for power management is deployed onto an Arduino. The multi-level PDU is realized as a three-level hierarchy PDU with six cascaded LDOs, three for 5 V regulation (L78L05, L7805ABV, and L78S05CV [65,67,68]) and three for 3.3 V regulation (L78L33, LD33V, and LM3940 [65,69,70]). Eight loads with diverse current consumption and voltage supply were used, as depicted in Table 2. Therefore, the exploration space consisted of 196 different States. At each decision step, the Q-agent executes an Action by physically enabling or disabling a specific LDO via the digital control bus to move to another State, thereby altering the total end-to-end efficiency $\eta_{tot}$ (Reward) of the system. By maximizing this Reward, the agent implicitly learns to select high-power regulators (e.g., L78S05CV) during high-load scenarios to avoid dropout and low-quiescent regulators (e.g., L78L05) during sleep scenarios to minimize idle losses. Following an action, the agent observes the Reward by calculating the real-time system efficiency $\eta_{tot}$ using data acquired from the input/output current and voltage sensors. This closed-loop structure allows the algorithm to directly map theoretical efficiency maximization to concrete hardware switching decisions.

The typical topology of an LDO described in [71] was utilized to implement various LDOs so as to provide currents of up to 1 A while delivering power with efficiency over 90% per LDO [72–74]. A set of four available LDOs is defined for a single load, as depicted in Figure 3a. Although all of the LDOs are connected to the load, the PDU enables only one LDO each time to utilize power gating [8], thereby yielding the maximum end-to-end power efficiency, as also illustrated in Figure 3b.

The proposed hierarchical Q-management system differs from conventional hardware-level efficiency techniques, such as interleaved DC–DC stages. While interleaving is highly effective for ripple reduction in high-power systems (>10 A), its application in ultra-low-power, battery-operated IoT nodes is limited by significant switching loss overhead at light loads. In contrast, this work focuses on the algorithmic coordination of a multi-level PDU. As demonstrated in Figure 13, the proposed method maintains an end-to-end efficiency over 80% for heterogeneous loads, even for scenarios involving ultra-low-power states of as

low as 1 µA. This result demonstrates a range of operations where traditional high-power topologies, such as interleaved stages, are not applicable.

Note that the efficiency curve presented in Figure 13 exhibits three distinct "discontinuities" due to the discrete switching events triggered by the Q-learning policy for the utilized four regulators (L78L05, LM2937, L7805ABV, and L78S05CV). As the load demand increases, the agent transitions the power path from L78L05 to LM2937, then to L7805ABV, and finally to L78S05CV. These three transitions are depicted as "discontinuities" in the efficiency plot, confirming that the Q-algorithm with power gating dynamically selects the optimal regulator for the given load current. However, the effective optimal load span for LM2937 (blue segment in Figure 13) is significantly smaller compared to other regulators. The load fragmentation into four sub-regions improves efficiency by almost 7%, but only across a limited range of loads between 0.1 A and 0.2 A, as shown in Figure 13. Conversely, incorporating a fourth regulator expands the Q-learning exploration space to 225 States. This increase in state-space complexity imposes a computational and power penalty that outweighs the efficiency benefit. Therefore, the final multi-level PDU effectively utilizes only three regulators (L78L05, L7805ABV, and L78S05CV) for the 5 V rail.

Therefore, the system displays up to a 60% increase in power efficiency, as reported in Table 3 for similar small loads (tens of µA), e.g., 66 µA, compared to a PDU utilizing a single LDO [48] with the same span of supported loads, as depicted in Figure 13. In addition, according to the analysis in Section 3.2, a higher power efficiency is demonstrated when compared to an autonomous PDU with coarser granularity for the DC–DC converters [5].

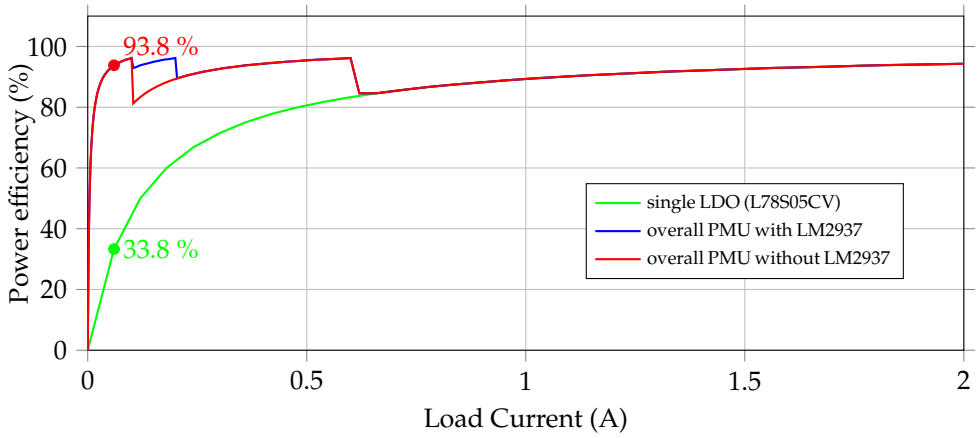



**Figure 13.** Overall power efficiency for homogeneous loads.

More importantly, the power efficiency is greatly increased for small loads, a few hundred µA, which is typical for embedded systems and systems that frequently enter power-saving modes, e.g., sleep mode [59,60]. Moreover, in order to keep the computational cost as low as possible, a search table for the most efficient converter may be used, depending on the output load, as efficiency does not need to be calculated every time a load is reassigned to a different converter. This search table reduces the runtime of the Q-algorithm and enhances the performance of the method, especially in the case of heterogeneous loads, as discussed in the next Section 5.2.

### *5.2. Three-Level Hierarchy PDU*

To demonstrate the usefulness of the Q-algorithm, a three-level hierarchy PDU with eight realistic loads was considered. The number of regulators upstream of the PDU tree decreases proportionally to the maximum current provided by these regulators. Therefore, at the top-most level (level 1) at the root of the tree, there is one battery that can provide up

to 4 A, while the lowest level (level 3) consists of several smaller regulators, as illustrated in Figure 14a.

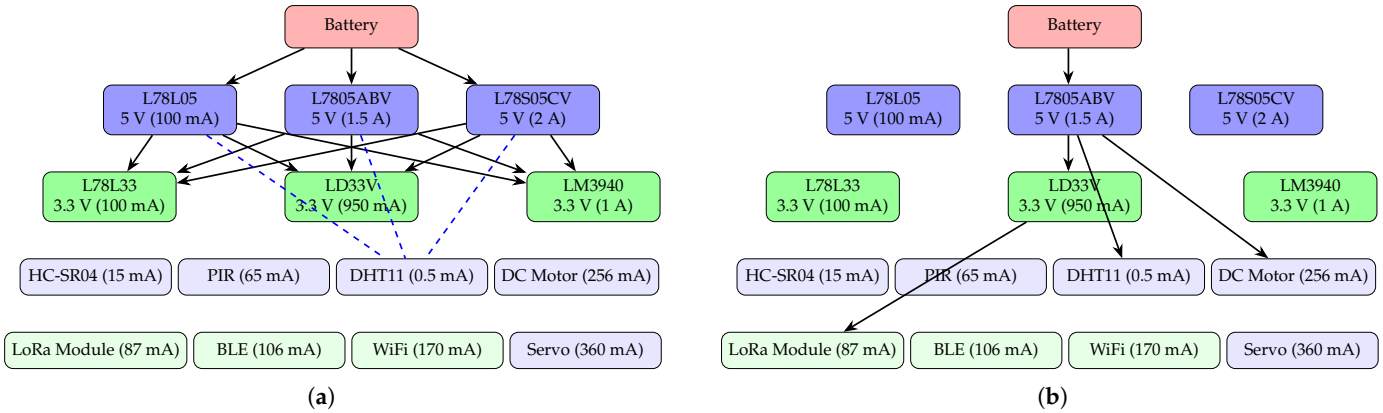



**Figure 14.** Three-level PDU supplying heterogeneous loads. (**a**) Initially, no load is assigned to any regulator. (**b**) After the Q-agent is trained, each load is assigned to a regulator to maximize overall efficiency.

Moreover, the loads correspond to current consumption at different operating modes of portable systems, e.g., 15 mA is needed in the modem-sleep mode of a WiFi MCU [59], and 5 mA is the current consumed by the radio transceiver for 1 Msps BLE of a multiprotocol SoC [60]. The number of iterations for the Q-algorithm is crucial to determine the optimum policy, since there are intermediate states that are far from optimal. For instance, the second epoch with five iterations yields an overall power efficiency of 75.9%, as shown in Figure 15. On average, seven iterations are typically needed for the proposed Q-algorithm to determine the optimum policy. Non-monotonic total efficiency indicates the exploration–exploitation trade-off in the middle epochs until the system converges to the optimal policy.

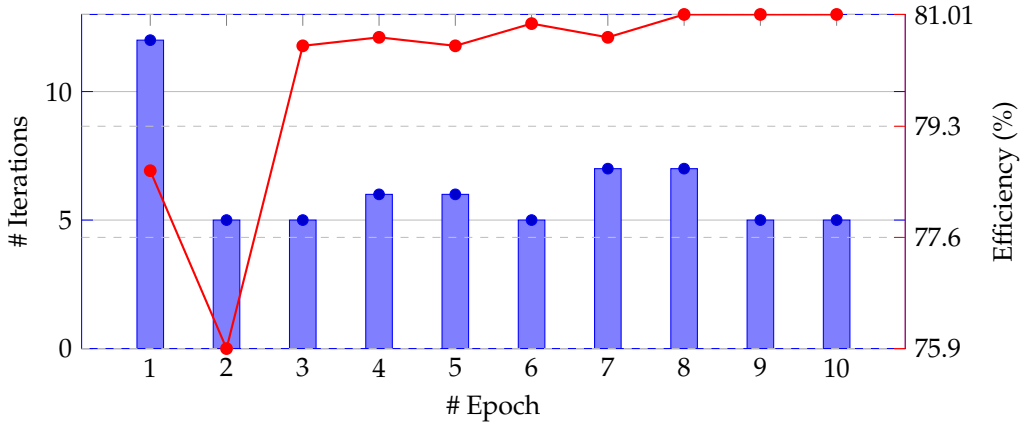



**Figure 15.** Scenario B (MAX currents)—Efficiency per epoch (red line) and iterations per epoch (blue bars). Total runtime: 48 ms.

The factor $\gamma$ in Algorithm 1 governs the trade-off between immediate and future rewards by scaling the bootstrap term in the Bellman update. Small $\gamma$ values induce myopic behavior and, typically, accelerate convergence due to a shortened effective planning horizon. Conversely, $\gamma \to 1$ promotes long-horizon reasoning and is required when optimality depends on delayed rewards, but it may slow learning and increase variance by amplifying bootstrapped estimates, particularly under function approximation. Based on a sensitivity analysis performed over a discrete grid of candidate values, $\gamma = 0.5$ and $\alpha = 0.8$ were finally selected. The discount factor $\gamma = 0.5$ provides an effective balance

between immediate and future rewards, which was observed to produce faster and more stable convergence than higher values (e.g., $\gamma \geq 0.8$) that introduced oscillatory behavior, and lower values (e.g., $\gamma \leq 0.2$) that yielded myopic policies. Similarly, $\alpha = 0.8$ was found to accelerate learning while maintaining numerical stability. Lower learning rates (e.g., $\alpha \leq 0.4$) decelerate convergence, whereas values close to 1 occasionally produce divergence. The chosen parameters thus represent the best trade-off between immediate and future efficiency with respect to convergence rate, stability, and final optimum policy.

The runtime of Algorithm 1 is highly affected when using the method that is used to obtain the load currents for calculating $\eta_{tot}$. A search table, with pre-measured currents, achieves a consistently low runtime of below 60 ms across all benchmark scenarios, as depicted in Figure 16. In contrast, the sensor-polling configuration incurs substantially higher computational cost due to the overhead associated with continuous current measurement. This effect is particularly pronounced in Scenario E, where the number of active loads is highest, resulting in up to a threefold ($\times 3$) increase in runtime, reaching 208 ms.

Power gating, discussed in Section 3.2, is also extended to heterogeneous loads to further enhance system power efficiency. Each DC–DC converter was implemented using multiple parallel LDOs, where only the most efficient LDO capable of supplying the required load current was enabled. Activating a single LDO allows converter trimming [75,76] and ensures compliance with the system constraint, limiting the number of active converters to 15. This configuration also preserves overall PDU stability, as discussed in Section 4.1.

Note that prior works developed specifically for heterogeneous loads [5] do not perform well for the scenarios described in Table 2, as these methods consider only the quiescent current when updating the PDU policy. As a result, existing PDUs exhibit $\eta_{tot}$ less than 26% when the loads are in *low power* mode, as depicted in Table 3. In contrast, the proposed Q-algorithm intelligently incorporates the power efficiency of each LDO in the PDU, as defined in (9), yielding a maximum $\eta_{tot}$ of over 80%.

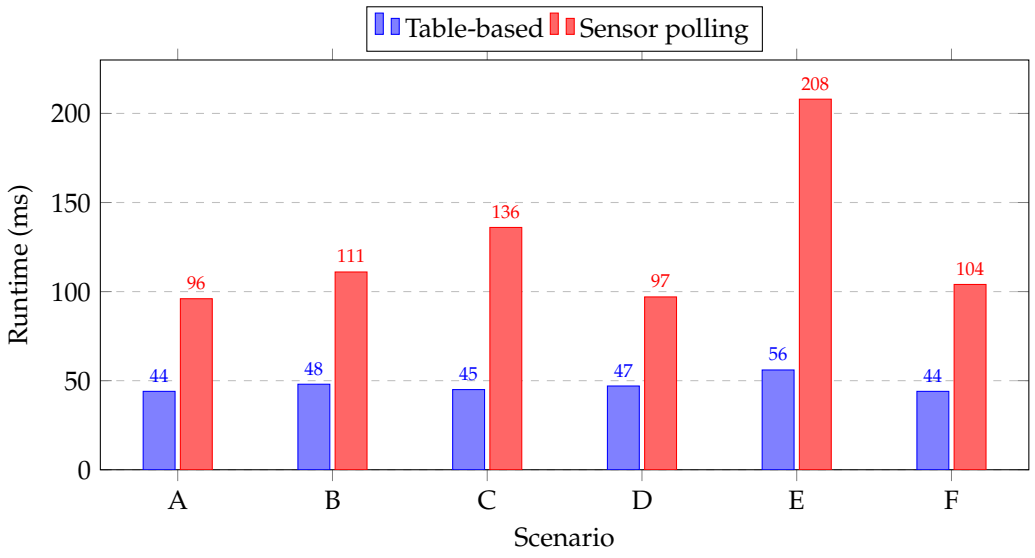



**Figure 16.** Runtime of the proposed Q-algorithm.

A comparative analysis of the efficiency achieved for a static, three-level PDU and a three-level PDU driven by the proposed Q-algorithm across the six application scenarios is depicted in Figure 17. The Q-algorithm consistently attains superior efficiency performance, outperforming the static PDU in all evaluated cases. The most notable improvement is observed in *Scenario A*, where the Q-algorithm yields a 12.67% increase in total efficiency compared to the static PDU. Comparable gains are evident in all cases, demonstrating an average increase of 5%. Even in scenarios where baseline performance is relatively

high, such as *Scenario C* and *Scenario E*, the proposed approach maintains a measurable efficiency advantage due to power gating.

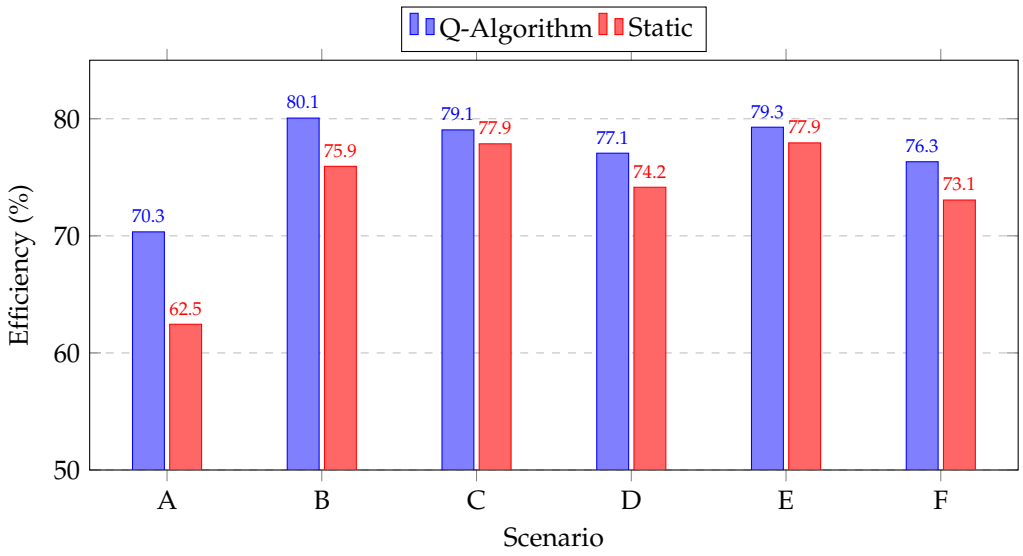



**Figure 17.** Efficiency of Q-algorithm PDU vs. Static PDU.

These results substantiate the robustness and superior efficacy of the proposed method across diverse application conditions, thereby validating its potential to manage hierarchical PMUs with heterogeneous loads encountered in present and upcoming portable systems.

## 6. Conclusions

The fragmentation of large DC–DC converters is not sufficient to achieve high power efficiency for heterogeneous loads. Rather, a hierarchical PDU that supports the dynamic utilization of different DC–DC converters improves the total end-to-end power efficiency $\eta_{tot}$ over a wide range of loads. However, for a hierarchical or, more generally, a PDU with many converters, using typical criteria for matching converters to loads does not suffice. A hierarchical PDU, augmented by a low-overhead and effective management technique, usefully bridges the efficiency gap across wide load ranges. This work presents a Q-algorithm for online learning of PDUs, which effectively adapts to the connected loads and incorporates any change in the supplied loads in real time without requiring prior modeling of the converter topology. The proposed power delivery approach, based on a properly tailored Q-algorithm and power gating, improves the overall efficiency while having low complexity. Furthermore, this implementation addresses the following critical constraints of portable systems:

- **Low overhead:** By utilizing a derived formula for power efficiency that relies solely on measured load currents and a pre-characterized regulator lookup table with quiescent current data, the system eliminates the need for continuous sensor polling. This approach significantly reduces the computational cost, improving runtime from 208 ms (sensor polling) to 56 ms (table-based) in high-density-load scenarios. This low overhead is highly appropriate for portable systems.
- **Stability:** The integration of power gating with a bounded number of regulators ensures the stability of the parallel LDO configuration while preserving the transient response required for fast wake-up modes.
- **Scalability:** The model-free Q-agent allows for the deployment to different PDU architectures for various applications to support diverse heterogeneous loads, from sensors to communication modules.

Six realistic IoT scenarios experimentally validate that the proposed method significantly outperforms static assignments. Specifically, the proposed Q-algorithm yields up to a 13% increase in total end-to-end power efficiency $\eta_{tot}$ compared to static PDUs, with an average improvement of 5% across all tested scenarios. Crucially, the system maintains high efficiency (>80%), even at light loads (tens of µA), a regime where traditional autonomous PDUs typically degrade to efficiencies below 26%.

**Author Contributions:** Conceptualization, A.T.; Methodology, A.T.; Software, A.T. and G.A.; Validation, A.T., G.A. and V.F.P.; Formal analysis, A.T. and V.F.P.; Investigation, A.T.; Resources, A.T., G.A. and V.F.P.; Data curation, A.T. and G.A.; Writing—original draft, A.T.; Writing—review and editing, A.T. and V.F.P.; Visualization, A.T.; Supervision, V.F.P.; Project administration, A.T. and V.F.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Prasad, A.; Chawda, P. Power Management Factors and Techniques for IoT Design Devices. In Proceedings of the International Symposium on Quality Electronic Design, Santa Clara, CA, USA, 13–14 March 2018; pp. 364–369. [CrossRef]
2. Wei, K.; Ma, D.B. A 10-MHz DAB Hysteretic Control Switching Power Converter for 5G IoT Power Delivery. *IEEE J. Solid-State Circuits* **2021**, *56*, 2113–2122. [CrossRef]
3. Ababneh, M.M.; Ugweje, O.; Jaesim, A. Optimized Power Management Unit for IoT Applications. In Proceedings of the International Conference on Electronics, Computer and Computation, Abuja, Nigeria, 10–12 December 2019; pp. 1–4. [CrossRef]
4. László-Zsolt, T.; Géza, C.; Csenteri, B. Power Management In IoT Weather Station. In Proceedings of the International Conference and Exposition on Electrical And Power Engineering, Iasi, Romania, 18–19 October 2018; pp. 133–138. [CrossRef]
5. Carreon-Bautista, S.; Huang, L.; Sanchez-Sinencio, E. An Autonomous Energy Harvesting Power Management Unit With Digital Regulation for IoT Applications. *IEEE J. Solid-State Circuits* **2016**, *51*, 1457–1474. [CrossRef]
6. Triki, M.; Wang, Y.; Ammari, A.; Pedram, M. Hierarchical Power Management of a System with Autonomously Power-Managed Components Using Reinforcement Learning. *Integration* **2015**, *48*, 10–20. [CrossRef]
7. Kose, S.; Friedman, E.G. Distributed On-Chip Power Delivery. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2012**, *2*, 704–713. [CrossRef]
8. Jiang, H.; Marek-Sadowska, M.; Nassif, S. Benefits and Costs of Power-Gating Technique. In Proceedings of the International Conference on Computer Design, San Jose, CA, USA, 2–5 October 2005; pp. 559–566. [CrossRef]
9. Hattori, T.; Irita, T.; Ito, M.; Yamamoto, E.; Kato, H.; Sado, G.; Yamada, T.; Nishiyama, K.; Yagi, H.; Koike, T.; et al. Hierarchical Power Distribution and Power Management Scheme for a Single Chip Mobile Processor. In Proceedings of the ACM/IEEE Design Automation Conference, San Francisco, CA, USA, 24–28 July 2006; pp. 292–295. [CrossRef]
10. Brown, J.K.; Abdallah, D.; Boley, J.; Collins, N.; Craig, K.; Glennon, G.; Huang, K.K.; Lukas, C.J.; Moore, W.; Sawyer, R.K.; et al. A 65 nm Energy-Harvesting ULP SoC with 256 kB Cortex-M0 Enabling an 89.1 µW Continuous Machine Health Monitoring Wireless Self-Powered System. In Proceedings of the IEEE International Solid-State Circuits Conference, San Francisco, CA, USA, 16–20 February 2020; pp. 420–422. [CrossRef]
11. Kim, S.; Vaidya, V.; Schaef, C.; Lines, A.; Krishnamurthy, H.; Weng, S.; Liu, X.; Kurian, D.; Karnik, T.; Ravichandran, K.; et al. A Single-Stage, Single-Inductor, 6-Input 9-Output Multi-Modal Energy Harvesting Power Management IC for 100 µW–120 MW Battery-Powered IoT Edge Nodes. In Proceedings of the IEEE Symposium on VLSI Circuits, Honolulu, HI, USA, 18–22 June 2018; pp. 195–196. [CrossRef]
12. Benini, L.; De Micheli, G. *Dynamic Power Management: Design Techniques and CAD Tools*; Springer: Berlin/Heidelberg, Germany, 1998. [CrossRef]
13. Paleologo, G.; Benini, L.; Bogliolo, A.; De Micheli, G. Policy optimization for dynamic power management. In Proceedings of the Design and Automation Conference, 35th DAC (Cat. No.98CH36175), San Francisco, CA, USA, 15–19 June 1998; pp. 182–187. [CrossRef]
14. Benini, L.; Bogliolo, A.; Paleologo, G.; De Micheli, G. Policy optimization for dynamic power management. *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* **1999**, *18*, 813–833. [CrossRef]

15. Ishihara, T.; Yasuura, H. Voltage scheduling problem for dynamically variable voltage processors. In Proceedings of the International Symposium on Low Power Electronics and Design (IEEE Cat. No.98TH8379), Monterey, CA, USA, 10–12 August 1998; pp. 197–202. [CrossRef]
16. Simunic, T.; Benini, L.; Acquaviva, A.; Glynn, P.; de Micheli, G. Dynamic voltage scaling and power management for portable systems. In Proceedings of the Design Automation Conference (IEEE Cat. No.01CH37232), Las Vegas, NA, USA, 22 June 2001; pp. 524–529. [CrossRef]
17. Kim, C.; Roy, K. Dynamic V/sub TH/scaling scheme for active leakage power reduction. In Proceedings of the Design, Automation and Test in Europe Conference and Exhibition, Paris, France, 4–8 March 2002; pp. 163–167. [CrossRef]
18. Qiu, Q.; Pedram, M. Dynamic Power Management Based on Continuous-Time Markov Decision Processes. In Proceedings of the Design Automation Conference, New Orleans, LA, USA, 21–25 June 1999; pp. 555–561. [CrossRef]
19. Dhiman, G.; Rosing, T.S. Dynamic Power Management Using Machine Learning. In Proceedings of the IEEE/ACM International Conference on Computer Aided Design, San Jose, CA, USA, 5–9 November 2006; pp. 747–754. [CrossRef]
20. Yue, S.; Zhu, D.; Wang, Y.; Pedram, M. Reinforcement Learning Based Dynamic Power Management with a Hybrid Power Supply. In Proceedings of the IEEE International Conference on Computer Design, Montreal, QC, Canada, 30 September–3 October 2012; pp. 81–86. [CrossRef]
21. Liu, W.; Tan, Y.; Qiu, Q. Enhanced Q-learning algorithm for Dynamic Power Management with Performance Constraint. In Proceedings of the Design, Automation Test in Europe Conference Exhibition, Dresden, Germany, 8–12 March 2010; pp. 602–605. [CrossRef]
22. Debizet, Y.; Lallement, G.; Abouzeid, F.; Roche, P.; Autran, J.L. Q-learning-based Adaptive Power Management for IoT System-on-Chips with Embedded Power States. In Proceedings of the IEEE International Symposium on Circuits and Systems, Florence, Italy, 27–30 May 2018; pp. 1–5. [CrossRef]
23. Gupta, U.; Mandal, S.K.; Mao, M.; Chakrabarti, C.; Ogras, U.Y. A Deep Q-Learning Approach for Dynamic Management of Heterogeneous Processors. *IEEE Comput. Archit. Lett.* **2019**, *18*, 14–17. [CrossRef]
24. Li, H.; Tian, Z.; Xu, J.; Maeda, R.K.V.; Wang, Z.; Wang, Z. Chip-Specific Power Delivery and Consumption Co-Management for Process-Variation-Aware Manycore Systems Using Reinforcement Learning. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **2020**, *28*, 1150–1163. [CrossRef]
25. Vaisband, I.; Friedman, E.G. Energy Efficient Adaptive Clustering of On-Chip Power Delivery Systems. *Integr. VLSI J.* **2015**, *48*, 1–9. [CrossRef]
26. Tan, Y.; Liu, W.; Qiu, Q. Adaptive Power Management Using Reinforcement Learning. In Proceedings of the IEEE/ACM International Conference on Computer-Aided Design, San Jose, CA, USA, 2–5 November 2009; pp. 461–467.
27. Sutton, R.S.; Barto, A.G. *Reinforcement Learning*; M-Press: Kansas City, MI, USA , 2014; Chapter 3, pp. 79–80.
28. Chen, Z.; Marculescu, D. Distributed reinforcement learning for power limited many-core system performance optimization. In Proceedings of the 2015 Design, Automation Test in Europe Conference Exhibition (DATE), Grenoble, France, 9–13 March 2015; pp. 1521–1526.
29. Ge, Y.; Qiu, Q. Dynamic thermal management for multimedia applications using machine learning. In Proceedings of the ACM/EDAC/IEEE Design Automation Conference, San Diego, CA, USA, 5–9 June 2011; pp. 95–100.
30. ul Islam, F.M.M.; Lin, M. Hybrid DVFS Scheduling for Real-Time Systems Based on Reinforcement Learning. *IEEE Syst. J.* **2017**, *11*, 931–940. [CrossRef]
31. Chen, Y.M.; Chen, C.J. An Event-Driven Self-Clocked Digital Low-Dropout Regulator with Adaptive Frequency Control. *Energies* **2023**, *16*, 4749 . [CrossRef]
32. Wang, D.; Gao, N.; Liu, D.; Li, J.; Lewis, F.L. Recent Progress in Reinforcement Learning and Adaptive Dynamic Programming for Advanced Control Applications. *IEEE/CAA J. Autom. Sin.* **2024**, *11*, 18–36. [CrossRef]
33. Xie, D.; Li, H. Deep Reinforcement Learning Based Collaborative Energy Management for Base Station and Microgrid. In Proceedings of the International Conference on Electronic Information Engineering and Computer Communication (EIECC), Wuhan, China, 27–29 December 2024; pp. 160–163. [CrossRef]
34. Wan, Z.; Huang, Y.; Wu, L.; Liu, C. ADPA Optimization for Real-Time Energy Management Using Deep Learning. *Energies* **2024**, *17*, 4821. [CrossRef]
35. Chen, T.; Dai, Z.; Shan, X.; Li, Z.; Hu, C.; Xue, Y.; Xu, K. Reactive Power Optimization Method of Power Network Based on Deep Reinforcement Learning Considering Topology Characteristics. *Energies* **2024**, *17*, 6454. [CrossRef]
36. Lee, Y.; Bang, S.; Lee, I.; Kim, Y.; Kim, G.; Ghaed, M.H.; Pannuto, P.; Dutta, P.; Sylvester, D.; Blaauw, D. A Modular 1 $mm^3$ Die-Stacked Sensing Platform With Low Power $I^2C$ Inter-Die Communication and Multi-Modal Energy Harvesting. *IEEE J. Solid-State Circuits* **2013**, *48*, 229–243. [CrossRef]
37. Urgaonkar, R.; Kozat, U.C.; Igarashi, K.; Neely, M.J. Dynamic resource allocation and power management in virtualized data centers. In Proceedings of the IEEE Network Operations and Management Symposium, Osaka, Japan, 19–23 April 2010; pp. 479–486. [CrossRef]

38. Kwon, E.; Han, S.; Park, Y.; Kim, Y.H.; Kang, S. Late Breaking Results: Reinforcement Learning-based Power Management Policy for Mobile Device Systems. In Proceedings of the 2020 57th ACM/IEEE Design Automation Conference (DAC), San Francisco, CA, USA, 20–24 July 2020; pp. 1–2. [CrossRef]
39. Kwon, E.; Han, S.; Park, Y.; Yoon, J.; Kang, S. Reinforcement Learning-Based Power Management Policy for Mobile Device Systems. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2021**, *68*, 4156–4169. [CrossRef]
40. Giardino, M.; Schwyn, D.; Ferri, B.; Ferri, A. Low-Overhead Reinforcement Learning-Based Power Management Using 2QoSM. *J. Low Power Electron. Appl.* **2022**, *12*, 29. [CrossRef]
41. Watkins, C.J.C.H.; Dayan, P. Q-learning. *Mach. Learn.* **1992**, *8*, 279–292. [CrossRef]
42. Oh, C.H.; Nakashima, T.; Ishibuchi, H. Initialization of Q-values by Fuzzy Rules for Accelerating Q-learning. In Proceedings of the IEEE International Joint Conference on Neural Networks, Anchorage, AK, USA, 4–9 May 1998; Volume 3, pp. 2051–2056. [CrossRef]
43. Song, Y.; Li, Y.B.; Li, C.H.; Zhang, G.F. An Efficient Initialization Approach of Q-learning for Mobile Robots. *Int. J. Control. Autom. Syst.* **2012**, *10*, 166–172. [CrossRef]
44. Mandal, S.K.; Bhat, G.; Doppa, J.R.; Pande, P.P.; Ogras, U.Y. An Energy-Aware Online Learning Framework for Resource Management in Heterogeneous Platforms. *ACM Trans. Des. Autom. Electron. Syst.* **2020**, *25*, 28. [CrossRef]
45. Ciprut, A.; Friedman, E.G. Stability of On-Chip Power Delivery Systems with Multiple Low-Dropout Regulators. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **2019**, *27*, 1779–1789. [CrossRef]
46. Chang, T.S.; Ramiah, H.; Jiang, Y.; Lim, C.C.; Lai, N.S.; Mak, P.I.; Martins, R.P. Design and Implementation of Hybrid DC-DC Converter: A Review. *IEEE Access* **2023**, *11*, 30498–30514. [CrossRef]
47. Abrar Akram, M.; Ali Wahla, I.; Kim, K.S.; Hwang, I.C. A Four-Phase Digital Buck Converter With MDLL-Based Adaptive Switching Frequency Compensation. *IEEE Access* **2024**, *12*, 180404–180414. [CrossRef]
48. TPS659037 Datasheet. Available online: https://www.ti.com/lit/ds/symlink/tps659037.pdf (accessed on 14 November 2025).
49. Tsiougkos, A.; Pavlidis, V.F. A PWM-free DC-DC Boost Converter with 0.43 V Input for Extended Battery Use in IoT Applications. In Proceedings of the IEEE International Midwest Symposium on Circuits and Systems, Lansing, MI, USA, 9–11 August 2021; pp. 479–483. [CrossRef]
50. MAX17710 Datasheet. Available online: https://datasheets.maximintegrated.com/en/ds/MAX17710.pdf (accessed on 12 November 2025).
51. SPV1050 Datasheet. Available online: https://www.st.com/resource/en/datasheet/spv1050.pdf (accessed on 17 November 2025).
52. Chen, C.; Sun, M.; Wang, L.; Huang, T.; Xu, M. A Fast Transient Response Capacitor-Less LDO with Transient Enhancement Technology. *Micromachines* **2024**, *15*, 299. [CrossRef] [PubMed]
53. Zachos, N.; Gogolou, V.; Noulis, T. A Fully Integrated 1.8 V Low-Power LDO Regulator with Dynamic Transient Control for SoC Applications. *Electronics* **2024**, *13*, 4734. [CrossRef]
54. Arévalos, D.; Marin, J.; Herman, K.; Gomez, J.; Wallentowitz, S.; Rojas, C.A. A Topology-Independent and Scalable Methodology for Automated LDO Design Using Open PDKs. *Electronics* **2025**, *14*, 3448. [CrossRef]
55. Okamura, L.; Morishita, F.; Arimoto, K.; Yoshihara, T. High efficiency Autonomous Controlled Cascaded LDOs for Green Battery system. In Proceedings of the IEEE International Conference on ASIC, Changsha, China, 20–23 October 2009; pp. 336–339. [CrossRef]
56. Oh, T.J.; Hwang, I.C. A 110-nm CMOS 0.7-V Input Transient-Enhanced Digital Low-Dropout Regulator with 99.98% Current Efficiency at 80-mA Load. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **2015**, *23*, 1281–1286. [CrossRef]
57. Rincon-Mora, G.A. Current Efficient, Low Voltage, Low Drop-Out Regulators. Ph.D. Thesis, Department Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, CA, USA, 1996.
58. Ciprut, A.; Friedman, E.G. On the Stability of Distributed On-Chip Low Dropout Regulators. In Proceedings of the IEEE International Midwest Symposium on Circuits and Systems, Boston, MA, USA, 6–9 August 2017; pp. 217–220. [CrossRef]
59. ESP8266EX Datasheet. Available online: https://www.espressif.com/sites/default/files/documentation/0a-esp8266ex_datasheet_en.pdf (accessed on 17 November 2025).
60. nRF52832 Product Specification. Available online: https://infocenter.nordicsemi.com/pdf/nRF52832_PS_v1.4.pdf (accessed on 14 November 2025).
61. Rincón-Mora, G.A. *Analog IC Design with Low-Dropout Regulators*; McGraw-Hill: New York, NY, USA, 2009.
62. Tijsma, A.D.; Drugan, M.M.; Wiering, M.A. Comparing Exploration Strategies for Q-learning in Random Stochastic Mazes. In Proceedings of the IEEE Symposium Series on Computational Intelligence, Athens, Greece, 6–9 December 2016; pp. 1–8. [CrossRef]
63. Chen, J.; Yin, M.; Duan, X.; Jiao, B. Q-Learning Based Selection Strategies for Load Balance and Energy Balance in Heterogeneous Networks. In Proceedings of the International Conference on Computer and Communication Systems, Shanghai, China, 15–18 May 2020; pp. 728–732. [CrossRef]

64. INA226 36V, 16-Bit, Ultra-Precise I2C Output Current, Voltage, and Power Monitor with Alert. Available online: https://www.ti.com/lit/ds/symlink/ina226.pdf?ts=1768878771120&ref_url=https%253A%252F%252Fwww.google.com%252F (accessed on 19 September 2025).
65. L78L Series: Positive Voltage Regulators. Available online: https://www.tme.eu/Document/b4d1ce27007259e9e680165d5f11d754/l78l.pdf (accessed on 19 September 2025).
66. LM2937 500mA Low Dropout Regulator. Available online: https://www.ti.com/lit/ds/symlink/lm2937.pdf (accessed on 19 September 2025).
67. L78 Series: Positive Voltage Regulators. Available online: https://eu.mouser.com/datasheet/2/389/l78-1849632.pdf (accessed on 19 September 2025).
68. L78S Series: 2A Positive Voltage Regulators. Available online: https://gr.mouser.com/datasheet/2/389/l78s-1849452.pdf (accessed on 19 September 2025).
69. LD33V (LD1117 Series): Low Drop Fixed and Adjustable Positive Voltage Regulators. Available online: http://www.st.com/st-web-ui/static/active/en/resource/technical/document/datasheet/CD00000544.pdf (accessed on 19 September 2025).
70. LM3940: 1-A Low Dropout Regulator for 5V to 3.3V Conversion. Available online: https://www.ti.com/lit/ds/symlink/lm3940.pdf (accessed on 19 September 2025).
71. Huang, C.H.; Ma, Y.T.; Liao, W.C. Design of a Low-Voltage Low-Dropout Regulator. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **2014**, *22*, 1308–1313. [CrossRef]
72. Cheah, M.; Mandal, D.; Bakkaloglu, B.; Kiaei, S. A 100-mA, 99.11% Current Efficiency, 2-mVpp Ripple Digitally Controlled LDO With Active Ripple Suppression. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **2017**, *25*, 696–704. [CrossRef]
73. Laleni, N.; Tsiougkos, A.; Pavlidis, V. Hybrid Capacitor-less LDO with Switched-Mode Dead-Zone Control. In Proceedings of the International Conference on Synthesis, Modeling, Analysis and Simulation Methods, and Applications to Circuit Design, Online, 19–22 July 2021.
74. Zarate-Roldan, J.; Wang, M.; Torres, J.; Sánchez-Sinencio, E. A Capacitor-Less LDO With High-Frequency PSR Suitable for a Wide Range of On-Chip Capacitive Loads. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **2016**, *24*, 2970–2982. [CrossRef]
75. Luria, K.; Shor, J.; Zelikson, M.; Lyakhov, A. 8.7 Dual-Use Low-Drop-Out Regulator/Power Gate with Linear and On-Off Conduction Modes for Microprocessor on-Die Supply Voltages in 14nm. In Proceedings of the IEEE International Solid-State Circuits Conference, San Francisco, CA, USA, 22–26 February 2015; pp. 1–3. [CrossRef]
76. Liu, X.; Krishnamurthy, H.K.; Barrera, C.P.; Han, J.; Bhatla, R.M.N.; Chiu, S.; Ahmed, Z.K.; Ravichandran, K.; Tschanz, J.W.; De, V. A Dual-Rail Hybrid Analog/Digital LDO with Dynamic Current Steering for Tunable High PSRR and High Efficiency. In Proceedings of the IEEE Symposium on VLSI Circuits, Honolulu, HI, USA, 16–19 June 2020; pp. 1–2. [CrossRef]


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.