



Efficient Data Reduction in Multisource Hybrid Electrical Systems Using Frequency-Domain Decimation Techniques

Mariam Hadj Salem¹, Yemna Bensalem¹, Karim Mansouri², Mounir Bouzguenda³,
Ezzeddine Touti^{4,*}

¹MACS LR16ES22, University of Gabes, Gabes, Tunisia,

(Mariem) mariam.hadsalem@outlook.com and (Yemna) bensalem.yemna@macs.tn

²ESEO-Tech, University of Angers, Angers, karim.mansouri@eseo.fr

³ Department of Electrical Engineering, College of Engineering, King Faisal University,
Al Ahsa, 31982, Saudi Arabia, mbuzganda@kfu.edu.sa.

⁴ Center for Scientific Research and Entrepreneurship, Northern Border University, Arar
73213, Saudi Arabia, esseddine.touti@nbu.edu.sa

*Correspondence: esseddine.touti@nbu.edu.sa

Abstract:- This work addresses the challenge of reducing data size without compromising analytical accuracy in simulations of a hybrid multisource electrical system. The system generates large volumes of time-series data—voltage, current, and power waveforms whose storage and processing can become computationally expensive. To mitigate this, we propose a reduction strategy based on the Fast Fourier Transform (FFT). A decimation coefficient is computed for each variable in the frequency domain, and the signals are then reconstructed using the Inverse FFT (IFFT). Simulation results show that the method achieves a substantial decrease in data volume while preserving the essential features of the original signals, ensuring the reliability of subsequent analyses and machine-learning-based fault diagnostics.

Keywords: FFT; IFFT; Optimization; Multisource system; Hybrid system; Data size; Data reduction

1. INTRODUCTION

Real-time monitoring and control are essential for optimizing the performance and reliability of multisource hybrid electrical systems, which often power critical infrastructure such as microgrids, electric vehicles, and industrial plants[1], [2], [3], [4]. These systems rely on a diverse array of data sources, including voltage and current sensors, power quality analyzers, and environmental monitoring equipment[5], [6], [7], [8]. However, the high sampling rates and extensive sensor networks result in a data deluge that can overwhelm processing capabilities and limit the effectiveness of control algorithms. [9], [10], [11], [12].

The contemporary digital landscape is characterized by an exponential surge in data, often referred to as "big data," which presents significant challenges related to storage, transmission, and computational efficiency across a wide range of fields [13], [14], [15], [16]. In recent years, the sheer volume, velocity, and variety of data have necessitated a paradigm shift from traditional handling methods toward more sophisticated optimization and reduction techniques [17], [18], [19], [20]. This issue is particularly pronounced in data-intensive domains such as



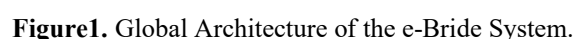
cloud computing, artificial intelligence, and genomics. For instance, in healthcare, the integration of Internet of Things (IoT) devices generates a continuous stream of sensitive data that must be efficiently managed and stored [21]. Similarly, the training of machine learning and deep learning models requires vast datasets, leading to high energy consumption and computational costs [22], [23], [24], [25], [26], [27]. Another critical area is genomics, where high-throughput technologies like nanopore sequencing produce immense raw signal data, posing significant bottlenecks for analysis and storage [28], [29], [30], [31], [32], [33].

To combat these challenges, researchers are developing innovative methods that go beyond traditional compression. For instance, Authors in [27] [34], [35], [36] explored and evaluated several strategies, including dimensionality and numerosity reduction, to minimize dataset size, thereby reducing computational cost and energy consumption. Another notable approach used in [37] provides a comprehensive survey of data reduction methods in the context of big data [28], [38], [39], [40], [41]. The paper systematically classifies various techniques into categories such as sampling, dimensionality reduction, data compression, and methods based on data mining and deep learning. Authors in [42] [43], [44], [45], [46], [47], [48], [49] explore the optimization of SQL databases for big data workloads through a case study in the cable manufacturing industry. The authors demonstrate the effectiveness of strategic normalization and advanced analytical techniques within Microsoft SQL Server. Their findings reveal that these methods significantly reduce data redundancy and dramatically improve query execution times, offering a practical framework for using relational databases to manage large-scale data in an industrial setting. These examples illustrate the diverse and specialized solutions being employed to tackle the problem of big data, showcasing a trend toward domain-specific and purpose-built algorithms to optimize data handling. Building on these advances, our research addresses a big-data challenge that emerged during the modeling and simulation of multiple faults in a hybrid multi-source electrical system. In this context, the need to simulate numerous fault scenarios over extended time horizons generated massive datasets, creating two critical issues: high storage requirements and lengthy simulation times. These datasets are subsequently used to train machine-learning algorithms for automated fault diagnosis, so preserving the essential information while reducing their size is crucial. To overcome these constraints, we propose a data-reduction methodology based on the Fast Fourier Transform (FFT). By exploiting the frequency-domain characteristics of the simulated electrical signals, our approach identifies and removes redundant information while retaining the features necessary for accurate fault analysis and machine-learning tasks. This work, therefore, focuses on achieving significant reductions in both storage space and computational time, paving the way for more efficient diagnostics of faults in complex hybrid energy systems.

The Fourier Transform, a mathematical concept first introduced by Jean-Baptiste Joseph Fourier, has become a cornerstone for addressing computational complexity and is a

This paper is organized as follows. Section 2, “The Multisource System Under Study,” presents the architecture and key components of the hybrid energy system. Section 3, “Problem Statement and Formulation,” defines the data-size and simulation-time challenges and formalizes the objectives of the study. Section 4, “Fast Fourier Transform,” describes the FFT-based method used for data reduction. Section 5, “Simulation Results,” reports and discusses the outcomes of applying the proposed approach. Finally, Section 6, “Conclusion,” summarizes the main findings and highlights the benefits of the proposed strategy for efficient data storage and reduced computational time.

The architecture of the multi-source system studied in this work is integrated within the framework of an e-Bride project. Figure 1 shows the general architecture of the system.





This architecture consists of four main stages:

- (a) Sources Stage: Allows for two types of sources, conventional and renewable, and for storing them in a battery.
- (b) Energy Conversion Stage: Composed of converters such as choppers, rectifiers, etc.
- (c) Energy Management System (EMS): To control the flow of energy.
- (d) Human-Machine Interface (HMI): Necessary for practical testing.

In its basic version, the test bench implements:

- ✓ A generator set (GE) consisting of an alternator driven by a diesel engine, here, emulated by a DC motor driving a synchronous machine.
- ✓ A photovoltaic (PV) panel.
- ✓ A lead-acid storage battery.

It is designed to be scalable by modifying existing elements or adding new sources. All sources share their electrical energy on a direct current (DC) bus. Energy flows are controlled by converters driven according to the orders of the Energy Management System (EMS). The user can configure each component or algorithm and monitor the operation of all equipment through the Human-Machine Interface (HMI). The latter also allows for recording of all useful values provided by the multiple instruments during the test. When solar energy is insufficient, the diesel generator serves as a backup power source. However, due to its slow startup, the quality of energy deteriorates because of the lack of power. For this reason, a supercapacitor was added during the startup phase of the diesel generator to ensure power balance and maintain a stable DC bus voltage. This component was chosen because of its fast response and high-power density. Finally, we obtain an architecture as presented in Fig.2.

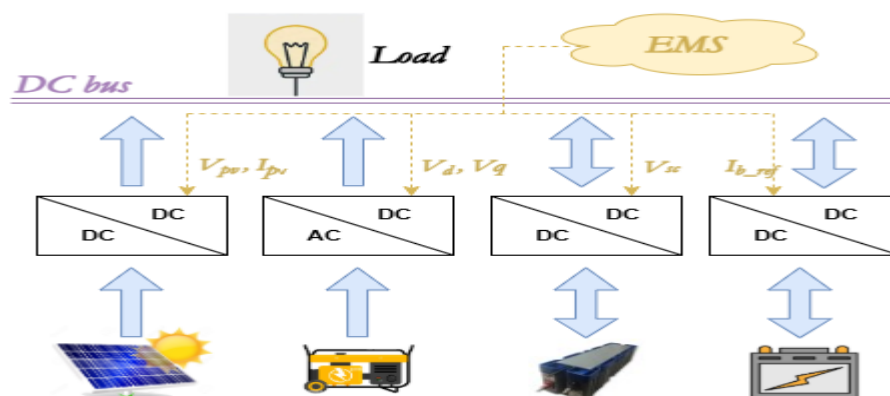


Figure 2. Architecture of the studied multi-source system.



3. PROBLEM STATEMENT AND FORMULATION

A. PROBLEM FORMULATION

The main objective of this work is to test several faults on our hybrid system and collect the simulation data for later use in the diagnostic phase. Indeed, the in-depth evaluation revealed the necessity of performing many simulations, thereby generating a considerable amount of data to be stored, exceeding our available capacities in terms of time and space. Faced with this reality, this section proposes a strategic approach to address this dilemma by opting for an efficient data reduction method. The central idea of this transition is to strike a balance between the need to retain relevant data and the practical constraints imposed by limited resources. The approach consists of using the Fast Fourier Transform (FFT) algorithm to intelligently compress the data while preserving the essence of the recorded waveforms. This innovative method aims to maintain the representativeness of the data, thus enabling a significant reduction in the size of the recordings without compromising the quality of the essential information. In this context, the storage issue has been addressed as an optimization problem, where the Fast Fourier Transform (FFT) serves as the core objective function. By adopting this approach, we ensure an efficient balance between data reduction and information preservation.

B. PROPOSED DATA REDUCTION METHOD

An optimization problem involves finding the best solution that maximizes or minimizes an objective function while satisfying a set of constraints [62], [63]. In other words, the goal is to determine the optimal configuration or combination of decision variables that yields the best outcome according to predefined criteria, without violating any imposed limits.

The optimization problem in this study aims to find the optimal balance between data size reduction and result accuracy for a multisource system simulation. The objective is to reduce the overall data size of several key variables—such as P_{load} , P_{batt} , P_{pv} , $SOCB$, and various currents and voltages—by applying a decimation coefficient.

The data reduction problem is formulated as a multi-objective optimization problem. The goal is to find a decimated signal $y(m)$ that simultaneously minimizes both data size and reconstruction error with respect to the original signal $x(n)$.

Let $x(n)$ be the original time-domain signal, where $n=0,1,\dots,N-1$ and N is the total number of samples. Let $y(m)$ be the decimated signal, where $m=0,1,\dots,M-1$ and M is the number of samples after decimation, with $M < N$.

The objective is to minimize the cost function $J(y)$, defined as follows:

$$\text{Minimize } J(y) = \alpha \cdot \text{Size}(y) + (1 - \alpha) \cdot \text{Error}(x, \tilde{x}) \quad (1)$$

Where:



- $Size(y)$ is a function representing the size of the decimated signal $y(m)$.
- $Error(x, \tilde{x})$ is a function that quantifies the difference between the original signal $x(n)$ and the reconstructed signal $\tilde{x}(n)$ (obtained by the Inverse Fourier Transform of the decimated signal).
- α is a weighting coefficient, $0 \leq \alpha \leq 1$, which adjusts the trade-off between data size reduction and reconstruction error.

The reconstruction error can be measured by the Mean Squared Error (MSE):

$$Error(x, \tilde{x}) = \frac{1}{N} \sum_{n=0}^{N-1} (x(n) - \tilde{x}(n))^2 \quad (2)$$

The constraints of this problem are:

- ❖ The decimation coefficient C must be a positive integer, $C \in \mathbb{Z}^+$.
- ❖ The size of the decimated signal M must be less than the size of the original signal N .
- ❖ The reconstruction error must be below a predefined tolerance threshold, $Error(x, \tilde{x}) < \epsilon$

The chosen method for this optimization is the FFT. The FFT algorithm is used to transform the data from the time domain to the frequency domain. This allows for the identification and removal of less significant frequency components, which effectively compresses the data while preserving its essential characteristics. The strategic application of FFT enables the efficient reduction of data size without compromising the quality of results.

4. PREPARE FAST FOURIER TRANSFORM

The principle of FFT relies on the efficient decomposition of a time-domain signal into its frequency components, enabling an in-depth analysis of the signal's spectral content. In practical terms, the FFT converts a discrete signal in the time domain into its equivalent in the frequency domain, highlighting frequencies and their amplitudes [64]. The fundamental idea behind FFT is to leverage the properties of complex unit roots to group terms in the calculation of the Fourier transform [65]. By cleverly performing these groupings, the FFT achieves a significant reduction in the number of operations required to obtain the Fourier transform. The Discrete Fourier Transform (DFT) of a time-domain signal $x(n)$ is given by:

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot e^{-j2\pi kn/N} \quad (3)$$

Where:

- $X(k)$ is the Fourier coefficient corresponding to frequency k/N .
- $k=0, 1, \dots, N-1$.



The Inverse Discrete Fourier Transform (IDFT), which allows for reconstructing the time-domain signal from the Fourier coefficients, is given by:

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) \cdot e^{j2\pi kn/N} \quad (4)$$

The FFT is an algorithm that computes the DFT much more efficiently.

In the context of this work, we propose a frequency-based decimation process. So let $X(k)$ be the FFT of the original signal $x(n)$. The decimation process consists of selecting a subset of the Fourier coefficients $X(k)$ based on their magnitude. Coefficients with an amplitude below a certain threshold are set to zero.

Let $X'(k)$ be the decimated Fourier coefficients. The reconstructed time-domain signal $\tilde{x}(n)$ is obtained by applying the IDFT to $X'(k)$

$$\tilde{x}(n) = \frac{1}{N} \sum_{k=0}^{N-1} X'(k) \cdot e^{j2\pi kn/N} \quad (5)$$

The decimation coefficient is implicitly defined by the number of $X'(k)$ coefficients that are retained. Given that the primary purpose of employing this algorithm is to calculate a decimation coefficient for each variable, the calculation principle remains consistent across all variables.

The formula utilized to calculate the decimation coefficient is established as follows:

A. SAMPLING FREQUENCY (F_{ech}):

According to the Nyquist-Shannon sampling theorem, for a signal with a maximum frequency of F_{max} , the sampling frequency must be at least twice as high to avoid aliasing.

$$F_{ech} \geq 2 \times F_{max} \quad (6)$$

B. SAMPLING PERIOD (T_{ech}):

The sampling period is the inverse of the sampling frequency.

$$T_{ech} = \frac{1}{F_{ech}} \quad (7)$$

C. Decimation Coefficient (C):

The decimation coefficient is the ratio of the new sampling period (T_{ech}) to the original sampling period (Δt). It's a reduction factor that determines how many samples to keep.

$$C = \frac{T_{ech}}{\Delta t} \quad (8)$$

Applying FFT within our context allows us to trace the spectrum of the signal. This process involves setting a maximum frequency (F_{max}) according to the Shannon theorem, which entails that the sampling frequency (F_{ech}) must be at least twice the maximum frequency to



ensure accurate signal representation. With F_{ech} established, the time interval between samples (T_{ech}) can be determined by taking the reciprocal of F_{ech} . Furthermore, by considering the pre-FFT sampling period (Δt), the decimation coefficient can be computed by dividing T_{ech} by Δt . This coefficient signifies the degree of data reduction achieved while transitioning from the time domain to the frequency domain. Therefore, the general formula derived from this process encapsulates the steps of determining F_{max} , calculating F_{ech} , deriving T_{ech} , and ultimately evaluating the decimation coefficient based on the defined Δt . This systematic approach facilitates the conversion of signal data into the frequency domain, granting insights into the spectral characteristics critical for analysis

5. SIMULATION RESULTS

In pursuit of optimizing data storage for a hybrid system, the focal objective of this study is to effectively manage the recorded data footprint. To achieve this, a comprehensive approach has been adopted, involving the simulation of the system's behavior under the influence of various faults. This emulation of unstable signal states is pivotal as it provides insights into scenarios where the system's dynamics become non-stationary. By exposing the system to these fluctuations, the optimization process aims to derive a decimation coefficient that remains compatible and effective across a spectrum of potential system perturbations. The ultimate aspiration is to tailor the decimation process such that it seamlessly accommodates the dynamic nature of the system while conserving storage space without compromising the fidelity of critical information. Given the wide range of variables examined in this study, our emphasis will be placed on exploring select examples among this set of variables. Also, we conducted two distinct simulations to validate the algorithm followed in this study.

A- SIMULATION 1

This simulation involved a short-circuit (CC) event in the diode of the PV boost converter between 5 and 6 seconds, followed by a commutation event (CO) of the same diode between 15 and 16 seconds. The simulation was carried out with a stop time of 20 seconds, utilizing the ode23tb solver. The total simulation time amounted to 2087.75 seconds. Fig. 3 depicts the multifaceted analysis of current I_b dynamics: a) time evolution, b) limited inverse FFT, c) frequency spectrum.

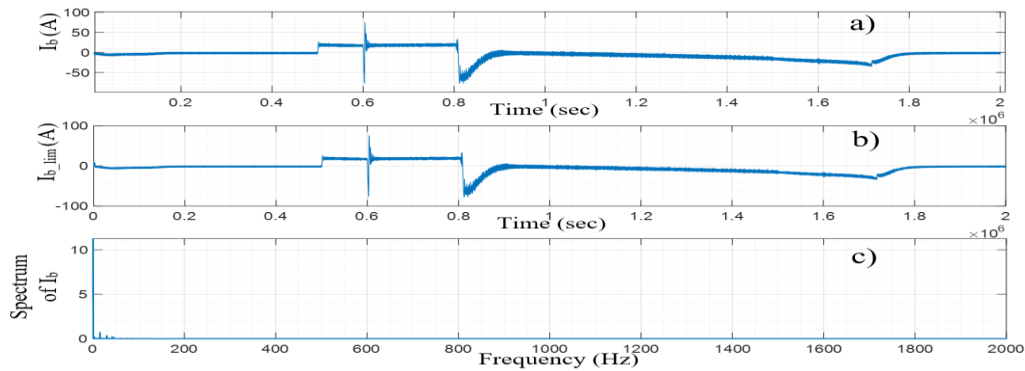


Figure 3. Multifaceted Analysis of Current I_b Dynamics: a) Time Evolution, b) Limited Inverse FFT, c) Frequency Spectrum.

In the visualization of results, three distinct curves provide insights into the behavior of the current variable, I_b (Fig. 4). The uppermost curve illustrates the progression of current I_b over time. The middle curve showcases the outcome of a limited inverse FFT applied to $I_b(t)$, shedding light on its frequency components. Lastly, the lowermost curve represents half of the amplitude spectrum derived from $I_b(t)$, revealing characteristics of the signal in the time domain. This comprehensive representation facilitates a comprehensive understanding of I_b 's dynamics from various analytical perspectives. Through the curve of half the amplitude spectrum of $I_b(t)$, the choice of $F_{\max}$ is determined.

If we choose $F_{\max} = 50$ Hz, and according to the Shannon theorem, we have:

$F_{ech} \geq 2 \times F_{\max} = 100$ Hz, $T_{ech} = \frac{1}{F_{ech}} = 10$ ms and $\Delta t = 10$ μ s, the sampling period before FFT

The decimation = $\frac{T_{ech}}{\Delta t}$

For I_b : the decimation coefficient is $t = 1000$

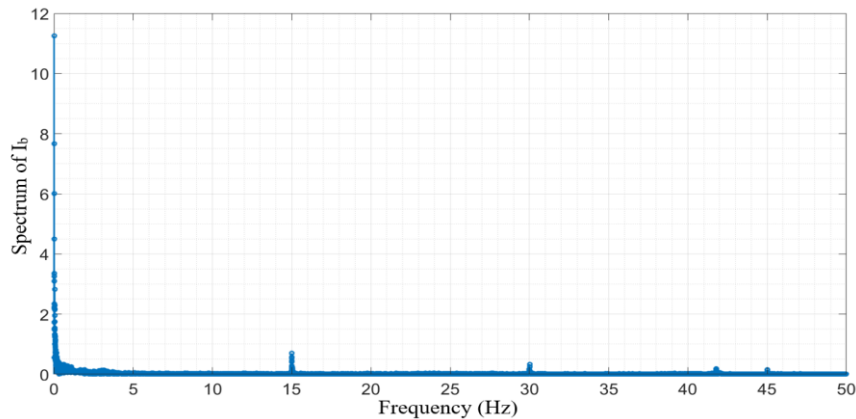


Figure 4. Spectrum of $I_b(f)$ over an interval $[0, F_{\max}]$.



In the context of validating the selection of $F_{\max}=50$ Hz for I_b , the spectrum curve serves as a crucial analytical tool. This curve visually presents the distribution of frequency components within the I_b signal. By examining the spectrum over the interval $[0, F_{\max}]$, it becomes possible to discern the dominant frequency content within this range. In essence, the spectrum curve allows us to identify whether significant frequency components exist below $F_{\max}=50$ Hz, thus justifying the choice of this value. Peaks and variations within the curve provide insights into the presence of distinct frequencies, harmonics, or other relevant characteristics of the signal. Through this analysis, we can gain confidence in the appropriateness of the chosen $F_{\max}$ value and its alignment with the signal's spectral characteristics. As such, the spectrum curve serves as a valuable tool to empirically support the decision-making process regarding parameter choices in the optimization strategy.

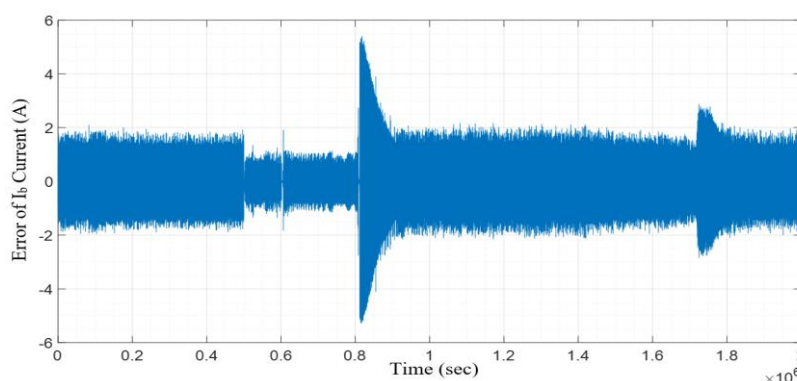


Figure 5. The difference between $I_b(t)$ and $I_{b_lim}(t)$.

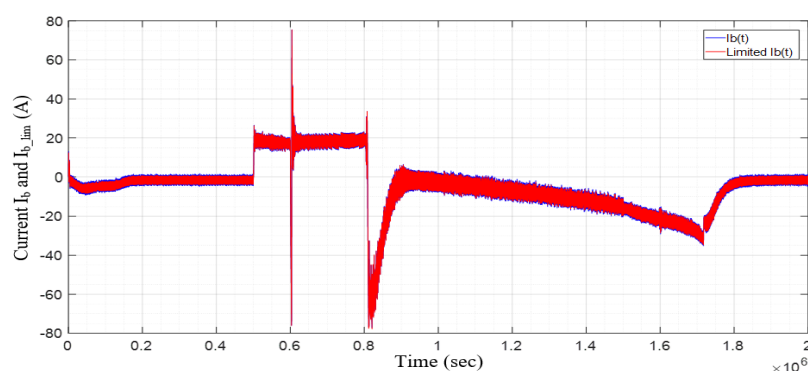


Figure 6. $I_b(t)$ and $I_{b_lim}(t)$ curve

To further validate this choice, we plotted the difference between the original time-domain curve of I_b and the limited I_b curve after the application of FFT (Fig.5). Upon inspection, the difference curve exhibits oscillations around nearly zero. While these oscillations may be negligible due to their average value being close to zero, the traces of the I_b and I_{b_lim} curves in Fig. 6 also exhibit minimal disparities. This is evident in both the overall shape and in



pronounced peaks. These results collectively indicate that the transformation introduced by the FFT preserves the essential characteristics of the signal, as the variations primarily converge around an average of zero. This outcome underscores the successful selection of F_{max} and the appropriateness of the chosen decimation process, as the fidelity of the signal is well-maintained, even in critical instances.

B- PV power

The same process is applied to the solar panel's power. Subsequently, we depicted the evolution curves of P_{pv} as well as the process of selecting F_{max} as seen in Fig.7.

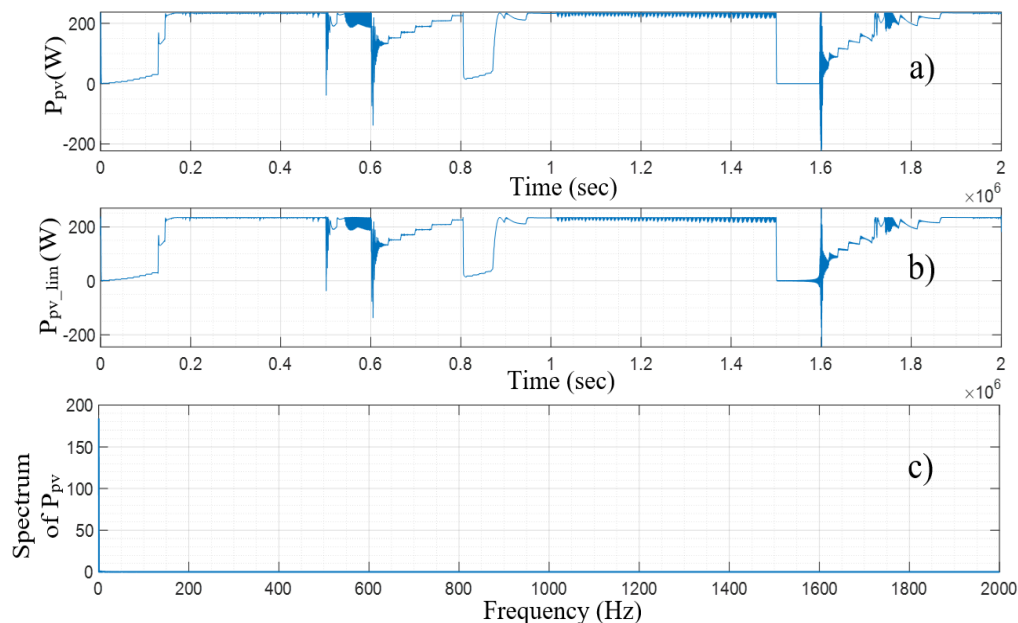


Fig. 7. Multifaceted Analysis of P_{pv} Dynamics: a) Time Evolution, b) and Limited Inverse FFT, c) Frequency Spectrum.

We always have the same curves: the top curve shows the evolution of power P_{pv} over time. Whereas the middle curve represents the limited inverse FFT of $P_{pv}(t)$ and the bottom curve shows the single-sided amplitude spectrum of $P_{pv}(t)$. If $F_{max}=10$ Hz, and according to the Shannon theorem, we get: $F_{ech} \geq 2 \times F_{max} = 20\text{Hz}$, $T_{ech} = 0.05$ s and $\Delta t = 10$ μs , the sampling period before FFT. For P_{pv} : decimation coefficient = 5000.

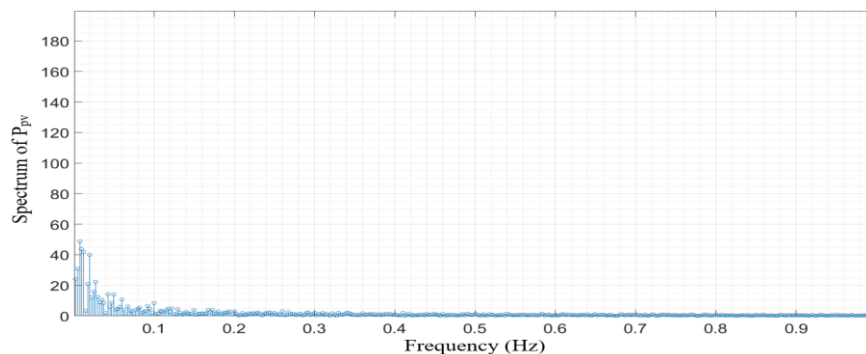


Figure 8. Zoomed-in Spectrum of $P_{pv}(f)$ within $[0, F_{max}]$.

The application of Fast Fourier Transform (FFT) to the power variable P_{pv} yielded results similar to those observed for the I_b variable. Upon examining the P_{pv} spectrum curve over the interval $[0, F_{max}]$ in Fig.8, we found that this curve reinforces the relevance of the chosen F_{max} , confirming its suitability for our analysis.

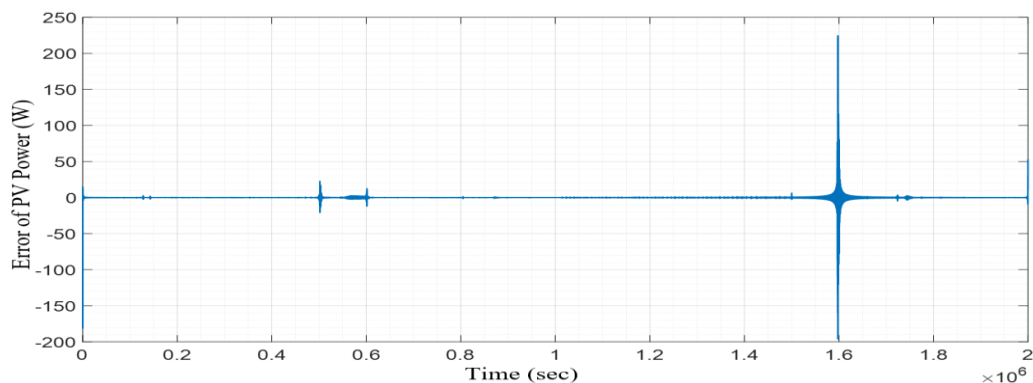


Figure 9. difference between $P_{pv}(t)$ and $P_{pv_lim}(t)$.

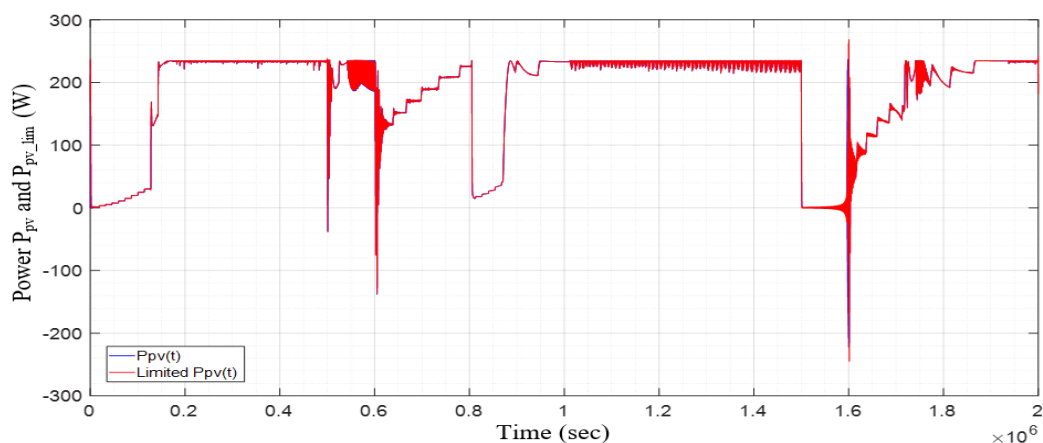


Figure 10. $P_{pv}(t)$ and $P_{pv_lim}(t)$ curve.



Furthermore, by comparing the difference curves between P_{pv} and P_{pv_lim} , along with the overlaid plots of P_{pv} and P_{pv_lim} (Fig.9 and Fig.10), a striking agreement becomes evident. The variations between the two curves converge around an average difference close to zero, reaffirming the quality of the F_{max} choice. These observations highlight FFT's ability to preserve essential signal characteristics, bolstering our confidence in the selected decimation methodology and its capability to maintain signal fidelity across various scenarios.

B. SIMULATION 2

Simulation 2 was designed as a direct extension of the first scenario to more thoroughly evaluate the proposed method. In addition to the short-circuit and commutation faults applied to the PV boost-converter diode in Simulation 1, a quantitative temperature fault was introduced on the photovoltaic panel. This fault imposed three distinct temperature levels 45°C, 50°C, and 75°C occurring at 4 s, 5 s, and 6 s, respectively. To capture the system's response to these successive thermal changes, the model automatically executed three separate runs, each reflecting one of the specified temperature values. The recorded execution times for these runs were 1123.41 s, 1959.15 s, and 1026.79 s, demonstrating the additional computational load required when incorporating quantitative variations into the fault-diagnosis study.

1. PV power results

Because all three additional faults are applied directly to the photovoltaic (PV) panel, the most significant variations are observed in the panel's own output. In a multisource system, other energy sources continue to supply the load even if one source is degraded, so their contributions remain largely stable, with only temporary fluctuations appearing in their measurements. Consequently, while a full comparison of all system variables is presented later in the paper, this section focuses specifically on a detailed analysis of the PV panel's power response, highlighting the primary effects of the introduced faults.

For the same analysis applied to P_{pv} and with the choice of $F_{max}=10$ Hz, we observe the following results. By examining the difference curve between P_{pv} and P_{pv_lim} in Fig.11, as well as the concurrent plots of P_{pv} and P_{pv_lim} , we reinforce the validity of the chosen F_{max} in varying scenarios. The difference curve exhibits minimal variations, primarily close to zero, attesting to the robustness of the maximum frequency choice despite changes in simulation parameters and the introduction of additional faults.

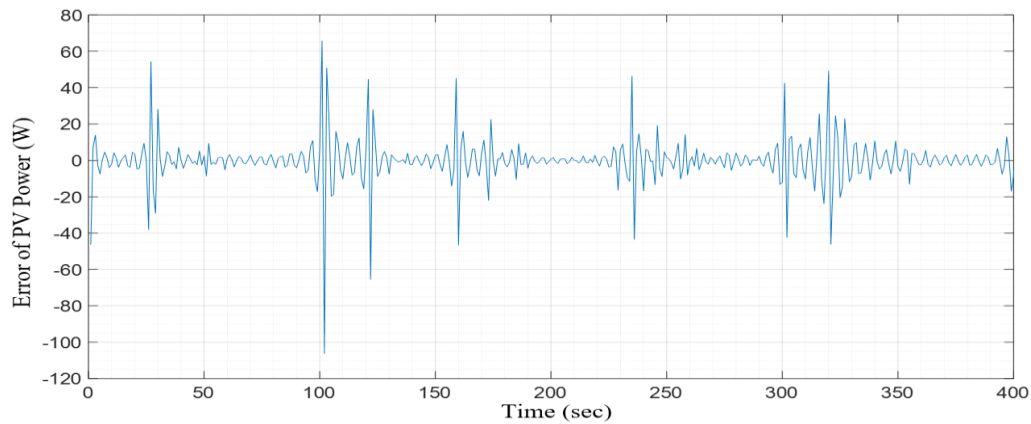


Figure 11. The difference between $P_{pv}(t)$ and $P_{pv_lim}(t)$.

Upon observing the overlaid traces of P_{pv} and P_{pv_lim} (Fig.12), a striking resemblance becomes evident. This coherence demonstrates that the selected F_{max} consistently captures the essential signal characteristics, regardless of disruptions or introduced variations. This in-depth analysis bolsters confidence in the chosen F_{max} as an appropriate frequency threshold, confirming its relevance in diverse evolving contexts.

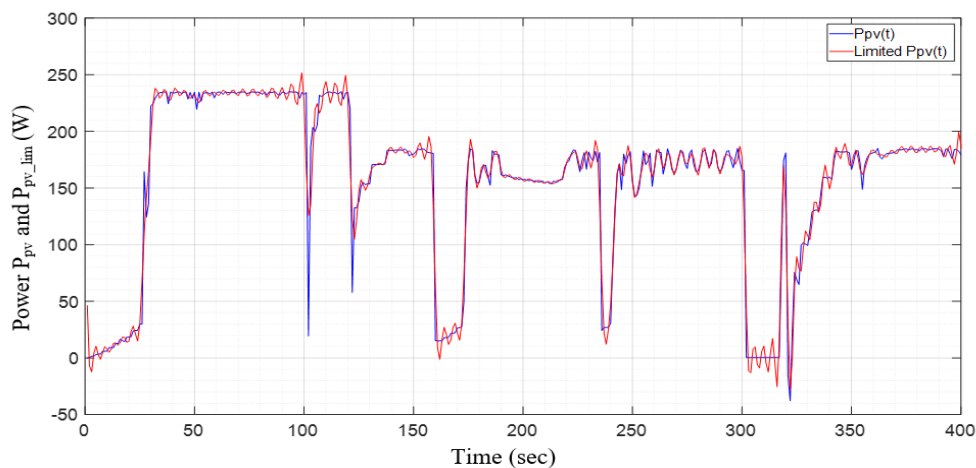


Figure 12. $P_{pv}(t)$ and $P_{pv_lim}(t)$ curve.

C. SUMMARY

We have compiled the decimation coefficients for each variable into a table to offer a comprehensive overview of their utilization within our methodology. While we have detailed the calculation of these coefficients for the I_b and P_{pv} variables, it is important to note that the underlying principle remains consistent for all variables.

The calculation approach is tailored based on the specific properties of each variable, while maintaining the coherence of the optimization process. This systematic approach ensures that



each variable benefits from appropriate decimation, aligned with data size requirements, while preserving result fidelity. Table 1 thus provides a convenient summary of the methodology applied across all variables, reinforcing the consistency and rigor of our optimization approach.

TABLE 1. DIFFERENT DECIMATION FACTORS.

Components		Decimation factor
The load power P_{load}		1000
The battery power P_{batt}		250
The DG power P_{red}		5000
The PV power P_{pv}		5000
The battery SOC		10 000
The DC bus voltage V_{DC}		1000
The battery current I_b		1000
The quadratic current I_q		1060
The direct current I_d		2500
The current/voltage references	I_{bref}	1000
	I_{dref}	2500
	I_{qref}	1060
	V_{dcref}	1000

To validate our choices of maximum frequency and decimation, we are reapplying them in a different simulation context. Utilizing the same frequency threshold and decimation values.

In the final step of validation, we present the waveforms of all variables with and without decimation under the conditions of the first simulation.

Based on the results obtained the comparison between the graphs with and without decimation shows that the decimation technique effectively reduces the amount of data while preserving the key characteristics of the signals (Fig. 13-Fig. 19). The decimation process, which involves downsampling the data, leads to a compressed time scale on the x-axis, but the overall trends, peak values, and signal shapes remain remarkably similar. This suggests the chosen decimation parameters are appropriate for this application.

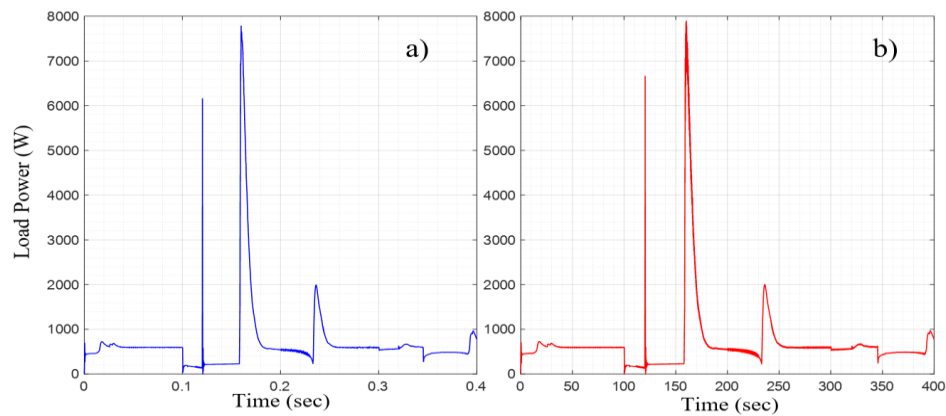


Figure 13. The load power variation: a) Without decimation, b) With decimation.

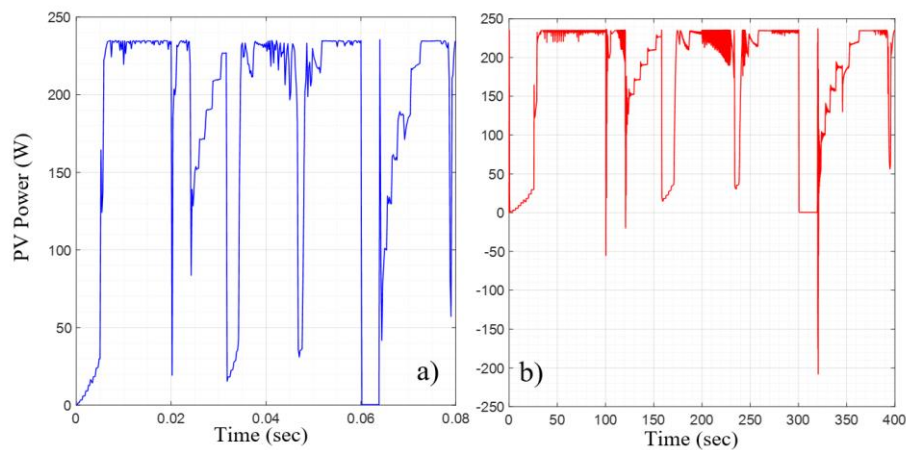


Fig. 14. The Ppv power variation: a) Without decimation, b) With decimation.

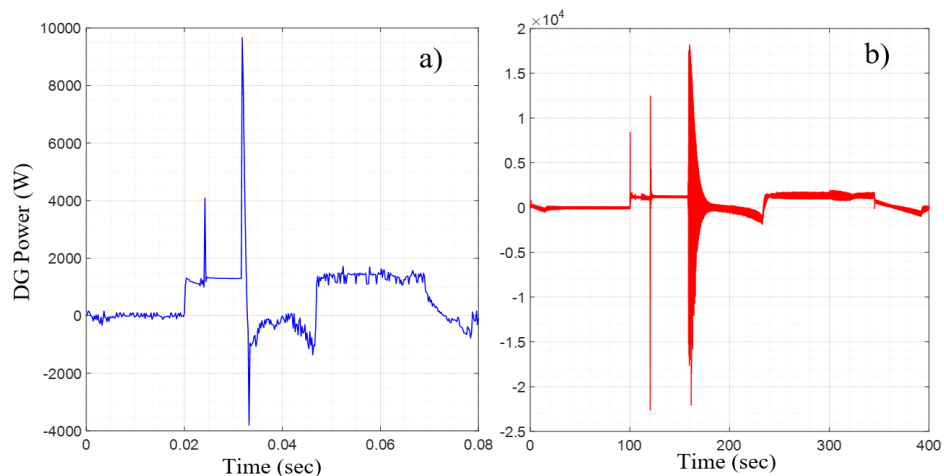


Figure 15. The DG power variation: a) Without decimation, b) With decimation.

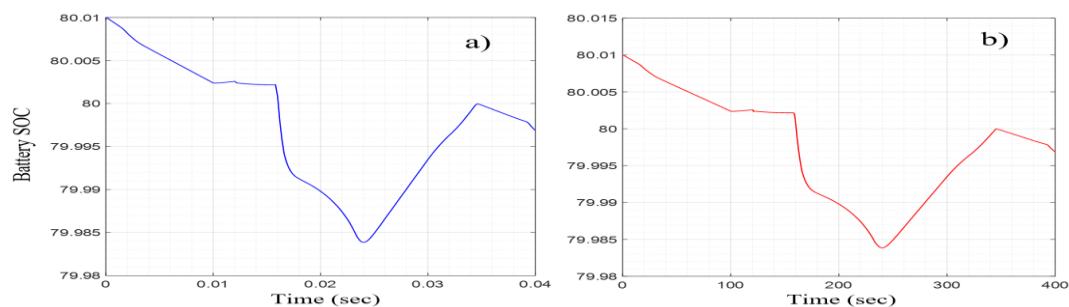


Figure 16. The battery's SOC variation: a) Without decimation, b) With decimation.

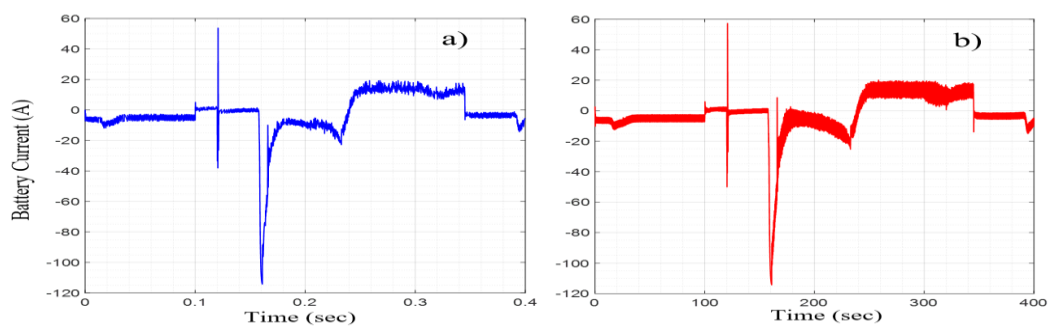


Figure 17. The battery's current variation: a) Without decimation, b) With decimation.

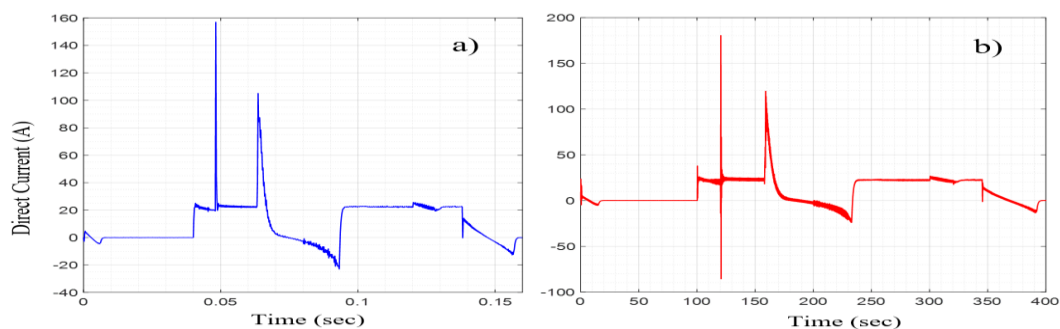


Figure 18. The direct current variation: a) Without decimation, b) With decimation.

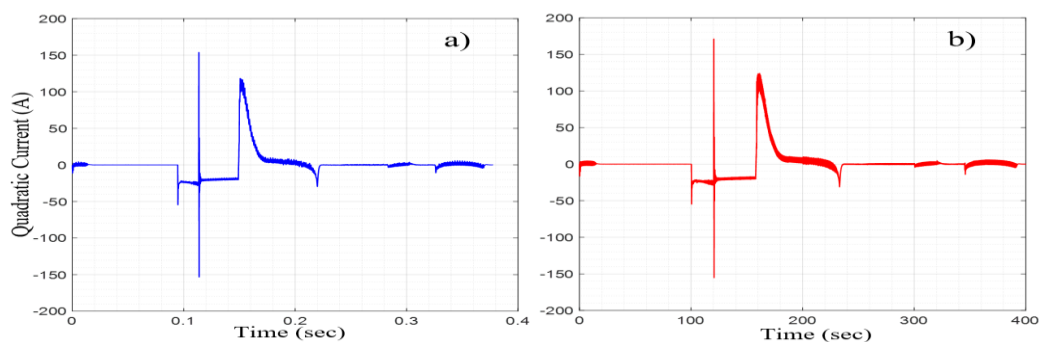


Figure 19. The quadratic current variation: a) Without decimation, b) With decimation.



D. OBSERVATIONS AND INTERPRETATIONS

The graphs illustrate the system's behavior with and without decimation, highlighting the relationship between PV panel faults, battery compensation, and overall power flow. The most significant and dynamic changes are observed in the PV power (PPV) and battery current (Ib) waveforms. This is a direct consequence of the simulation's fault conditions, which are specifically applied to the PV panel. When a fault causes a sudden drop in PV power, the battery is immediately called upon to fill the power gap and maintain the load supply. This quick compensation is visually represented by the large, abrupt negative spikes in the battery current, signifying discharge. The effectiveness of the decimation is evident across all parameters. The graphs for Load Power, Direct Current (Id), and Quadratic Current (Iq) demonstrate that the overall trends and peak values are accurately maintained, even with the compressed time scale. The battery SOC plot, despite its narrow range, perfectly mirrors the subtle discharge and recharge cycles. A key finding from the Id and Iq plots is that their values remain close to zero throughout the simulation. This indicates that the diesel generator is turned off in this scenario, confirming that the battery is the sole component responsible for compensating for the PV faults.

E. Detailed Analysis of the Waveforms

Each set of plots tells a consistent story about the system's resilience. The PV power graphs show multiple instances of sharp power drops, representing the applied faults. These are critical events, and the decimated waveform accurately captures their timing and magnitude. In direct response, the battery current exhibits large negative spikes. These are the "hero" moments of the battery, as it steps in to prevent system instability. This direct cause-and-effect relationship between the PV power drops and the battery current spikes is flawlessly preserved in the decimated data. Furthermore, the Load Power graph shows the demand on the system, with several peaks indicating a variable load. The decimation method successfully captures these peaks, ensuring that the analysis of the system's performance under changing load conditions remains reliable. The near-zero values for the direct (Id) and quadratic (Iq) currents further underscore the system's operational mode, where the diesel generator is inactive and all regulation relies on the PV-battery combination. The integrity of these signals across both datasets validates that the decimation technique doesn't compromise critical information.

F. Advantages of Decimation

The successful application of this decimation approach provides significant practical benefits, most notably a drastic reduction in storage space. By down-sampling the data, the number of stored data points is significantly reduced, which is a major advantage for long-duration simulations that would otherwise generate massive, unwieldy datasets. This gain in storage efficiency is achieved without sacrificing the crucial details of the system's dynamic behavior,



as evidenced by the consistent results between the original and decimated plots. The marginal differences observed are minimal and do not impact the core analysis. Ultimately, this comprehensive comparison reinforces the robustness and reliability of the proposed decimation method. It demonstrates that the technique is effective for capturing the essence of the signals across various scenarios, including dynamic fault conditions, while simultaneously providing a highly efficient solution for data management.

TABLE 2. COMPARISON OF DATA SIZES FOR SIMULATION VARIABLES WITH AND WITHOUT DECIMATION.

Variables	Size without decimation	Size with decimation
The load power	2001340	2002
The PV power	2001340	401
The battery power	2001340	8006
The DG power	2001340	401
The battery's current	2000001	2001
The battery's SOC	2001340	201
The direct current Id	2001340	801
The quadratic current Iq	2001340	1889

The addition of Table 2 provides crucial quantitative evidence of the decimation's effectiveness. It clearly shows the dramatic reduction in data size for each variable, moving from millions of data points to just thousands or hundreds. This translates directly to significant savings in storage space and improved processing efficiency. For instance, as shown in Table 1, the size of the PV power data is reduced by a factor of nearly 5,000, from 2,001,340 data points to just 401. This tangible data reinforces that decimation is not just a theoretical concept but a highly practical and beneficial approach for managing large-scale simulation data.

G. Conclusion

In this paper, we successfully addressed the critical challenge of optimizing data storage without compromising the integrity of results in a multi-source system simulation. By leveraging the Fast Fourier Transform (FFT) for frequency analysis, we developed and validated a decimation technique that dramatically reduces data size. Our comprehensive analysis, supported by visual evidence and quantitative data, confirms that this approach effectively preserves the essential characteristics of all signals, including transient events such as PV panel faults and the corresponding dynamic response of the battery. The near-zero



discrepancies between the original and decimated waveforms demonstrate the robustness and accuracy of our methodology.

Acknowledgment

The authors extend their appreciation to the Northern Border University, Saudi Arabia for supporting this work through project number "NBU-CRP-2025-2448".

References

- [1] N. Fesakis, G. Falekas, I. Palaiologou, G. E. Lazaridou, and A. Karlis, "Integration and optimization of multisource electric vehicles: A critical review of hybrid energy systems, topologies, and control algorithms," *Energies (Basel)*, vol. 17, no. 17, p. 4364, 2024.
- [2] A. Q. Al-Shetwi, I. E. Atawi, M. A. El-Hameed, and A. Abuelrub, "Digital Twin Technology for Renewable Energy, Smart Grids, Energy Storage and Vehicle-to-Grid Integration: Advancements, Applications, Key Players, Challenges and Future Perspectives in Modernising Sustainable Grids," *IET Smart Grid*, vol. 8, no. 1, p. e70026, 2025.
- [3] A. Khan, M. Bressel, A. Davigny, D. Abbes, and B. Ould Bouamama, "Comprehensive Review of Hybrid Energy Systems: Challenges, Applications, and Optimization Strategies," *Energies (Basel)*, vol. 18, no. 10, p. 2612, 2025.
- [4] A. Khan, M. Bressel, A. Davigny, D. Abbes, and B. Ould Bouamama, "Comprehensive Review of Hybrid Energy Systems: Challenges, Applications, and Optimization Strategies," *Energies (Basel)*, vol. 18, no. 10, p. 2612, 2025.
- [5] S. Verma, Y. L. Kameswari, and S. Kumar, "A review on environmental parameters monitoring systems for power generation estimation from renewable energy systems," *Bionanoscience*, vol. 14, no. 4, pp. 3864–3888, 2024.
- [6] S. Kryltcov, A. Makhovikov, and M. Korobitcyna, "Novel approach to collect and process power quality data in medium-voltage distribution grids," *Symmetry (Basel)*, vol. 13, no. 3, p. 460, 2021.
- [7] R. K. Beniwal, M. K. Saini, A. Nayyar, B. Qureshi, and A. Aggarwal, "A critical analysis of methodologies for detection and classification of power quality events in smart grid," *IEEE Access*, vol. 9, pp. 83507–83534, 2021.
- [8] A. Swain, E. Abdellatif, A. Mousa, and P. W. T. Pong, "Sensor technologies for transmission and distribution systems: A review of the latest developments," *Energies (Basel)*, vol. 15, no. 19, p. 7339, 2022.



- [9] N. Okafor, "Advances and Challenges in IoT Sensors Data Handling and Processing in Environmental Monitoring Systems," Authorea Preprints, 2023.
- [10] M. Trigka and E. Dritsas, "Wireless Sensor Networks: From Fundamentals and Applications to Innovations and Future Trends," IEEE Access, 2025.
- [11] S. Ismail, D. W. Dawoud, and H. Reza, "Securing wireless sensor networks using machine learning and blockchain: A review," Future Internet, vol. 15, no. 6, p. 200, 2023.
- [12] S. Salmi and L. Oughdir, "Cnn-lstm based approach for dos attacks detection in wireless sensor networks," International Journal of Advanced Computer Science and Applications, vol. 13, no. 4, 2022.
- [13] Z. N. Al-kateeb and D. B. Abdullah, "Unlocking the potential: synergizing IoT, cloud computing, and big data for a bright future," Iraqi Journal for Computer Science and Mathematics, vol. 5, no. 3, p. 25, 2024.
- [14] E. Cofas, "The role of big data in digitalizing information.," Scientific Papers Series Management, Economic Engineering in Agriculture & Rural Development, vol. 23, no. 3, 2023.
- [15] S. Gathu, "High-performance computing and big data: Emerging trends in advanced computing systems for data-intensive applications," Journal of Advanced Computing Systems, vol. 4, no. 8, pp. 22–35, 2024.
- [16] T. Dubuc, F. Stahl, and E. B. Roesch, "Mapping the big data landscape: technologies, platforms and paradigms for real-time analytics of data streams," IEEE Access, vol. 9, pp. 15351–15374, 2020.
- [17] T. T. Khoei and A. Singh, "Data reduction in big data: a survey of methods, challenges and future directions," Int J Data Sci Anal, pp. 1–40, 2024.
- [18] S. Jebreili and A. Goli, "Optimization and computing using intelligent data-driven," Optimization and computing using intelligent data-driven approaches for decision-making: optimization applications, vol. 90, no. 4, 2024.
- [19] B. Wang, "A Comprehensive Review of Optimal Power Flow in Integrated Energy Systems: A Shift from Traditional to Data-Driven Technologies," 2025.
- [20] S. Selvarajan, "A comprehensive study on modern optimization techniques for engineering applications," Artif Intell Rev, vol. 57, no. 8, p. 194, 2024.
- [21] Y. Guo, S. Ganti, and Y. Wu, "Enhancing energy efficiency in telehealth internet of things systems through fog and cloud computing integration: simulation study," JMIR Biomed Eng, vol. 9, no. 1, p. e50175, 2024.



- [22] A. Ramachandran, “Powering Intelligence The Future of AI Hardware for Training, Inference, and Innovation,” Jan. 2025.
 - [23] D. Patil, N. Rane, P. Desai, and J. Rane, “Machine learning and deep learning: Methods, techniques, applications, challenges, and future research opportunities,” 2024, pp. 28–81.
 - [24] M. Nabeel, “AI-Enhanced Project Management Systems for Optimizing Resource Allocation and Risk Mitigation: Leveraging Big Data Analysis to Predict Project Outcomes and Improve Decision-Making Processes in Complex Projects,” *Asian Journal of Multidisciplinary Research & Review*, vol. 5, pp. 53–91, Oct. 2024.
 - [25] M. Patricia and M. Alhassan, *Optimizing Energy Consumption in AI-Driven Financial Analytics: A Machine Learning Approach for Enterprise Intelligent Finance*. 2024.
 - [26] A. Eaturu and K. P. Vadrevu, “Evaluation of machine learning and deep learning algorithms for fire prediction in Southeast Asia,” *Sci Rep*, vol. 15, no. 1, p. 18807, 2025.
 - [27] J. Perera-Lago, V. Toscano-Duran, E. Paluzo-Hidalgo, R. Gonzalez-Diaz, M. A. Gutiérrez-Naranjo, and M. Rucco, “An in-depth analysis of data reduction methods for sustainable deep learning,” *Open Research Europe*, vol. 4, p. 101, 2024.
 - [28] K. Jayasooriya et al., *A new compression strategy to reduce the size of nanopore sequencing data*. 2024.
 - [29] O. Mutlu and C. Firtina, *Invited: Accelerating Genome Analysis via Algorithm-Architecture Co-Design*. 2023.
 - [30] H. Gamaarachchi et al., “GPU accelerated adaptive banded event alignment for rapid comparative nanopore signal analysis,” *BMC Bioinformatics*, vol. 21, Aug. 2020.
 - [31] H. Samarakoon, J. Ferguson, H. Gamaarachchi, and I. Deveson, *Accelerated nanopore basecalling with SLOW5 data format*. 2023.
 - [32] D. Senol Cali, J. Kim, S. Ghose, C. Alkan, and O. Mutlu, “Nanopore Sequencing Technology and Tools for Genome Assembly: Computational Analysis of the Current State, Bottlenecks, and Future Directions,” *Brief Bioinform*, Nov. 2017.
 - [33] K. Jayasooriya et al., “A new compression strategy to reduce the size of nanopore sequencing data,” *Genome Res*, vol. 35, no. 7, pp. 1574–1582, 2025.
- R. Xiang, W. Wang, L. Yang, S. Wang, C. Xu, and X. Chen, “A comparison for dimensionality reduction methods of single-cell RNA-seq data,” *Front Genet*, vol. 12, p. 646936, 2021.



- [34] V. Fernandes, G. Carvalho, V. Pereira, and J. Bernardino, “Analyzing data reduction techniques: an experimental perspective,” *Applied Sciences*, vol. 14, no. 8, p. 3436, 2024.
- [35] W. Jia, M. Sun, J. Lian, and S. Hou, “Feature dimensionality reduction: a review,” *Complex & Intelligent Systems*, vol. 8, no. 3, pp. 2663–2693, 2022.
- [36] T. T. Khoei and A. Singh, “Data reduction in big data: a survey of methods, challenges and future directions,” *Int J Data Sci Anal*, pp. 1–40, 2024.
- [37] E. Chris, A. Qudus, and K. Sherifdeen, “The Evolution of Volumetric Video: A Comprehensive Survey of Smart Transcoding and Compression Techniques,” Dec. 2024.
- [38] T. Alghamdi and N. Javaid, “A Survey of Preprocessing Methods Used for Analysis of Big Data Originated From Smart Grids,” *IEEE Access*, vol. 10, Feb. 2022.
- [39] M. Hosseini, D. Pratas, and A. Pinho, “A Survey on Data Compression Methods for Biological Sequences,” *Information*, vol. 7, p. 56, Oct. 2016.
- [40] M. Kaur, “A Comprehensive Survey on Architecture for Big Data Processing in Mobile Edge Computing Environments: From Hype to Reality,” 2019, pp. 33–49.
- [41] A. Uzzaman, M. M. I. Jim, N. Nishat, and J. Nahar, “Optimizing SQL databases for big data workloads: techniques and best practices,” *Academic Journal on Business Administration, Innovation & Sustainability*, vol. 4, no. 3, pp. 15–29, 2024.
- [42] Y. Silva, I. Almeida, and M. Queiroz, *SQL: From Traditional Databases to Big Data*. 2016.
- [43] V. Panwar, “Optimizing Big Data Processing in SQL Server through Advanced Utilization of Stored Procedures,” *International Journal of Management IT and Engineering*, vol. 14, pp. 60–70, Feb. 2024.
- [44] B. Nevarez, *High Performance SQL Server: Consistent Response for Mission-Critical Applications*. 2021.
- [45] M. M. Rahman, S. Islam, M. Kamruzzaman, and Z. H. Joy, “advanced query optimization in sql databases for real-time big data analytics,” *academic journal on business administration, innovation & sustainability*, vol. 4, p. 1, Jun. 2024.
- [46] A. Uzzaman, M. M. I. Jim, N. Nishat, and J. Nahar, “optimizing sql databases for big data workloads: techniques and best practices,” *academic journal on business administration, innovation & sustainability*, vol. 4, pp. 15–29, Jun. 2024.
- [47] H. G. Kola and A. Taqa, “Available online,” vol. 11, pp. 2853–3006, Nov. 2023.



- [48] S. Altinişik and T. Bilgin, “Optimizing Big Data Management on Microsoft SQL Server: Enhancing Performance through Normalization and Advanced Analytical Techniques,” *International Journal of Innovative Engineering Applications*, vol. 9, pp. 23–36, Jun. 2025.
- [49] M. H. Haron et al., “Image data compression using fast Fourier transform (FFT) technique for wireless sensor network,” in *AIP Conference Proceedings*, AIP Publishing LLC, 2024, p. 030032.
- [50] M. Abbaci and M. O. Zmirli, “Cooley-Tukey FFT Algorithm based on GDFT for Phasor Estimation by PMU under Power Quality Disturbances,” *Engineering, Technology & Applied Science Research*, vol. 13, no. 1, pp. 9906–9912, 2023.
- [51] S. Hu et al., “Multistatic MIMO sparse imaging based on FFT and low-rank matrix recovery techniques,” *IEEE Trans Microw Theory Tech*, vol. 71, no. 3, pp. 1285–1295, 2022.
- [52] X. Song, Z. Li, and Y. Liu, “MVB fault diagnosis based on time-frequency analysis and convolutional neural networks,” *Sci Rep*, vol. 15, no. 1, p. 5271, 2025.
- [53] Z. Hao, Y. Wang, F. Li, G. Ding, K. Fan, and Y. Gao, “Detection of vital signs based on millimeter wave radar,” *Sci Rep*, vol. 15, no. 1, p. 28112, 2025.
- [54] Rustamaji, S. Kliwati, and W. Widada, “The angle of arrival estimation of frequency-hopping cooperative object based on software-defined radio,” *Sci Rep*, vol. 14, no. 1, p. 7732, 2024.
- [55] M. F. Wahab, F. Gritti, and T. C. O’Haver, “Discrete Fourier transform techniques for noise reduction and digital enhancement of analytical signals,” *TrAC Trends in Analytical Chemistry*, vol. 143, p. 116354, 2021.
- [56] M. Ravi, “Complex Analysis and Fourier Transforms in signal processing,” *Mathematical Innovation*, p. 64, 2025.
- [57] R. Xiang, W. Wang, L. Yang, S. Wang, C. Xu, and X. Chen, “A comparison for dimensionality reduction methods of single-cell RNA-seq data,” *Front Genet*, vol. 12, p. 646936, 2021.
- [58] Y. Zhou, Z. Ma, and L. Fu, “A review of key signal processing techniques for structural health monitoring: Highlighting non-parametric time-frequency analysis, adaptive decomposition, and deconvolution,” *Algorithms*, vol. 18, no. 6, p. 318, 2025.
- [59] T. Shevgunov, E. Efimov, and O. Guschina, “Estimation of a Spectral Correlation Function Using a Time-Smoothing Cyclic Periodogram and FFT Interpolation—2N-FFT Algorithm,” *Sensors*, vol. 23, no. 1, p. 215, 2022.



- [60] M.-A. Delsuc and P. O’connor, “The Fourier transform in analytical science,” *Nature Reviews Methods Primers*, vol. 4, no. 1, p. 49, 2024.
- [61] J. L. J. Pereira, G. A. Oliver, M. B. Francisco, S. S. Cunha Jr, and G. F. Gomes, “A review of multi-objective optimization: methods and algorithms in mechanical engineering problems,” *Archives of Computational Methods in Engineering*, vol. 29, no. 4, pp. 2285–2308, 2022.
- [62] Q. Li, D. Zuo, Y. Feng, and D. Wen, “Research on High-Performance Fourier Transform Algorithms Based on the NPU,” *Applied Sciences*, vol. 14, no. 1, p. 405, 2024.
- [63] F. Peres and M. Castelli, “Combinatorial optimization problems and metaheuristics: Review, challenges, design, and development,” *Applied sciences*, vol. 11, no. 14, p. 6449, 2021.
- [64] X. Fang, J. Zheng, and B. Jiang, “A rolling bearing fault diagnosis method based on vibro-acoustic data fusion and fast Fourier transform (FFT),” *Int J Data Sci Anal*, pp. 1–10, 2024.
- [65] M. H. Salem, K. Mansouri, E. Cauveau, Y. Bensalem, and M. N. Abdelkrim, “Simulation of an Energy Management System and Control in a Multi-Source System,” in *2023 IEEE International Workshop on Mechatronic Systems Supervision (IW_MSS)*, IEEE, 2023, pp. 1–8.
- [66] M. Hadj Salem, K. Mansouri, E. Chauveau, Y. Ben Salem, and M. N. Abdelkrim, “Multi-Power System Electrical Source Fault Review,” *Energies (Basel)*, vol. 17, no. 5, p. 1187, 2024.