D4.4 Optimal selection of available flexibility v2

30 October 2025

OPENTUNITYproject.eu

Deliverable details

Title	WP	Version
Optimal selection of available flexibility v2	4	V5

Contractual delivery date	Actual delivery date	Delivery type*	Dissemination**
M34 (October 2025)	M34 (October 2025)	R	PU

*Delivery type: R: Document, report; DEM: Demonstrator, pilot, prototype; DEC: Websites, patent fillings, videos, etc; OTHER; ETHICS: Ethics requirement; ORDP: Open Research Data Pilot.
Dissemination Level: **PU: Public; **CO**: Confidential, only for members of the consortium (including the Commission Services)

Author(s)	Organization
Janez Gregor Golja	UL
Klemen Peter Kosovinc	SETUP
Damir Omrčen	Amibit

Version	Date	Person	Action	Status***
V0	1.7.2025	Janez Gregor Golja (UL)	ToC created	
V1	9.9.2025	Janez Gregor Golja (UL), Klemen Peter Kosovinc (KOL)	Added contributions from KOL	
V3	7.10.2025	Janez Gregor Golja (UL)	Document ready for peer-review	
V4	13.10.2025	Damir Omrčen (Amibit), Janez Gregor Golja (UL)	Integrated Amibit contribution	
V5	29.10.2025	Janez Gregor Golja (UL)	Final version ready for submission	

***Status: Draft, Final, Approved, Submitted (to European Commission).

Authors (organization)
Janez Gregor Golja (UL), Klemen Peter Kosovinc (SETUP), Damir Omrčen (Amibit), Miha Valentinčič (AVA), Matej Pečjak (IRI UL), Tomi Medved (UL), John Karakitsios (ICCS)

Keywords

Flexibility, aggregator, optimal selection, economic optimization

Executive Summary

This deliverable presents the second phase of the innovation activity **"Optimal Selection of Available Flexibility"**, developed within Task 4.4 of the OPENTUNITY project. It builds on the foundations established in the first phase, where data from electric vehicle (EV) charging stations and Home Energy Management Systems (HEMS) were analyzed, integrated into the KOL aggregation platform, and used to develop advanced baseline and flexibility forecasting models. These outputs now serve as key inputs for the work presented here.

The main goal of this task is to support **aggregators and flexibility service providers (FSPs)** in making optimal decisions on *where* to offer flexibility and *which assets* to activate for delivery. Accordingly, the deliverable covers two complementary algorithms:

- **Market Selection Algorithm**, which identifies the most profitable flexibility markets (day-ahead, intraday, TSO ancillary, or DSO local) considering technical and regulatory constraints and using forecasted consumption, flexibility potential, and price signals. It compares different participation strategies and demonstrates that combining day-ahead optimization with ancillary services participation can significantly increase revenues while maintaining operational feasibility. It also includes a brief module on implicit flexibility: a household-level tariff optimization case that shows how end-users can manage loads under the new Slovenian network tariff system.
- **Optimal Selection Algorithm**, which determines the least-cost feasible combination of assets to activate in response to flexibility requests, given available resources and constraints. Implemented via a genetic-algorithm framework, it selects the most cost-efficient asset set while satisfying technical and operational requirements. The framework is modular and adaptable to evolving market settings or new asset types.

Together, these components form a **comprehensive flexibility-management framework** — spanning the identification of profitable market opportunities, activation of distributed resources in a cost-efficient manner, and enabling household-level optimization under evolving tariff schemes.

Copyright statement

Index

1 INTRODUCTION 7

1.1 Purpose of the document 7

1.2 Scope of the document 7

1.3 Structure of the document 7

2 OPTIMAL MARKET SELECTION ALGORITHM 8

2.1 Market possibilities 8

- 2.1.1 Electricity price optimization 9
  - 2.1.1.1 Barriers and opportunities 9
- 2.1.2 TSO ancillary services markets 9
  - 2.1.2.1 Barriers and opportunities 9
- 2.1.3 DSO local flexibility markets 10
  - 2.1.3.1 Barriers and opportunities 10
- 2.1.4 Network tariff optimization 10
  - 2.1.4.1 Barriers and opportunities 11

2.2 Optimal market selection algorithm description – high level 11

2.3 Algorithm description and market gains calculation – granular level 14

2.4 Implicit flexibility optimization based on grid tariff 20

- 2.4.1 Context and motivation 20
- 2.4.2 Methodological concept 20
- 2.4.3 Algorithm testing approach and expected outcomes 22

3 OPTIMAL SELECTION OF AVAILABLE FLEXIBILITY 24

3.1 Background and literature review 24

- 3.1.1 Literature review 25
- 3.1.2 Genetic algorithms 26

3.2 Optimal selection algorithm design 27

- 3.2.1 Problem definition and objectives 27
- 3.2.2 High-level algorithm overview 28

3.2.3 Cost modelling 29
3.2.4 Per-asset models (technical constraints and SRMC) 31
3.2.4.1 EV chargers 31
3.2.4.2 Battery Energy Storage Systems (BESS) 32
3.2.4.3 PV generation 33
3.2.4.4 HVAC systems 34
3.3 Algorithm implementation 37
3.3.1 Practical example of algorithm operation 39
4 CONCLUSIONS 46
5 REFERENCES AND ACRONYMS 47
5.1 References 47
5.2 Acronyms 48

List of figures

Figure 1: High-level perspective of the optimal market selection algorithm. 12
Figure 2: Example of the DA price forecast. 13
Figure 3: Optimization results for an average day at different shares of maximum power. 16
Figure 4: Optimization results for an average day at different shares of maximum power. Added limitation of charging to a maximum of 4 hours during work and 4 hours during night. 17
Figure 5: Optimization results for scenarios of DA and DA & ancillary services markets combination. 19
Figure 6: Tariff optimization logic representation 21
Figure 7: Building thermal behavior modelling concept 22
Figure 6: High-level GA workflow 28
Figure 7: Algorithm results - signed stacked activations by asset 42
Figure 8: BESS activations and SoC evolution 42
Figure 9: HVAC baseline, setpoint, flexibility, and indoor temperature 43
Figure 10: SRMC values for upward and downward flexibility by asset 44

List of tables

Table 1: Optimization results for different capacity utilization levels, DA market. 15
Table 2: Optimization results for different capacity utilization levels, DA market. Added limitation of charging to a maximum of 4 hours during work and 4 hours during night 17

Table 3: Optimization results for different capacity utilization levels, DA market and ancillary services. Added limitation of charging to a maximum of 4 hours during work and 4 hours during night. 19
Table 4: Literature review findings 25
Table 5: Practical example, SRMC parameters for EV asset 39
Table 6: Practical example, SRMC parameters for PV asset 39
Table 7: Practical example, SRMC parameters for BESS asset 40
Table 8: Practical example, SRMC parameters for HVAC asset 40
Table 9: Acronyms 48

1 Introduction

1.1 Purpose of the document

This deliverable represents the second iteration of the report prepared under Task 4.4: Optimal Selection of Available Flexibility within the OPENTUNITY project. It builds on the outcomes of the first phase, which focused on analysing real data from EV charging stations and HEMS devices to develop baseline and flexibility forecasting models. The second phase extends this work by building upon the established forecasting and data integration framework to develop and implement two algorithms that address different stages of the aggregator's and FSP's workflow. The first is the Market Selection Algorithm, which identifies which flexibility markets an aggregator should participate in, based on forecasted consumption, flexibility potential, price signals, and technical or regulatory constraints. Within this framework, a complementary tariff-optimization module is also presented, demonstrating how similar optimization principles can be applied at household level to reduce network-tariff costs using the Reduxi HEMS system. The second algorithm is the Optimal Selection Algorithm, which focuses on the internal decision-making of the aggregator during flexibility activation, determining the least-cost combination of assets to deliver the requested flexibility while respecting technical and operational limits.

1.2 Scope of the document

The document describes the approach adopted in the second phase of Task 4.4, detailing the theoretical formulation, design, and initial testing of both algorithms. For the Optimal Market Selection Algorithm, the document presents the methodology and the results obtained using real data from AVANTCAR's EV charging stations. In the same section, the tariff-optimization module that addresses implicit flexibility at the household level is presented. For the Optimal Selection Algorithm, it introduces the algorithmic framework and provides a demonstration example using a small, simplified portfolio to showcase its operation and internal decision logic.

1.3 Structure of the document

The document is structured in the following sections:

- Section 2 provides an overview of the Optimal Market Selection Algorithm, describing the main market opportunities available to aggregators and FSPs and the methodology used to assess their profitability. It includes the theoretical formulation of the algorithm and the results obtained from its application using real data from AVANTCAR's EV charging stations. It also includes a complementary tariff-optimization sub-algorithm, which applies the same optimization principles at the household level through the Reduxi HEMS platform under Slovenia's dynamic network-tariff system.
- Section 3 focuses on the Optimal Selection Algorithm, which supports the decision-making process of an aggregator during flexibility activation. It presents the background and literature review, the design and implementation of the algorithm, and the modelling of both technical constraints and cost components for different asset types. The section concludes with a demonstration case illustrating how the algorithm operates on a simplified portfolio.

2 Optimal market selection algorithm

The flexibility of energy assets can be offered across a range of energy markets, each ruled by technical, regulatory, and economic requirements. To navigate this complexity, aggregators or flexibility service providers (FSPs) must assess how the specific characteristics and limitations of their assets align with the conditions of each market.

In Deliverable D4.3, we addressed two foundational aspects of this assessment: First, we focused on baseline forecasting of the assets. This is essential because both system operators and market platforms depend on accurate baseline estimates to validate whether the contracted or activated amount of flexibility was or will be delivered as agreed. If the baseline forecast is not calculated accurately, the FSP is penalized either because of not delivered energy in the case of ancillary services and local distribution system operator (DSO) markets or in the form of imbalance in the case of day-ahead (DA) and intraday (ID) markets.

Second, we have developed models for flexibility forecasting, which estimate the potential power and energy that can be offered across different markets. These forecasts are based on an evaluation of the assets' technical feasibility, considering factors such as response time, activation duration, availability windows, and operational constraints.

Both baseline and flexibility forecasting are essential for the development of the optimal market selection algorithm. This algorithm is designed to determine, in advance, the most advantageous combination of markets for each flexibility asset—guiding decisions on where and how to offer or place bids for flexibility in order to maximize the value of the asset or portfolio of assets. The algorithm estimates the relative value of different market opportunities—such as DA, ID, transmission system operator (TSO) ancillary services like manual frequency restoration reserve (mFRR), local DSO markets or network tariffs — and suggests how flexibility should be allocated across them.

For example, offering 1 MW of flexibility in the form of a capacity bid exclusively to the mFRR market and waiting all day for activation may result in underutilization and lost revenue opportunities. Instead, it could be more profitable to place bids in the DA market, adjust in the ID market based on updated forecasts and prices, and only offer flexibility to mFRR as energy bids. The algorithm aims to automate and optimize such strategic decisions, maximizing revenue while respecting technical constraints and market rules.

Therefore, in this Chapter, we will introduce different markets, which were reviewed and considered for market selection in the scope of OPENTUNITY. Furthermore, we will describe the optimal market selection algorithm and provide an explanation of the demo use case, along with actual results of the algorithm for the electric vehicle (EV) charging points fleet of AVANTCAR.

2.1 Market possibilities

The following sub-sections provide an overview of the primary flexibility utilization options available to FSPs. These include TSO-level ancillary services for frequency regulation, DSO-operated local flexibility markets and optimization strategies in the day-ahead and intraday markets, as well as network tariff-based flexibility.

2.1.1 Electricity price optimization

Electricity price optimization for end users—such as AVANTCAR's charging stations — can be achieved by strategically participating in the DA and ID markets. The approach involves scheduling asset operation based on DA price forecasts to minimize energy costs, while leveraging flexibility to exploit extreme market signals (very high or low prices in ID market) to further reduce the energy costs. Furthermore, adjustments in the ID market are relevant, where trades closer to real time help reduce imbalance costs. Additionally, analyzing the system's expected power balance enables smarter decisions in the ID market, for instance, not selling surplus energy even when prices are high due to anticipated system shortages and due to the forecast indicating that imbalance costs will be even higher.

2.1.1.1 Barriers and opportunities

Barriers:

- Baseline, flexibility and price forecast inaccuracy.
- Auction and gate-closure deadlines (e.g., more than 12 hours for DA, one hour before delivery for ID).
- Minimum trade sizes and market liquidity constraints.
- Data and computation requirements for real-time optimization.

Opportunities:

- Revenue from price arbitrage (charging in low-price periods, discharging/selling in high-price periods).
- Reduced imbalance costs by closely matching DA forecast with actual consumption.

2.1.2 TSO ancillary services markets

The TSO ancillary services market enables system operators to procure services like automatic frequency restoration reserve (aFRR) and mFRR to maintain grid stability, with flexibility offered through either capacity or energy bids. Avantcar's EV charging stations and Amibit's HEMS devices could participate in these services, particularly through energy bids, which allow targeted offering during specific hours when flexibility is available. However, a key barrier is the minimum bid size, for example 1 MW in most European countries including Slovenia, Spain and Greece and for example, 5 MW in Switzerland. This bid size exceeds the flexibility capacity of individual portfolios—necessitating aggregation of multiple assets. While direct participation is not yet feasible (flexibility power of both portfolios has not reached 1 MW yet. For example, turning off all AVANTCAR charging stations could provide up to 200 kW), future growth in charging infrastructure and HEMS devices could open valuable opportunities for market entry.

2.1.2.1 Barriers and opportunities

Barriers:

- High minimum bid sizes (e.g., 1 MW).
- Fast activation times and performance requirements (e.g. 12,5-min in mFRR and 5-min in aFRR).

- Strict metering and communication standards.
- Penalties for non-delivery.

Opportunities:

- Capacity payments for standing reserve bids.
- Energy payments when actually activated.
- Enhanced system reliability and potential premium prices during scarcity.

2.1.3 DSO local flexibility markets

Local flexibility markets are emerging as a potential cost-effective solution for DSOs to manage grid issues caused by the rising integration of distributed energy resources (DERs), such as photovoltaics (PVs), heat pumps (HPs), and EVs. Unlike traditional infrastructure upgrades, these markets allow DSOs to procure location-specific flexibility from prosumers or aggregators to resolve local congestion and voltage problems. Supported by EU legislation and upcoming regulatory frameworks, they offer lower participation thresholds compared to TSO-level ancillary services, making them accessible to smaller assets. However, their broader adoption is still hindered by regulatory uncertainty, limited standardization, and the need for real-time data exchange and responsiveness from participating assets.

2.1.3.1 Barriers and opportunities

Barriers:

- Regulatory uncertainty and differing national/local rules.
- Definition of geographic flexibility zones and asset registration processes.
- Limited market liquidity and pilot-phase participation.
- Communication protocols and real-time settlement requirements.

Opportunities:

- Capacity payments for standing reserve bids.
- Energy payments when actually activated.
- Engagement of small-scale assets that cannot meet TSO bid sizes.

2.1.4 Network tariff optimization

Electricity costs for consumers consist of supply charges, network fees, and regulatory charges, with network tariffs covering the cost of using the grid. Since October 2024, Slovenia has adopted a dynamic tariff model [1] that charges based on the highest 15-minute average power demand, introducing five time-based tariffs that can increase costs for users exceeding their limits. Traditionally, in other EU countries, for example, Austria, the network tariff consists of a fixed fee for connected power and different fees for energy – summer (from April till October) and winter (from November till March) high (from 06:00 till 22:00) and low tariff (from 22:00 till 06:00). For example, in new tariff system in Slovenia, penalties are determined if the peak power exceeds the defined power at the start of each month. The new network tariff system introduces 15-minute interval metering with seasonal and time-of-day pricing, emphasizing contracted power levels and applying surcharges for exceeding them, thereby shifting the focus from total energy use to managing peak demand. Optimal utilization of the flexibility of an asset can lower network tariff costs if high peaks can be avoided in the expensive time blocks. More detailed description of the new tariff system in Slovenia can be found in section 2.4, which presents a solution co-developed in OPENTUNITY, which utilizes the flexibility of heat pumps to lower the consumer's tariff costs.

2.1.4.1 Barriers and opportunities

Barriers (relevant for Slovenia):

- Regulatory uncertainty and differing national/local rules (e.g. Slovenia adopted the dynamic model and changed it after a few months of operation. A New change is proposed for 2026).
- The defined power needs to be reported one month in advance, when the baseline and flexibility forecasts are still uncertain.

Opportunities:

- Reduced network tariff costs.
- Engagement of small-scale assets that cannot meet TSO or DSO local market bid sizes.

2.2 Optimal market selection algorithm description – high level

High-level perspective of the optimal market selection algorithm is presented in Figure 1. To initiate the optimal market selection process, the algorithm requires comprehensive asset information as a fundamental input. As outlined earlier—and detailed in Deliverable D4.3—this includes both the baseline forecast (the asset's expected consumption or generation without external signals) and the flexibility forecast (the potential deviation from the baseline that can be offered as flexibility). However, this information alone is not sufficient. The algorithm also takes into account a set of technical, operational, and security constraints, which play a critical role in determining market eligibility.

These constraints define the physical and operational boundaries of the asset and directly influence where and how flexibility can be offered. For instance, an asset with a ramp-up time of 20 minutes would be too slow to respond to activation signals in fast-response ancillary services like mFRR or aFRR, which typically require activation within 12.5 and 5 minutes, respectively. Similarly, an asset with a minimum continuous runtime of 12 hours may lack the operational flexibility required for participation in short-duration balancing markets, where activation windows may last only a few minutes. Conversely, assets with a ramp-up time of less than 30 seconds and no restrictive runtime constraints would be technically capable of participating in nearly all relevant markets, including the most demanding ancillary service products.

By integrating these parameters—baseline behavior, available flexibility, and technical limitations—the algorithm performs a pre-screening step to determine the initial set of markets where each asset could potentially participate. This mapping is essential for ensuring that the asset is not only economically viable but also technically compliant with the rules and requirements of each market. These carefully evaluated constraints serve as the foundation for the optimal market selection algorithm, enabling it to recommend where and how assets should offer their flexibility to maximize value while remaining within their operational capabilities.

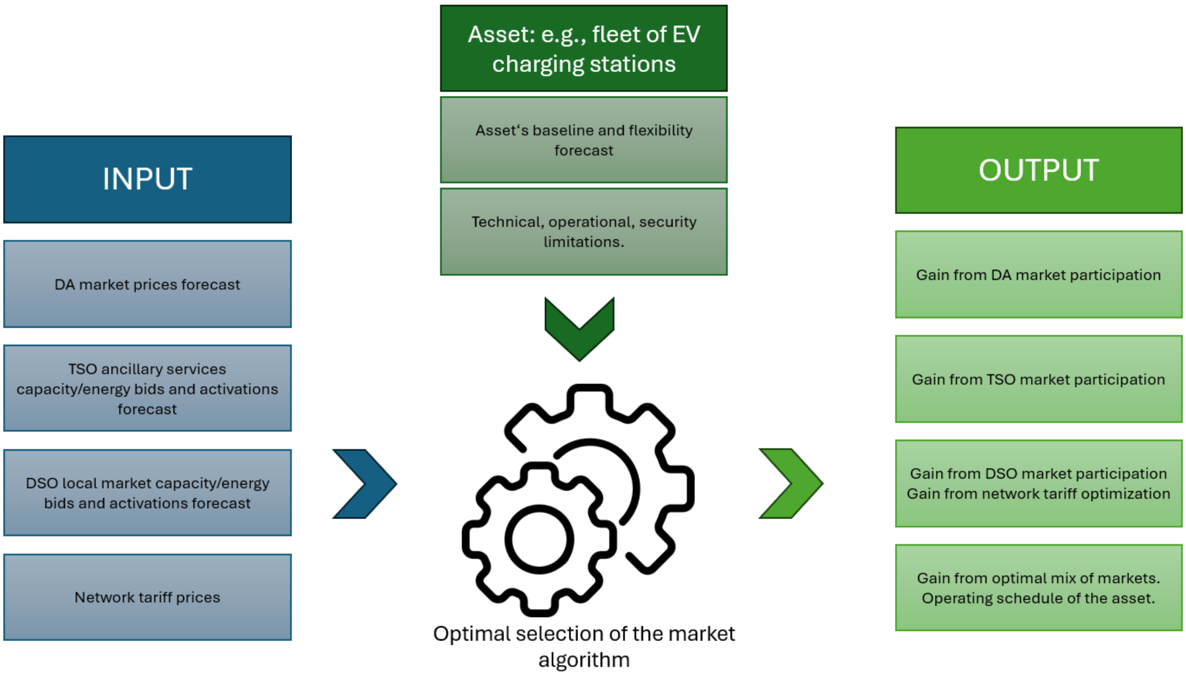

Figure 1: High-level perspective of the optimal market selection algorithm.

The second critical set of inputs to the optimal market selection algorithm pertains to market parameters and price forecasts. These inputs are essential for evaluating the economic potential of each market and estimating how much revenue or cost savings can be achieved by offering flexibility from either individual assets or aggregated portfolios. Among the most fundamental of these inputs is DA market price forecast, which provide anticipated hourly electricity prices for the next day. The relevance of DA market price forecasts lies in the anticipation of the market and, therefore, the calculation of the gain an asset can achieve in the market. The economic gain is the crucial factor for an asset owner to decide whether he is willing to lose the comfort by, for example, not charging at the most convenient times. The prices forecasting enables the algorithm to assess opportunities for cost minimization (in the case of flexible loads) or revenue maximization (in the case of dispatchable assets). An illustration of a two-day DA price forecast is presented in Figure 2, highlighting potential price spreads that can be exploited for energy arbitrage.

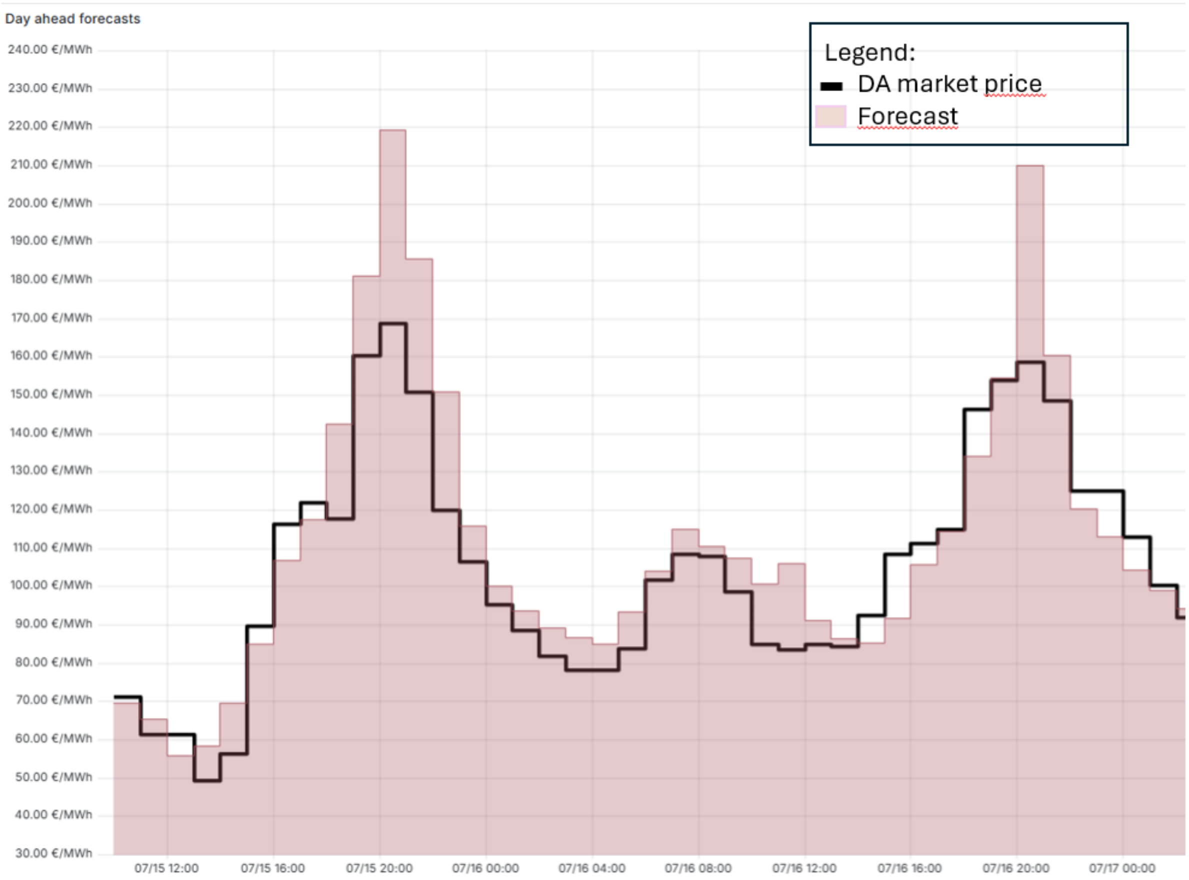

Figure 2: Example of the DA price forecast.

In addition to the DA market, capacity, energy, and activation forecasts are provided for TSO-level ancillary services (such as aFRR and mFRR) and emerging DSO-level local flexibility markets. These forecasts allow the algorithm to anticipate how likely it is that flexibility bids will be accepted and activated, which is especially important when assessing revenue potential from capacity payments (for reserved flexibility) or activation payments (for delivered energy). Moreover, network tariff structures, including both energy-based and power-based charges, are included as inputs. These prices influence the total cost of operating assets and are particularly relevant under dynamic pricing schemes, such as Slovenia's new tariff model introduced in October 2024 [1].

Once all necessary inputs are collected, including baseline forecasts, flexibility profiles, technical constraints, market prices, and tariff information, the algorithm performs the optimization and market evaluation process (described in detail in the next subchapter). This process generates quantitative assessments for each relevant market, providing both standalone and combined market participation strategies. For instance, consider a 1 MW / 2 MWh battery energy storage system (BESS). If the forecasted DA price spread is 120 EUR/MWh and the battery is cycled twice per day, the algorithm identifies a potential gain of 480 EUR per day (2 MWh × 2 cycles × 120 EUR/MWh) through DA market arbitrage.

While DA market gains can be calculated relatively directly, assessing potential revenues in other markets—such as ancillary services or local flexibility schemes—is more complex, requiring probabilistic estimations of activation likelihood, acceptance rates, and revenue caps. The algorithm accounts for these factors and compares all options to determine the most economically attractive strategy for each asset or portfolio.

Ultimately, the algorithm delivers a comprehensive output that includes:

- The optimal market (or combination of markets) for each asset,
- The expected revenue or cost savings for each evaluated option,
- And the optimal operating schedule for the asset, specifying when and how flexibility should be deployed to maximize returns.

By synthesizing technical feasibility with dynamic market and tariff inputs, the algorithm empowers aggregators and flexibility service providers to make informed, data-driven decisions that maximize the value of their energy assets.

2.3 Algorithm description and market gains calculation – granular level

Optimizing the cost of electricity consumption is an example of a linear optimization problem. In our case, we are looking for an optimal consumption vector $\vec{E}$ that minimizes the total cost of consumption across discrete time periods

$$\min C_{\text{consumption}} = \min \vec{p}(t) \cdot \vec{E}(t),$$

where vector $\vec{p}(t)$ represents the electricity price and $\vec{E}(t)$ is the vector of energy consumption at each time period t. The optimization attempts to shift consumption to periods with lower prices to reduce the overall cost while meeting energy requirements.

We introduce a physical restriction of non-negative energy consumption

$$\vec{E}(t) \geq 0 \qquad \forall t \in \{1, 2, \dots T\},$$

while on the other hand, we bound it by the availability of energy assets, in Slovenian pilot case, the charging stations E_{M}

$$\vec{E}(t) \leq \vec{E}_{\mathrm{M}}(t) \qquad \forall t \in \{1, 2, \dots T\}$$

Here E_{M} represents the maximum energy capacity that can be consumed based on the technical limitations and the number of active charging stations in each time interval.

An important constraint ensures that total consumed energy over the optimization period meets the required energy demand. We represent this by the equality constraint

$$A_{\text{eq}} \cdot \vec{E} = \vec{b}_{\text{eq}}$$

where A_{eq} is a constraint matrix that defines how consumption in every time interval relates to total energy requirement, and $\vec{b}_{\text{eq}}$ represents the vector of energy demand requirements that must be satisfied.

The vector $\vec{b}_{\mathrm{eq}}$ can represent different types of constraints depending on the optimization objective. For example, the vector can be the actual energy consumed each day, ensuring that the optimized consumption profile $\vec{E}(t)$ matches specific energy requirements. Alternatively, $\vec{b}_{\mathrm{eq}}$ can represent operational constraints such as the number of hours in a 12-hour time period that the stations can be activated, effectively limiting the operational duration rather than direct energy quantities.

The choice of what $\vec{b}_{\mathrm{eq}}$ represents directly determines the structure of the matrix A_{eq}. When $\vec{b}_{\mathrm{eq}}$ represents energy consumption, A_{eq} typically functions as a summation matrix that aggregates consumption across time periods. However, when $\vec{b}_{\mathrm{eq}}$ represents operational time constraints, the matrix A_{eq} must be modified accordingly to relate energy consumption to operational hours, incorporating factors such as charging rates, station capacity, and the relationship between active time and energy delivered. This flexibility allows the same linear programming framework to accommodate various energy management objectives while maintaining mathematical consistency.

We analyze three distinct optimization scenarios. In the first case we optimize cost by shifting consumption to hours with lower day-ahead prices $\vec{p}_{DA}$. We ensure that the total energy consumed each day remains constant through the equality constraint $A_{\mathrm{eq}} \cdot \vec{E} = \vec{b}_{\mathrm{eq}}$ where $\vec{b}_{\mathrm{eq}}$ represents constant energy consumed each day. The matrix A_{eq} is structured as a block diagonal matrix

$$A_{\mathrm{eq}} = \begin{pmatrix} \vec{d}_1 & 0 & 0 & \cdots & 0 \\ 0 & \vec{d}_2 & 0 & \cdots & 0 \\ 0 & 0 & \vec{d}_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \vec{d}_n \end{pmatrix}$$

where $\vec{d}_i = (1, 1, \dots 1)$ are row vectors with length equal to the number of hours in day i. For a better overview, each day spans from 6AM to 6AM the following day (in the further cases, we will divide the day into working and non-working hours).

In our analysis, we have implemented the presented algorithm using the linprog [2] function from the *scipy.optimize* submodule. We have chosen to limit the time span to 8 months, from January 1st, 2025, to September 1st, 2025. The original full dataset (including data from 2024) creates an optimization problem with thousands of variables and constraints, which can lead to numerical instability, memory constraints, and excessive solution times. Initial attempts to solve this as a single optimization problem frequently failed to converge, exhibited numerical instability, and required long solution times – often without reaching any solution despite feasible solutions probably existing. This challenge is inherent to large-scale linear programming problems, not specific to our implementation. By reducing the timeframe and implementing a chunked optimization strategy (dividing the period into segments), we significantly improve convergence reliability, enable error isolation between chunks, and reduce the computational burden while still capturing representative seasonal variation in energy demand and price patterns. Table 1 presents the optimization results for different capacity utilization levels, showing how varying the upper power boundary from 100% down to 30% of maximum availability affects both the achievable cost savings.

Table 1: Optimization results for different capacity utilization levels, DA market.

Share of max power	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**

Original cost [EUR]	31,637							
Optimized cost [EUR]	21,110	18,318	16,566	15,485	14,728	14,185	13,779	13,452
Absolute difference [EUR]	10,526	13,319	15,070	16,152	16,908	17,452	17,858	18,184
Relative difference [%]	33	42	48	51	54	55	57	58

Figure 3: Optimization results for an average day at different shares of maximum power.

shows energy consumption of AVANTCAR's electric stations portfolio for an average day. Based on the figure, the original consumption of portfolio of charging points is between 15% and 20% of the maximum power. As utilization increases and a larger share of available power is deployed, the load curve becomes more peaked, with the majority of consumption shifting toward midday hours, when electricity prices are lowest.

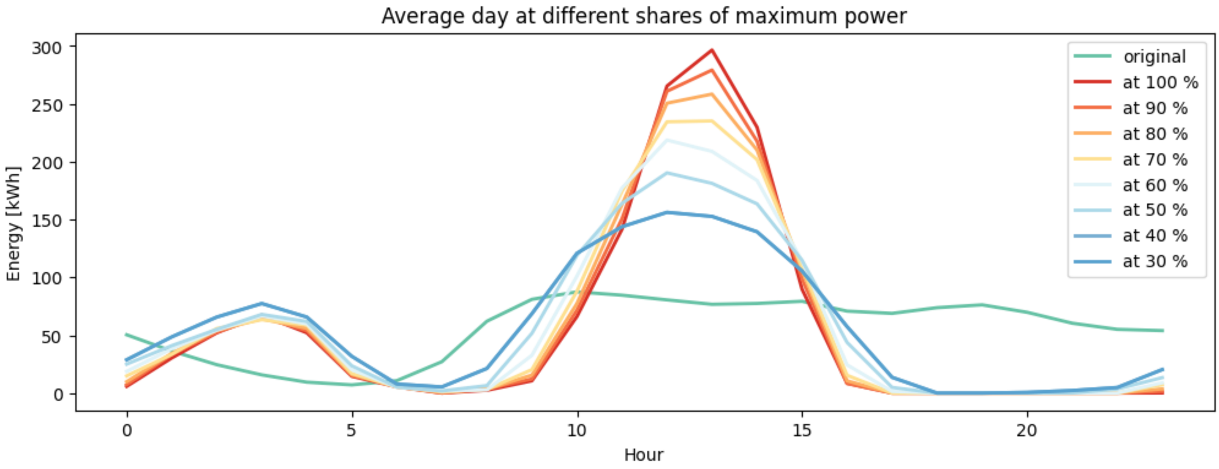

Figure 3: Optimization results for an average day at different shares of maximum power.

Building on top of the first case, we introduce operational restrictions limiting charging station activation to a maximum of 4 hours during the work hours (6 am to 6 pm) and 4 hours during the night hours (6 pm to 6 am). The 4 hours were chosen as this is the duration which should, if empty, fully charge the vehicle in the majority of cases. This is especially relevant for the night hours. This introduces an additional inequality constraint:

$$A_{\mathrm{ub}} \cdot \vec{E} \leq \vec{b}_{\mathrm{ub}},$$

here $\vec{b}_{\mathrm{ub}} = (4, \dots 4)$ with length twice the number of days. The matrix A_{ub} separates day and night periods:

$$A_{\mathrm{ub}} = \begin{pmatrix} \vec{d}_1^1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \vec{d}_1^2 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \vec{d}_2^1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \vec{d}_2^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \vec{d}_n^1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & \vec{d}_n^2 \end{pmatrix},$$

where vectors $\vec{d}_i^{(1,2)}$ correspond to daytime and nighttime hours respectively, with elements equal to the inverse upper limits of available energy $\vec{d}_i^{(1,2)} = {E_{\mathrm{M}}}_i^{-1}$.
Table 2 was made in the same way as in the previous case, but we have only included those percents of total power that were successfully optimized. For example, if only 30% of max power was available, we would not be able to consume the same amount of energy as the fleet did during this same 12-hours period.

Table 2: Optimization results for different capacity utilization levels, DA market. Added limitation of charging to a maximum of 4 hours during work and 4 hours during night

Share of max power	**60%**	**70%**	**80%**	**90%**	**100%**
Original cost [EUR]	31,637				
Optimized cost [EUR]	17,792	16,073	14,830	14,031	13,525
Absolute difference [EUR]	13,844	15,564	16,807	17,606	18,111
Relative difference [%]	44	49	53	56	57

Figure 4 shows energy consumption of AVANTCAR's electric stations portfolio for an average day with additional limitations.

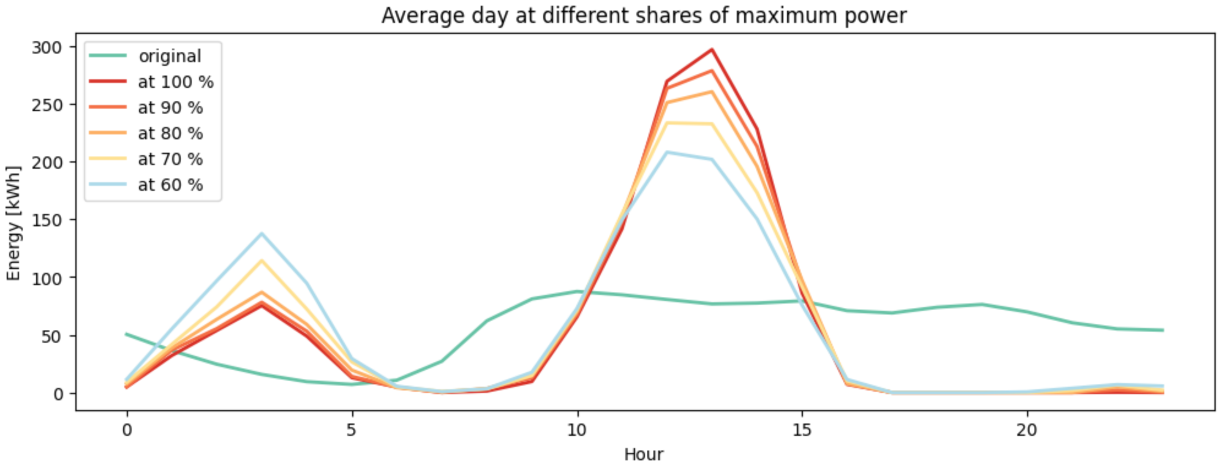

Figure 4: Optimization results for an average day at different shares of maximum power. Added limitation of charging to a maximum of 4 hours during work and 4 hours during night.

As mentioned in the previous section, the third case is not as straightforward as optimizing based on a vector of prices. When participating in ancillary services markets, the optimization problem must account for participation uncertainty – successful activation in the reserve market is not always guaranteed. This introduces a stochastic element: future energy procurement and associated costs depending on probabilistic outcomes. In general, solving this problem would mean tackling a stochastic optimization problem with a large scenario tree. Rather than that, we have simplified this model using probability-weighted averages and scenario sampling.

For our analysis we have decided to analyze a mixed scenario for the ancillary market participation. The combined scenario operates by allocating energy storage participation between the daytime day-ahead market (between 6 AM to 6 PM) focused on energy shifting, and nighttime ancillary services participation (between 6 PM and 6 AM) when the system provides secondary reserve capacity. The reason for the division lies in the fact that charging during the night can be very flexible – we can exploit the period from 6pm to 6am and try to charge the EVs with negative aFRR price signals. While during the day, some flexibility is possible, but time limitation makes it not aFRR viable. For EV charging stations, this combination is particularly optimal because daytime DA market participation allows the system to defer charging during peak price periods (when most EVs are parked at work and don't need immediate charging), while nighttime operations can strategically wait for the start of charging to provide demand response for ancillary services while ensuring vehicles are fully charged by morning.
The basis in this case is the optimized energy profile from the previous case, with the limitation to 4 operational hours in each 12-hour time slot. We divide the profile into daily and nightly components:

$$E_{\text{opt}} = E_{\text{day}} + E_{\text{night}}$$

Energy consumed during the daytime is bought at the DA market at price p_{DA}, while during the nighttime, we assume a price of the secondary reserve p_{sec}, which we simplify with the following expression $p_{\text{sec}} = 0.2 \cdot p_{\text{DA}}$ (the lowest price which can be offered for energy in aFRR market in Slovenia). We assume a 50% energy received (the number is a pessimistic guess, assuming that we would be able to get 50% of the energy consumed in the ancillary services market. We cannot predict the amount of activations, but based on history data, as we act as an aggregator, we could get more than that.) in the ancillary market, which means that half of the allocated hours would be won and paid at the price p_{sec}, while the other half would then have to be paid at the imbalance prices, p_{imb}. In vector form, we will describe this as E_{sec} , which is the same as E_{night}, but has zeros at time stamps of those 50 % not-activated cases. Total supplier's cost can now be written as

$$C_{\text{supplier}} = \sum_t E_{\text{day}}^t \cdot p_{\text{DA}}^t + E_{\text{sec}}^t \cdot p_{\text{sec}}^t + \left[E_{\text{night}}^t - E_{\text{sec}}^t\right] \cdot p_{\text{imb}}^t$$

We can define a difference between the cost of the original not-optimized profile as C_0 and the supplier's cost as $\Delta = C_0 - C_{\text{supplier}}$. The relative difference is then $\sigma_C = \Delta / C_0$. To account for the stochastic nature of the problem, we execute the optimization algorithm several thousand times, each iteration randomly sampling different activation scenarios based on the assumed probability distribution. For each simulation run, we randomly determined which nighttime hours would be successfully activated for ancillary services (receiving payment at p_{sec}) and which would result in imbalance costs (charged at p_{imb}). This approach allows us to generate probability distributions of cost savings rather than point estimates, and it highlights the issue of uncertainty in the market. The results present the average outcomes across all simulation runs, providing a more realistic assessment of expected performance under uncertain market conditions.

Table 3 showcases the average results from 10,000 runs of the random scenarios.

Table 3: Optimization results for different capacity utilization levels, DA market and ancillary services. Added limitation of charging to a maximum of 4 hours during work and 4 hours during night.

Share of max power	**60%**	**70%**	**80%**	**90%**	**100%**
Original cost [EUR]	31,637				
Optimized cost [EUR]	13,974	13,026	12,324	11,799	11,416
Absolute difference [EUR]	17,663	18,611	19,313	19,838	20,221
Relative difference [%]	56	59	61	63	64

Figure 5 shows the comparison of different scenarios: the DA scenario with work and night-hour limitations, and the combined DA and ancillary services market scenario. The results indicate that combining DA with ancillary services can enhance savings by around 10% on average compared to DA-only optimization. Additionally, increasing the share of utilized charging points does not significantly improve savings, suggesting that operating above 60% of maximum power is unnecessary. Notably, at 60% of maximum power, savings of 56% compared to original costs can be achieved.

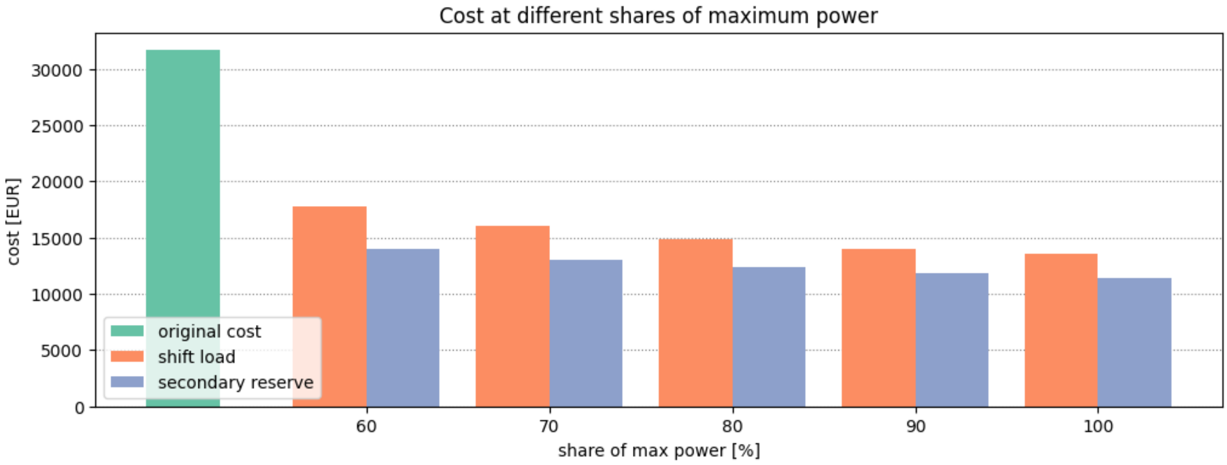

Figure 5: Optimization results for scenarios of DA and DA & ancillary services markets combination.

2.4 Implicit flexibility optimization based on grid tariff

2.4.1 Context and motivation

While the previous sections address explicit flexibility trading through participation in DA, ID, and ancillary service markets, flexibility can also be exploited in an implicit manner—by optimizing end-user consumption patterns locally without direct market participation. Within OPENTUNITY, this approach is demonstrated through Amibit's Reduxi [3] HEMS platform, which enables intelligent control of household loads to reduce electricity costs and support grid efficiency.

New Slovenian network tariff system

Since 1 October 2024, Slovenia has implemented a new network-tariff scheme [1] that drastically changes how distribution charges are calculated. The new model bills users according to the highest 15-minute average power demand measured within defined time blocks, shifting the emphasis from total monthly energy consumption to instantaneous power usage.

Key features of the system include:

- **Five daily time blocks** reflecting different grid-load conditions, each with its own tariff.
- **Two seasonal periods**: a high season (November–February) and a low season (March–October).
- **Contracted capacity per time block**, declared in advance and based on historical consumption.
- **Excess-capacity surcharges** whenever measured power exceeds the contracted value within any 15-minute interval.
- **Shift of emphasis from energy to power**: the new billing system places greater weight on instantaneous power (kW) rather than total energy consumed (kWh), incentivising users to reduce and smooth their peak demand.

Under this regime, households are incentivized to flatten demand peaks and shift flexible loads toward lower-tariff periods. The system promotes energy efficiency and network stability but also introduces new challenges, particularly the risk of exceeding contracted power limits, which results in penalties.

Within OPENTUNITY, the existing functionalities of the Reduxi HEMS were enhanced to address these challenges by integrating the forecasting and optimization capabilities developed in Phase 1 of Task 4.4. Building on these developments, a tariff optimization algorithm was designed to anticipate household consumption, evaluate tariff impacts, and automatically control flexible devices (mainly heat pumps) to minimise network charges while maintaining user comfort.

The following subsection presents the conceptual design and operation of this algorithm.

2.4.2 Methodological concept

The tariff optimization algorithm developed within OPENTUNITY builds on the Phase-1 forecasting and data-preparation framework. Its objective is to minimise total household electricity cost under the new network tariff while preserving thermal comfort and respecting contracted power limits. The approach focuses primarily on heat pumps (HPs), given their widespread adoption, thermal inertia, and strong link between power consumption and indoor comfort.

The algorithm functions as a predictive control routine: it uses external weather forecasts and household-load forecasts to anticipate future cost implications, and combines this with a data-driven thermal model of the building-HP system to determine when to operate the HP in order to shift load away from high-tariff or capacity-penalty periods and pre-heat intelligently. This process is illustrated in Figure 6.

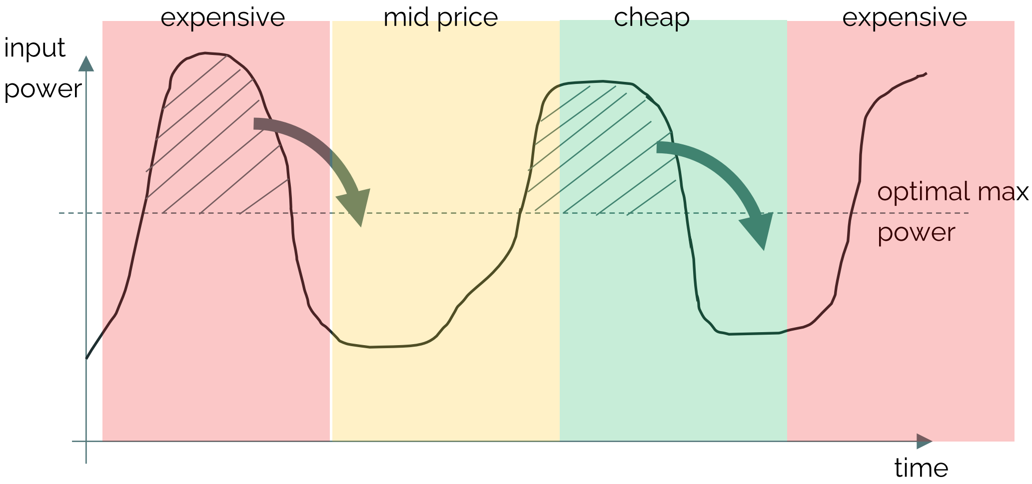

Figure 6: Tariff optimization logic representation

The optimization problem is formulated as a constrained minimisation of a cost function:

$$minimise\ C_{total} = C_{energy} + C_{capacity} + C_{comfort}$$

Where:

- C_{energy} represents the cost of energy consumption in each tariff block
- $C_{capacity}$ represents potential penalties for exceeding contracted power and
- $C_{comfort}$ quantifies the deviations of indoor temperature from the preferred comfort range (typically 0.5–1 °C from the preferred comfort level).

The algorithm uses the following inputs:

- **Forecasted household consumption**, using the forecasting methodology developed in Phase 1.
- **Tariff structure data**, including time-block definitions, seasonal factors, and contracted power levels.
- **Comfort parameters**, defining the acceptable indoor temperature range

The algorithm follows the next steps:

1. The **baseline demand** is forecasted
2. **Thermal state prediction** with the data-driven thermal model picks candidate HP schedules

3. The **cost model evaluates** each candidate schedule by combining energy cost, capacity penalty risk and comfort deviation
4. **Optimization** picks the schedule that minimises total cost while respecting contracting power and comfort constraints

Data-driven thermal model (DNN with optional ODE model)

Instead of detailed physics-based building model (which is impractical at scale) or a hand crafter first-order thermal model, the thermal dynamics of each building are learned directly from data. For the model a deep neural network (DNN) f_θ is trained using historical load data, as well as measured indoor and outdoor temperatures. It can be expressed in a form of a function:

$$T_{t+1} \approx f_\theta(T_{t:t-k}, P_{t:t-k}, T_{t:t-k}^{out})$$

where T is the indoor temperature, P is the electrical power of HP and T^{out} is the outdoor temperature and the look-back window k captures thermal inertia and past dynamics. To enhance generalisation and enforce physically plausible dynamics (e.g., smoothing, inertia, monotonic cooling/heating behaviour), the DNN can be augmented with a lightweight ODE-style residual sub-model g_ϕ (a lightweight, physics-inspired term). The combined hybrid model becomes:

$$T_{t+1} \approx f_\theta(\cdot) + g_\phi(\cdot)$$

Here g_ϕ encodes a simplified version of a thermal-dynamics equation (for example derived from an RC network model) but with limited parameters and trained in parallel or integrated with the DNN. A schematic representation of this modelling concept is shown in Figure 7.

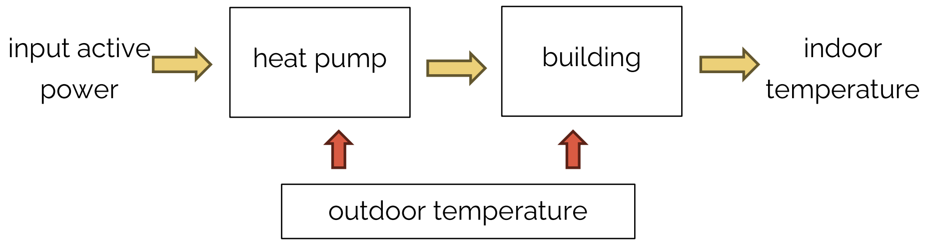

Figure 7: Building thermal behavior modelling concept

By adopting this modelling approach, the optimization algorithm can evaluate the indoor-temperature outcomes of candidate HP schedules reliably and efficiently, enabling proactive shifting or pre-heating that reduces tariff costs and avoids exceeding contracted power limits.

2.4.3 Algorithm testing approach and expected outcomes

The algorithm will be evaluated using data from approximately 30 households equipped with Reduxi HEMS devices and heat pumps. Because full live control may not initially be feasible, the evaluation will rely on simulated control performance using real-world measurement and device-status data. Each selected household will provide at least one full year of historical power-consumption and heat-pump operating-state data. The collected parameters include active power consumption in 15-minute intervals, heat-pump operating statuses (e.g., OFF, ECO, Normal, Boost), set-point temperatures (where available), and outdoor temperature (where available). Prior to analysis, the dataset will be cleaned to remove incomplete or inconsistent records, thereby ensuring the reliability of the results.

Collected data serves two primary purposes: firstly, to train the data-driven thermal model described in Section 2.4.2 for each household (or household cluster); secondly, to support simulation of the control algorithm under the new tariff structure. For each household simulation, a baseline scenario representing normal user operation (without optimization) will be compared with an optimized scenario (with the algorithm applied).

The main expected outcome is that the algorithm will successfully reduce total household electricity costs under the new tariff system, by intelligently timing consumption and avoiding unnecessary exceedance of contracted power limits. Additionally, the evaluation will test how suitable heat pumps are as flexibility assets and how their controlled operation impacts indoor temperature states and comfort levels.

3 Optimal selection of available flexibility

3.1 Background and literature review

In modern electricity systems, an aggregator or FSP acts as an intermediary that coordinates a portfolio of DERs in order to jointly participate in flexibility and ancillary services markets. The aggregator bundles "behind-the-meter" assets whose individual size or heterogeneity would otherwise make them non-viable for direct market engagement, turning them into a coherent virtual resource.

Under the EU's Clean Energy Package (notably Directive 2019/944 [4]), independent aggregators have been explicitly recognized and granted the right to enter electricity markets on behalf of their assets, without requiring the consent of retail or generation incumbents. This shift is meant to unlock demand response, distributed flexibility, and grid services from smaller actors that historically could not engage with system operators or wholesale markets.

The process by which an aggregator participates in flexibility markets and manages its portfolio can be described as a sequence of key stages, outlined below.

1. **Portfolio onboarding and pre-qualification**
   The process begins with contracting customers and their flexible devices, capturing their technical parameters (e.g. maximum/minimum power, ramp rates, state-of-charge or temperature bounds), and deploying the necessary metering and control infrastructure. If required by local rules, the aggregator must also pre-qualify its portfolio with the DSO and/or TSO (e.g. proving controllability, meeting minimum bid sizes, ensuring metering granularity). In many cases, the aggregator may group assets by network zones or market segments (e.g. DSO flexibility zones).
2. **Forecasting, baseline definition and availability**
   In this step, the aggregator builds forecasts of generation and load for each asset, estimates when devices are available for flexibility, and determines the flexibility envelope of the portfolio under technical constraints. It also defines baselines (the counterfactual "no-activation" profile) for later verification and settlement.
3. **Product selection and bid formation**
   Based on forecasts and flexibility potential, the aggregator chooses which markets or products to enter (e.g. day-ahead/intraday energy markets, balancing services, local DSO flexibility). It then formulates bids or offers that comply with each market's rules (minimum volumes, durations, ramp limits, gate closures). This stage often involves **value stacking**, deciding how to allocate the same flexibility to different overlapping opportunities to maximize revenue.
4. **Validation and nominations**
   Before finalizing the bids, validation is performed to ensure network feasibility (DSO checks for congestion and voltage constraints) and market consistency (balance responsible parties/TSO checks to ensure the portfolio remains balanced). Once validated, the aggregator issues nominations or confirmations indicating which flexibility it will deliver.
5. **Activation, disaggregation and real-time control**
   When the flexibility request (activation) arrives, the aggregator must disaggregate the requested profile into per-asset setpoints, respecting device constraints and time-coupled

dynamics (e.g. battery state of charge (SoC), Heating, Ventilation and Air-Conditioning (HVAC) thermal inertia). Real-time telemetry monitors compliance, and intraday re-dispatch or corrective trades may be executed to address deviations.

6. **Measurement and verification**
   Post-event, the aggregator compares the actual delivered response to the baseline and contracted activation profile. It attributes performance to individual assets or zones, produces auditable records, and prepares the data required for settlement, penalties, or bonuses.
7. **Settlement, reporting and learning loop**
   The market operator calculates revenues or penalties resulting from deviations, and the aggregator receives payments, which must then be fairly distributed among the participating customers. The aggregator also updates forecasts, recalibrates cost and penalty models, and incorporates performance feedback to improve future bidding and dispatch decisions.

The optimal selection algorithm is positioned at step 5 (Activation, Disaggregation & Real-time Control). At this point, the aggregator has received a flexibility request (upward or downward) to deliver and must decide which subset of its assets, in which time slots, should provide that service. This decision is complicated by:

- The combinatorial nature of the problem (many assets and many time slots),
- Inter-temporal coupling (e.g. battery SoC evolution, HVAC thermal inertia, minimum up/down times),
- Real-time uncertainty and drift (forecast errors, device dropouts),
- The requirement to remain feasible under all constraints while minimizing cost or revenue loss.

3.1.1 Literature review

To identify solution approaches capable of handling the complex disaggregation and dispatch problem, we conducted a targeted review of the literature on energy systems, DER scheduling, flexibility markets, and combinatorial optimization under uncertainty. Several methodological groups emerged, which are summarized in Table 4 together with their main strengths and limitations.

Table 4: Literature review findings

Method / Group	Strengths / Advantages	Weaknesses / Limitations
Genetic Algorithms /Evolutionary Methods [5], [6], [7]	• Flexible encoding of heterogeneous decision variables (assets × timeslots) • Good at exploring non-convex, discrete solution spaces • Can support multi-objective formulations (cost vs reliability vs fairness)	• Requires parameter tuning (population size, crossover/mutation rates) • Convergence may be slow for large-scale instances • Risk of premature stagnation or getting stuck in local optima
Hybrid / Memetic & Parallel Metaheuristics [8], [9]	•Improves convergence speed via local search, heuristics, or hybrid operators • Better fine-tuning / exploitation combined with global search	• More complex to design and tune • Needs domain-specific heuristics • Overhead of coordination / parallel infrastructure may outweigh gains for smaller problems

	•Parallel implementations scale better	
Exact / Relaxation / Decomposition [10]	• Can sometimes deliver optimal or bounded solutions • More predictable or provable performance • Useful for small-to-medium problem sizes or subproblems	• Often doesn't scale to high-dimensional, non-convex, discrete problems • Relaxations may introduce large optimality gaps • Hard to incorporate real-time recourse/uncertainty elegantly
Stochastic /Adaptive Optimization / MPC [11], [12]	• Explicit modeling of uncertainty, recourse, drift • Greater robustness to deviations and forecast errors • Adaptability via rolling horizon or feedback control	• Scenario explosion or large computational complexity • Hard to integrate discrete constraints, nonlinearity, device coupling • May become overly conservative if uncertainty sets are broad

From this comparison, it is clear that while exact and decomposition methods offer formal guarantees, they are often difficult to maintain and scale when the problem involves a large number of assets, time intervals, and operational constraints. Stochastic and robust methods provide resilience to uncertainty but impose heavy computational burdens, especially for heterogeneous portfolios. Hybrid metaheuristics can mitigate some of these issues but add design complexity.

It is acknowledged that the underlying problem could also be formulated as a Mixed-Integer Linear Program (MILP), and modern solvers such as HiGHS or Gurobi are capable of handling problems of comparable size. However, in the context of this task, where multiple asset types, evolving constraints, and dynamic cost representations are expected, the use of a Genetic Algorithm offers greater modelling flexibility. This approach avoids the need to reformulate the entire optimization problem when new devices or constraints are added.

Considering the high dimensionality, inter-temporal coupling, and need for iterative testing within the project, the GA approach was selected as a practical balance between flexibility, transparency, and computational tractability.

3.1.2 Genetic algorithms

Genetic Algorithms (GAs) [13], a well-established class of evolutionary computation methods, are particularly suitable for optimization problems characterized by complex, rugged search spaces and multiple, often conflicting objectives. They operate on a population of candidate solutions and iteratively evolve these solutions using biologically inspired operators such as selection, crossover, and mutation. Over successive generations, the population converges toward high-quality solutions without the need for gradient information or convexity assumptions.
Several features make GAs especially relevant to the flexibility disaggregation and scheduling problem:

- **Flexible encoding of heterogeneous assets**: GAs can represent diverse decision variables (e.g., different DER types and time-indexed constraints) using adaptable encoding schemes ranging from simple binary strings to structured arrays.

- **Support for multi-objective optimization**: Aggregators must balance revenue, operational costs, degradation, and delivery reliability. GAs naturally accommodate composite objective functions or Pareto-based approaches to explore these trade-offs.
- **Robustness in non-convex landscapes**: Unlike classical optimization techniques, GAs remain effective on discontinuous, nonlinear, or multi-modal problems where traditional solvers often struggle.

3.2 Optimal selection algorithm design

3.2.1 Problem definition and objectives

Aggregators participating in flexibility markets face the challenge of scheduling a portfolio of heterogeneous DERs in order to deliver explicit flexibility activations at minimum cost. The complexity arises from the large number of assets and time intervals, diverse technical constraints, user comfort requirements, and uncertainty in availability and forecasts.

The algorithm developed in this work is designed to address this scheduling problem in a way that can be applied across different flexibility products. By adapting input parameters such as size, duration, and ramping rules, it can operate in a variety of market settings. At its core, the algorithm dynamically selects which assets to activate in each interval in order to minimize delivery cost, while ensuring that:

- all technical and operational limits are respected,
- customer comfort and asset health are preserved, and
- uncertainty in availability and forecasts is handled through a rolling, adaptive operation.

The algorithms main inputs are the following:

- **Flexibility forecasts** (upward/downward potential per asset and availability windows),
- **Asset metadata** (rated capacity, maximum power, ramping limits, rest periods etc),
- **Market signals** (prices, accepted bids, product requirements),
- **Real-time asset states** (availability, SoC, indoor temperature, PV availability), and
- **Cost information** (activation costs per asset).

The decision variables of the algorithm are the dispatch set-points $x_{i,t}[kW]$ for each asset i and each time interval t.

The algorithm also includes several feasibility constraints:

- Per-asset bounds (technical capabilities, availability windows, SoC/comfort limits) – the detailed definition is included in section 3.2.4,
- Portfolio or site-level limits (e.g., feeder cap, total site rating),
- Delivery balance: aggregate dispatch must match the required flexibility profile, or deviations are penalized.

The primary objective of the optimal selection algorithm is to minimize the total delivery cost of flexibility over the optimization horizon.

$$\min_{X} \sum_{t\in W} \left(\sum_{i\in I} c_{i,t} \cdot x_{i,t} \cdot \Delta t + Penalty(s_t) \right)$$

Where $c_{i,t}$ represents the per-asset activation cost, $x_{i,t}$ the dispatch set-point and s_t any shortfall in meeting the requested activation.

3.2.2 High-level algorithm overview

The operation of the optimal selection algorithm can be summarized as a sequence of steps that are repeated for each window in the rolling horizon. Figure 8 illustrates the workflow at a high level, showing how candidate solutions are generated, evaluated, evolved, and ultimately applied in real time.

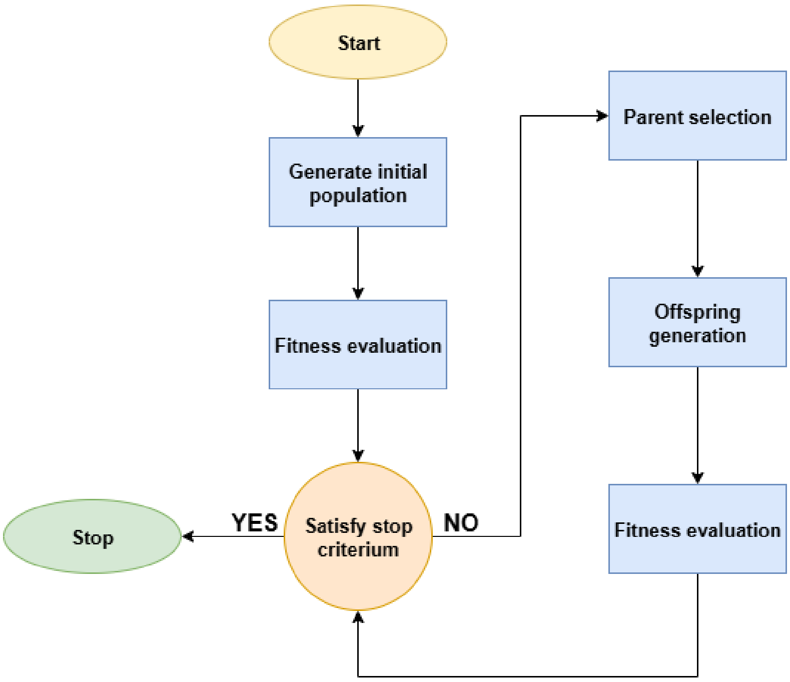

Figure 8: High-level GA workflow

The main steps are as follows:

1. **Generate initial population:**
   A population of candidate dispatch plans (assets × intervals in the look-ahead window) is created. The initial population consists of a simple merit-order plan and several randomly generated feasible plans, each of which is repaired to respect basic technical limits before evaluation.
2. **Fitness evaluation:**
   For each candidate plan, feasibility is enforced by clipping set-points to instantaneous caps and device rules (e.g., SoC, comfort bands, timers, ramp rates). Within each interval, the plan is repaired to match the requested total where possible; if capacity is insufficient, set-points are left at caps and the resulting shortfall is recorded. Fitness is then calculated as the **total delivery cost over the window**, including any shortfall penalties.

$$F(X) = \sum_{t \in W} \sum_{i \in I} c_{i,t} \cdot x_{i,t} \cdot \Delta t + \sum_{t \in W} C^{pen}(s_t) \cdot \Delta t$$

3. **Select parents**
   Candidate plans with lower cost are more likely to be chosen as parents, typically through tournament selection, while maintaining sufficient diversity in the population.
4. **Generate offspring**
   - **Crossover**: Parents are recombined by exchanging coherent blocks of their encoding (for example, swapping entire time blocks between plans or exchanging dispatch vectors of certain assets) so that promising traits are inherited by the offspring.
   - **Mutation**: Offspring are diversified by small random perturbations of set-points, such as shifting power within feasible limits or toggling inclusion in a given interval. This maintains diversity and encourages exploration of new solutions.
5. **Constraint handling and repair**
   After crossover and mutation, each new individual is passed again through the repair procedure to ensure that technical constraints are satisfied and that the per-interval flexibility target is met where possible (or the shortfall clearly recorded). This guarantees that the evolving population remains valid or near valid.
6. **Form new generation**
   A new population is formed from the selected parents and the newly created offspring, balancing the retention of strong solutions with the introduction of new variations.
7. **Stopping criterion**
   The cycle of evaluation, selection, crossover, mutation, and repair continues until the generation budget is reached or improvements plateau. The best candidate plan for the window is then selected.
8. **Dispatch & roll**
   Only the first interval of the best plan is executed. Asset states (e.g., SoC, temperatures, timers) are updated, the horizon is shifted forward by one interval, inputs are refreshed, and the procedure restarts from step 1.

3.2.3 Cost modelling

Aggregators, or flexibility service providers (FSPs), derive their revenues by coordinating a portfolio of DERs and offering flexibility services in different markets. Typically, they are compensated through two main payment streams:

- **Utilisation payments**, received per MWh of flexibility actually delivered during events.
- **Availability payments**, received per MW of capacity made available for activation, depending on the market product.

Once revenues are received, they are shared with asset owners according to different remuneration models:

- **Pay-for-performance**: compensation linked to both availability and actual delivery, sometimes including bonuses or penalties based on quality.
- **Revenue share**: asset owners receive a share of the aggregator's market revenues, sometimes combined with shared-savings clauses (e.g., when flexibility lowers electricity bills).

- **Tariffed incentive**: a simple fixed €/kWh rate for delivered flexibility, often used for residential assets.

These practices highlight that flexibility provision is not free: every activation has a cost, whether in energy terms, equipment wear, lost opportunities, or administrative overhead. To operate profitably and sustainably, an aggregator must therefore optimize not only its bidding strategy but also its activation decisions at delivery time.

In practice, activation costs can be defined in several ways. In some cases, contracts directly specify a simple activation cost per kWh, or a time-dependent tariff (for example, curtailing a heat pump when the building is unoccupied may be priced lower than during occupied hours). However, this approach is not common for low-capacity, residential-type assets, where explicit cost contracts are rarely used. Instead, it becomes essential for the aggregator to model activation costs internally in a way that captures the real technical and economic impacts of dispatch. This provides a consistent basis for optimization and ensures that the algorithm's decisions align with both profitability and user acceptance.

To capture these incremental costs in a consistent way, we introduce the concept of Short-Run Marginal Cost (SRMC). SRMC represents the true per-asset, per-interval cost of delivering flexibility and forms the basis of the optimization in our algorithm. In essence, the algorithm selects in each interval the combination of assets with the lowest SRMC, ensuring that the required flexibility is delivered at minimal cost while preserving asset owner value. This approach mirrors the well-established merit order principle used in wholesale energy markets.

SRMC is defined as:

$$SMRC_{i,t} = \Delta\lambda_{i,t} + A_{i,t} + O_{i,t} + R_{i,t} + F_i$$

where the components are:

- **Energy price differential** ($\Delta\lambda_{i,t}$): This term captures the cost or benefit of shifting energy in time relative to a baseline. For loads, it reflects whether consumption is moved into more expensive or cheaper price periods. For storage, it accounts for the net cost of discharging now and recharging later, including efficiency losses. Because energy costs are typically settled outside of local flexibility markets, this term is central to ensuring that dispatch decisions do not unintentionally increase energy bills or imbalance costs.
- **Degradation and wear** ($A_{i,t}$): DERs are not cost-free to use indefinitely. Every battery charge–discharge cycle contributes to capacity fade, and EV discharges for grid services may accelerate battery aging. This degradation can be monetized as a cost per kWh, derived from the asset's replacement cost and expected cycle life. By including it explicitly in SRMC, the algorithm avoids over-using assets in ways that harm long-term value and ensures that activations only occur when revenues exceed wear costs.
- **Opportunity and comfort penalties** ($O_{i,t}$): Many flexible assets have implicit "soft" costs tied to user satisfaction or foregone opportunities. For HVAC, excessive curtailment may lead to discomfort if temperatures drift outside acceptable ranges. For EVs, flexibility may risk insufficient charge for mobility needs; for PV, curtailment reduces export revenues. By assigning a penalty value to such situations, SRMC internalizes user-facing impacts and ensures that the optimization balances financial gains with continued customer acceptance.
- **Forecast uncertainty cost** ($R_{i,t}$): Inaccuracies in forecasting the baseline operation of small-scale flexibility assets introduce a risk of non-delivery and potential imbalance costs. This

component reflects the historical forecasting performance of each asset, assigning higher marginal costs to those with greater baseline uncertainty. By incorporating this factor, optimization naturally favors assets with more reliable forecasts, reducing operational risk and improving delivery consistency.

- **Fixed and administrative costs** (F_i): Beyond energy and wear, each activation carries overheads such as telemetry, validation, platform fees, or contractual transaction costs. These may be modest on a per-activation basis but are significant when scaled across large portfolios or frequent activations. Incorporating them into SRMC guarantees that revenues cover both the direct and indirect costs of flexibility provision, ensuring profitability and sustainability of aggregator operations.

3.2.4 Per-asset models (technical constraints and SRMC)

The genetic algorithm seeks to minimize the cost of delivering flexibility, but it can only do so within the technical capabilities of the underlying assets. Each resource type has its own physical and operational limits that define when and how it can provide upward or downward flexibility. These feasibility constraints are expressed as per-interval caps on activation, derived from asset baselines, internal states (e.g. state of charge or indoor temperature), and device-specific rules such as timers or ramp limits. In the optimization loop, candidate schedules are always clipped and rescaled against these caps during the repair step, ensuring that dispatch plans remain realistic and implementable.

At the same time, the optimizer must weigh the SRMC of activating each asset. SRMC translates the economic and technical side-effects of flexibility into a €/kWh value per asset and interval. It accounts for avoided or incurred energy costs (based on the day-ahead price curve, except for PV where a fixed feed-in tariff is used), degradation and wear, opportunity or comfort penalties, and administrative overheads. This allows the algorithm to prefer assets that are not only technically able to respond, but also cheapest when all relevant costs are considered.

In practice, the GA operates on three layers of information per asset:

- **Availability mask** $a_{i,t}$: a binary signal indicating whether the asset is controllable/available in a given slot t. If $a_{i,t} = 0$, the asset is excluded in that slot.
- **Feasibility caps** $\bar{p}_{i,t}$: the maximum activation an asset can deliver in slot t, ensuring compliance with physics and user limits
- **Marginal costs** $c_{i,t}$: the €/kWh value associated with that activation, which shapes the fitness function.

Together, these three layers define the feasible set $F_{i,t}$. The following sub-sections describe for each asset type both the technical constraints, as well as the SRMC model that quantifies their costs.

3.2.4.1 EV chargers

EV chargers provide upward flexibility by curtailing charging relative to the expected baseline demand. They cannot export power (no V2G), so their only option is to reduce consumption when requested.

Feasibility caps

The feasible upward activation p_t^{up} at slot t is limited by the forecasted baseline demand b_t^{EV}. In case the charging site is unavailable, the cap is equal to 0.

$$0 \leq p_t^{up} \leq \begin{cases} b_t^{EV}, if\ available\ at\ t \\ 0, \qquad otherwise \end{cases}$$

The device setpoint (charging limit) applied to the chargers is:

$$u_t = \max\{0, b_t^{EV} - p_t^{up}\}$$

Where:

- Baseline forecasted charging demand b_t [kW] per slot
- Availability mask ∈ $\{0,1\}$ per slot (indicating if the charging location is present and controllable in slot t).

SRMC model

The EV SRMC captures the balance between immediate energy savings from curtailment and the future cost of recharging that energy, along with forecast uncertainty and administrative overheads. The SRMC for EV chargers consists of the following components:

- **Energy term:** Avoided energy purchase at the DA price λ_t^{buy}, which represents the immediate cost saving achieved by curtailment.
- **Opportunity term:** Reflects the cost of recharging later. When charging is curtailed now, the same energy must be refilled later at a representative future price v_{refill}. This value is taken as low-percentile reference of the DA price curve within a fixed look-ahead window. The term is scaled by the charging efficiency η_{chg} to account for energy losses.
$$O_{EV,t} = \frac{1}{\eta_{chg}} v_{refill}$$
- **Forecast uncertainty term** ($R_{EV,t}$): Adjustment that accounts for baseline prediction errors. It is defined as:
$$R_{EV,t} = \gamma_{EV} \cdot \sigma_{EV}$$
where σ_{EV} is the historical baseline error and γ_{EV} a €/kWh scaling factor.
- **Administrative term** (F_{EV}): Fixed activation-related costs, including telemetry, validation, and platform fees.

Together they form:

$$c_t^{EV} = -\lambda_t^{buy} + O_{EV,t} + R_{EV,t} + F_{EV}$$

3.2.4.2 Battery Energy Storage Systems (BESS)

BESS can provide both upward flexibility (by discharging) and downward flexibility (by charging). In the implementation, the direction is fixed for each optimization window via a mode flag. The activation variable is always non-negative, and the signed setpoint is derived internally depending on whether the mode is “up” or “down.”

Feasibility caps

For each slot t, feasible activation is limited by both instantaneous power limits and the available state of charge (SoC).

- Up mode (discharge)

$$0 \leq p_t^{up} \leq min\left\{P_{dis}^{max}, \frac{(E_t - E^{min})\eta_{dis}}{\Delta t}\right\};\ E_{t+1} = E_t - \frac{p_t^{up}}{\eta_{dis}}\Delta t$$

- Down mode (charge)

$$0 \leq p_t^{down} \leq min\left\{P_{chg}^{max}, \frac{(E^{max} - E_t)}{\eta_{chg}\ \Delta t}\right\};\ E_{t+1} = E_t + p_t^{down}\eta_{chg}\Delta t$$

Where:

- Initial energy state E_t [kWh] with SoC bounds SoC^{min} and SoC^{max}.
- Power limits P_{chg}^{max}, P_{dis}^{max} [kW].
- Efficiencies $\eta_{chg}, \eta_{dis} \in (0,1]$.

Additionally, availability can set the cap to zero for selected slots (the BESS not available, or out of order). Conservative forward projection shrinks future caps if energy is scheduled early, preventing infeasible SoC later in the window.

SRMC model

The BESS SRMC balances instantaneous energy prices, future opportunities, degradation costs, and administrative overheads.

- **Energy term**: current day-ahead buy price λ_t^{buy}
- **Opportunity term**: explicit look-ahead comparing current action with expected future prices
  - v_{refill}: low future buy reference (minimum or low percentile) for recharge after discharge
  - v_{sell}: high future sell reference (maximum or high percentile) for resale after charging
  - Both are computed from the DA price curve
  - The opportunity terms are scaled by the round-trip efficiency $r = \eta_{chg}\eta_{dis}$ since energy lost in the cycle must be accounted for.
- **Degradation term** c_{deg}: constant cost, which is added to both charge and discharge. It reflects the wear and lifetime reduction from cycling.
- **Administrative term** F_{BESS}: fixed costs for telemetry, validation etc.

Formulas:

- Discharge (upward flexibility)

$$c_t^{BESS,up} = -\lambda_t^{buy} + \frac{1}{r}v_{refill} + c_{deg} + F_{BESS}$$

- Charge (downward flexibility)

$$c_t^{BESS,down} = \lambda_t^{buy} - r\ v_{sell} + c_{deg} + F_{BESS}$$

3.2.4.3 PV generation

PV can provide downward flexibility by curtailing active power relative to the PV baseline forecast.

Feasibility caps

Maximal curtailment at timeslot t equals the PV baseline forecast b_t^{PV}:

$$0 \leq p_t^{down} \leq b_t^{PV}$$

The applied setpoint is then:

$$u_t = b_t^{PV} - p_t^{down}$$

Key parameters:

- PV baseline b_t^{PV} [kW] per slot (production forecast).

SRMC model

Curtailing PV generation in slot t means forfeiting export revenue at that time. This is treated as a positive cost equal to the feed-in tariff (FIT) or selling price for that slot. The PV system provides only downward flexibility (curtailment), and no upward action is possible.

The SRMC for PV consists of the following components:

- **Energy term:** Lost export revenue, represented by the FIT or the time-dependent sell price $O_{PV,t}$.
- **Forecast uncertainty term** $R_{PV,t}$: Reflects the deviation between forecasted and actual PV generation due to weather variability. It is defined as:

  $$R_{PV,t} = \gamma_{PV} \cdot \sigma_{PV}$$

  where σ_{PV} is the historical baseline error and γ_{PV} a €/kWh scaling factor.
- **Administrative term** F_{PV}: Fixed costs related to control, monitoring, and verification of activation.

The SRMC expression is therefore defined as:

$$c_t^{PV} = O_{PV,t} + R_{PV,t} + F_{PV}$$

3.2.4.4 HVAC systems

HVAC units provide upward flexibility by curtailing consumption relative to a baseline (e.g., relax cooling/heating) and downward flexibility by boosting consumption (e.g., pre-cool/pre-heat). All actions are baseline-relative.

Feasibility caps

We enforce feasibility through two guard types that are evaluated every slot in the rolling window:

Timer/budget guard (meant for equipment protection)

- min_on and min_off (slots): once the unit turns ON (or OFF), it must remain in that state for at least the specified number of slots (this prevents short cycling)
- starts_per_hour (in the rolling window): limit how many ON switches are allowed per hour
- events_per_hour (in the rolling window): limit the number of flexibility events per hour

Thermal guard (comfort preservation)

Here we include a simplified first-order thermal model to preview the next slot indoor temperature and adapt the caps so that any chosen activation cannot violate the comfort band on the next step. The model is defined with the following term:

$$T_{t+1} = T_t + a(T_t^{out} - T_t) \pm b \cdot s_t$$

Where:

- T_t is the indoor temperature [°C] (measured or simulated)
- T_t^{out} the outdoor temperature [°C] (forecast)
- $a \in [0,1]$ is the reversion toward outdoor per slot.
- $b > 0$ is the temperature effect per slot at full power (sign depends on heating or cooling)
- P_{nom} is the nominal electrical power [kW]
- $s_t = \frac{p_t^{HVAC}}{P_{nom}} \in [0,1]$ is the normalized power.
- The last term depends on the mode (heating/cooling): minus for cooling mode (increasing power lowers T), plus for heating mode (increasing power increasing T)

The comfort band limits are defined as (calculated for the next step):

$$T^{min} \leq T_{t+1} \leq T^{max}$$

Where T^{min} and T^{max} are the lower and upper limits of the temperature comfort band.

SRMC model

The HVAC SRMC combines two main terms:

1. **Energy price term**

$$energy_t = \begin{cases} -\lambda_t^{buy}; \ \ if\ UP\ (discharge) \\ +\lambda_t^{buy}; if\ DOWN\ (charge) \end{cases}$$

2. **Linear comfort term**

This term encourages the algorithm to not push the indoor temperature towards the limits of the comfort band, based on this it either applies a penalty or reward based on the direction the activation pushes the temperature. The comfort band is defined via minimum and maximum temperature T^{min} and T^{max}. From this the midpoint can be calculated:

$$T_{mid} = \frac{1}{2}(T^{max} + T^{min})$$

And the half-width of the band:

$$w = \frac{1}{2}(T^{max} - T^{min})$$

We define a normalized deviation from comfort as:

$$z_t = \frac{T_t - T_{mid}}{w}$$

In practice this means in the band center T_{mid}, z_t equals 0 and increases up to 1 to both limits. We determine whether a candidate activation pushes the indoor temperature towards or away from the midpoint. The direction of the temperature effect depends on both the thermal mode and the service requested. In cooling mode, curtailment (UP) reduces cooling and therefore raises indoor temperature, whereas boosting (DOWN) increases cooling and lowers it. In heating mode, curtailment reduces heating and lowers indoor temperature, whereas boosting increases heating and raises it.

Based on this, effect_sign is defined, which is +1 if the action moves temperature away from comfort and -1 if it moves towards. Then:

- If $effect\_sign \cdot z_t > 0$; action is pushing the T away from midpoint/comfort -> the term becomes a penalty
- If $effect_{sign} \cdot z_t < 0$; action is pushing the T towards the midpoint/comfort -> the term becomes a reward (negative cost)
- If $z_t = 0$, comfort term is 0

To scale this, we define:

$$M = \kappa \cdot |z_t|$$

Where κ is a tuning parameter in €/kWh, which defines the slope of the comfort penalty. Then:

$$comfort\ term = \begin{cases} +M; if\ action\ moves\ away\ from\ T_{mid} \\ -\rho M; if\ action\ moves\ towards\ T_{mid} \\ \ \ 0;\ \ if\ neutral \end{cases}$$

Where $\rho \in [0,1]$ is a reward factor that limits how much credit the optimization can get for comfort improving activities.

3. Forecast uncertainty term

To account for potential deviations between the forecasted baseline consumption and actual load, a forecast uncertainty cost is also included. It is defined as:

$$R_{HVAC,t} = \gamma_{HVAC} \cdot \sigma_{HVAC}$$

Where γ_{HVAC} is the historical baseline error factor and σ_{HVAC} the cost scaling factor.

Putting all this together we get:

$$c_t^{HVAC} = \begin{cases} -\lambda_t^{buy} + comfort\ term + R_{HVAC,t} + F_{HVAC}; if\ curtail\ (UP) \\ +\lambda_t^{buy} + comfort\ term + R_{HVAC,t} + F_{HVAC}; if\ boost\ (DOWN) \end{cases}$$

Regarding the SRMC, the HVAC combines a price term with a directional comfort term, as well as the admin costs. The comfort term is signed – it can either be a penalty (positive cost) if the chosen direction of activation would push the temperature nearer to the limit of the comfort band and a reward (negative cost). For this a linear model is used to generate the cost inside the comfort band for both directions. The term for both upward and downward direction is the following:

$$c_t^{HVAC} = \pm\lambda_t^{buy} + dir_{sign} \times M(z_t, \kappa, \rho) + R_{HVAC,t} + F_{HVAC}$$

3.3 Algorithm implementation

The optimal selection algorithm is implemented in Python, making use of DEAP for the evolutionary processes together with standard scientific libraries such as NumPy and pandas. The design is modular, with a clear separation between the optimization core, the technical models of the assets, and the cost modelling layer.

Each controllable device type is represented as a standalone class. These classes encapsulate both the technical constraints of the asset and the evolution of its internal state during operation. For example, the BESS class tracks the state of charge, enforces charge and discharge limits, and provides the maximum feasible flexibility available in each slot. This approach ensures that after each simulation, the full operational history of every asset can be retrieved and analyzed.

Each asset class also handles its availability state, which is represented as a binary (or fractional) mask indicating whether the asset is controllable in a given slot. If an asset becomes unavailable (due to disconnection, maintenance, or local constraint violation) its feasible power cap is automatically set to zero. This behavior allows the simulation to represent realistic operational conditions, such as temporary dropouts or communication failures, without modifying the optimization logic. In a real deployment, this mechanism would operate continuously in real time as new telemetry data arrive; in the current implementation, it is simulated within each rolling-horizon step.

Because the interface between the optimizer and the assets is standardized, new device types can be added without altering the core algorithm—only the corresponding class needs to be defined.

On the cost side, a separate modelling layer computes the SRMCs for each asset and interval, producing cost matrices for upward and downward flexibility. These values serve as the main optimization input and ensure that all decisions reflect not only energy prices but also degradation, comfort penalties, and administrative overheads. The decomposition of SRMCs by component is also logged, enabling later diagnostic and sensitivity analysis.

The optimization itself is carried out with a genetic algorithm operating on a rolling look-ahead window of discrete intervals. Each candidate solution is encoded as a matrix of dispatch setpoints across assets and slots. The initial population is seeded with a mix of merit-order heuristics and random feasible schedules. Through successive generations, individuals are evaluated on the basis of their total cost over the window, and standard evolutionary operators such as selection, crossover, and mutation are applied to explore the solution space. After each operation, schedules are repaired against the technical caps published by the assets so that feasibility is always preserved.

A rolling-horizon approach is employed to maintain adaptability. In each iteration, the optimizer plans over a horizon of several slots, but only the first interval of the best schedule is dispatched. Asset states are then updated, the horizon is shifted forward by one slot, and the process repeats with refreshed data. This design allows the algorithm to incorporate updated information on availability, forecasts, and prices in real time while keeping problem complexity manageable.

The algorithm requires several inputs to run:

- the **time discretisation** (slot length, e.g. 15 minutes),
- the **look-ahead horizon length** (number of slots optimized at once),
- the **day-ahead price curve** (used for energy valuation), and
- the **flexibility request profile**, specifying how much upward or downward flexibility must be delivered in each slot.

In addition, the optimization itself is controlled by several parameters that influence search behaviour and performance:

- the **population size** (number of candidate solutions per generation),
- the **number of generations** (iterations before termination),
- the **selection rule** (e.g. tournament size), and
- the **crossover and mutation probabilities**, which regulate the balance between exploitation and exploration.

By adjusting these parameters, a trade-off can be struck between computational speed and solution quality. Smaller populations and fewer generations are sufficient for simple portfolios, while larger and more diverse portfolios benefit from broader search to avoid premature convergence.

The implementation is intended as a flexible testbed. By varying portfolio composition, price conditions, or GA parameters, a wide range of scenarios can be simulated. The resulting dispatch schedules, asset state trajectories, and cost outcomes provide a basis for assessing the robustness of the algorithm and for comparing its behavior under different operating conditions.

3.3.1 Practical example of algorithm operation

To illustrate the operation of the optimal selection algorithm, a practical example has been implemented and tested. The scenario simulates the coordination of a small portfolio consisting of four different assets: an EV charging site, a PV installation, a BESS, and a HVAC load.
The algorithm is executed over a four-hour horizon with 15min intervals (together this means 16 time slots). The flexibility request profile alternates between upward and downward flexibility testing the system's ability to adapt to changing requirements and portfolio conditions.

Each asset is defined by technical and cost-related parameters:

1. **EV charging site** can provide upward flexibility by curtailing charging relative to its baseline charging demand profile.
   - **Technical parameters**: The baseline profile is provided as input, which varies between 0 and 25 kW.
   - **SRMC parameters** are presented in Table 5.

Table 5: Practical example, SRMC parameters for EV asset

Term	Parameters
Energy price differential ($\Delta\lambda_{i,t}$)	• benefit of avoided purchase $-\lambda_t^{buy}$
Degradation and wear ($A_{i,t}$)	/
Opportunity and comfort penalties ($O_{i,t}$)	• Look-ahead window N=16 slots • Percentile low=0.10 • $\eta_{chg} = 0.95$.
Forecast uncertainty cost ($R_{i,t}$)	• Baseline error factor $\sigma_{EV} = 0.2$ • Scaling factor $\gamma_{EV} = 0.036$ €/kWh
Fixed and administrative costs (F_i)	• Administrative term $F_{EV} = 0.01$ €/kWh

2. **PV unit** can offer downward flexibility by limiting the generation output.
   - **Technical parameters**: the PV forecasted baseline follows a bell-shaped production curve with a peak near 50 kW. The maximum curtailment in each slot equals the forecasted baseline.
   - **SRMC parameters** are presented in Table 6.

Table 6: Practical example, SRMC parameters for PV asset

Term	Parameters
Energy price differential ($\Delta\lambda_{i,t}$)	/
Degradation and wear ($A_{i,t}$)	/
Opportunity and comfort penalties ($O_{i,t}$)	• Lost export revenue, $O_{i,t} = 0.08$ €/kWh
Forecast uncertainty cost ($R_{i,t}$)	• Baseline error factor $\sigma_{PV} = 0.1$ • Scaling factor $\gamma_{PV} = 0.036$ €/kWh
Fixed and administrative costs (F_i)	• Administrative term $F_{PV} = 0.01$ €/kWh

3. **BESS** provides flexibility in both directions

- **Technical parameters**:
    - Rated capacity: 100 kWh
    - Charge/discharge power limit: 30 kW
    - Efficiency: 95%
    - Initial SoC: 60%
    - SoC limits: 20%-95%
- **SRMC parameters** are presented in Table 7.

Table 7: Practical example, SRMC parameters for BESS asset

Term	Parameters
Energy price differential ($\Delta\lambda_{i,t}$)	• Current DA buy price λ_t^{buy}
Degradation and wear ($A_{i,t}$)	• Degradation $c_{deg} = 0.03$ €/kWh
Opportunity and comfort penalties ($O_{i,t}$)	• Look-ahead window N=16 • Percentiles: low=0.1 (buy), high=0.9 (sell)
Forecast uncertainty cost ($R_{i,t}$)	/
Fixed and administrative costs (F_i)	• Administrative term $F_{BESS} = 0.01$ €/kWh

4. **HVAC** represents a thermally flexible load, which is able to change consumption relative to the baseline, providing both upward and downward flexibility.
    - **Technical parameters**:
        - Mode: cooling (external temperature profile included as input)
        - Comfort band: 21-24 ˚C
        - Thermal coefficients: reversion factor $a = 0.1$, temperature effect $b = 0.5\,°C/slot$ at full power.
        - Nominal electrical power: 15 kW
        - Baseline profile provided for the simulated period (peaking at 12 kW)
        - Limits for short-cycling: the device can be toggled ON/OFF twice per hour.
    - **SRMC parameters** are presented in Table 8.

Table 8: Practical example, SRMC parameters for HVAC asset

Term	Parameters
Energy price differential ($\Delta\lambda_{i,t}$)	• $-\lambda_t^{buy}$ (up) • λ_t^{buy} (down)
Degradation and wear ($A_{i,t}$)	/
Opportunity and comfort penalties ($O_{i,t}$)	• comfort weight $\kappa = 0.08$ €/kWh • Soft band half-width 0.5 ˚C • Outside-band adder 0.20 €/kWh • Reward factor 1.0
Forecast uncertainty cost ($R_{i,t}$)	• Baseline error factor $\sigma_{HVAC} = 0.12$ Scaling factor $\gamma_{HVAC} = 0.036$ €/kWh
Fixed and administrative costs (F_i)	• Administrative term $F_{HVAC} = 0.01$ €/kWh

The algorithm operates with a time resolution of 15 minutes and a rolling horizon of four slots (one hour). A non-delivery penalty of 50 €/kWh is applied to ensure full delivery when feasible. The main genetic algorithm parameters are:

- random seed = 42,
- population size = 30,
- generations = 40,
- crossover probability = 0.9,
- mutation probability = 0.2, and
- tournament size = 0.3.

At each iteration, the optimizer determines the direction of required flexibility (up or down), generates the SRMC matrices and searches for the least-cost feasible combination of assets over the horizon, Only the first step of the plan is executed, after which the window moves forward and all asset states are updated.

Figure 9 below shows the total activations of all assets over the simulated horizon. The solid and dashed black lines represent the required upward and downward flexibility targets, respectively, while the coloured bars display the contributions of individual assets. Positive values indicate upward flexibility (increased generation or reduced consumption), whereas negative values represent downward flexibility (increased consumption or curtailed generation).

The flexibility request begins with an upward activation phase, switches to downward activation in the middle of the horizon, and returns to upward flexibility in the final slots. The BESS provides the largest share of flexibility in both directions, consistent with its technical capability of delivering up to ± 30 kW. During upward activation, additional support is provided by the EV site, which reduces charging, and the HVAC system, which temporarily curtails cooling. During downward activation, the BESS absorbs energy through charging, complemented by periods of PV curtailment in the beginning and increased HVAC consumption.

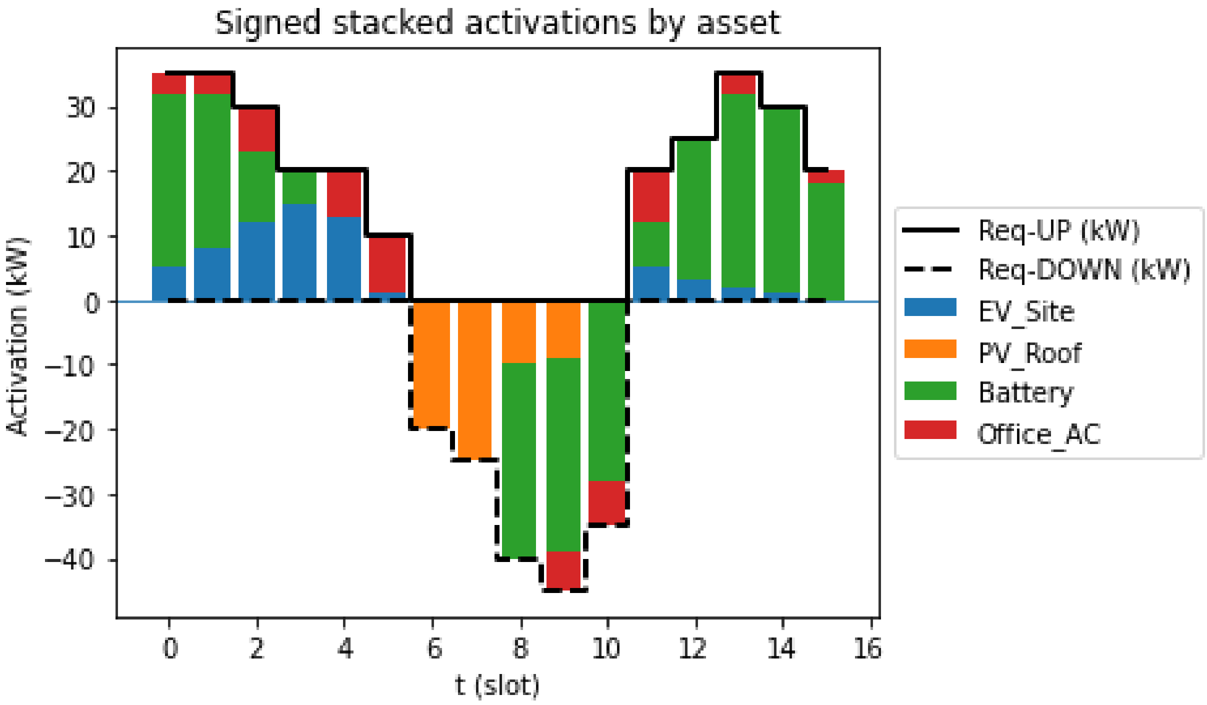

Figure 9: Algorithm results - signed stacked activations by asset

The operation of the BESS over the same horizon is shown in Figure 10. The blue bars denote activation power (positive = discharge, negative = charge), while the green line tracks the SoC. The red dashed lines mark the lower and upper SoC bounds.

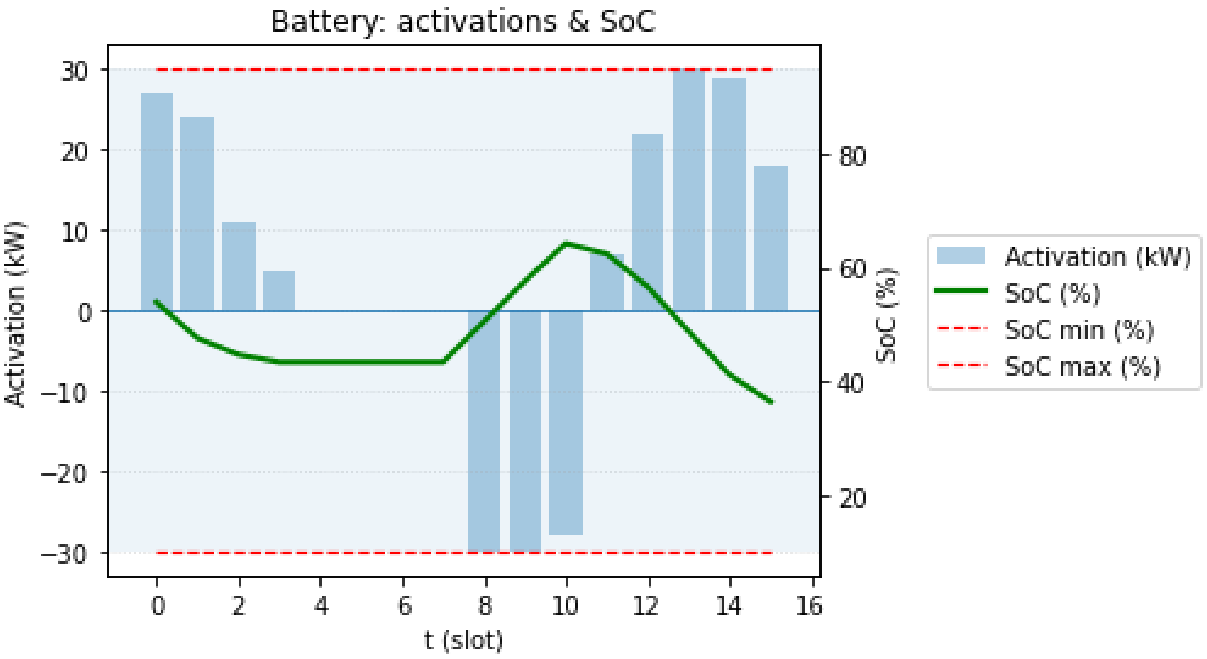

Figure 10: BESS activations and SoC evolution

In the first phase, the battery discharges to provide upward flexibility, gradually reducing its SoC. Once the direction of activation changes to downward flexibility, the BESS charges and its SoC increases. In the final upward phase, the BESS discharges again, returning toward its initial SoC. The

SoC remains consistently within the defined limits, confirming that the algorithm enforces feasibility at each step.

Figure 11 illustrates the operation of the HVAC unit. The blue line shows the baseline electrical consumption, the orange line the optimized setpoint, and the blue bars the flexibility activations (positive for curtailment, negative for boost). The green line represents indoor temperature evolution, bounded by the comfort limits shown in red dashed lines.

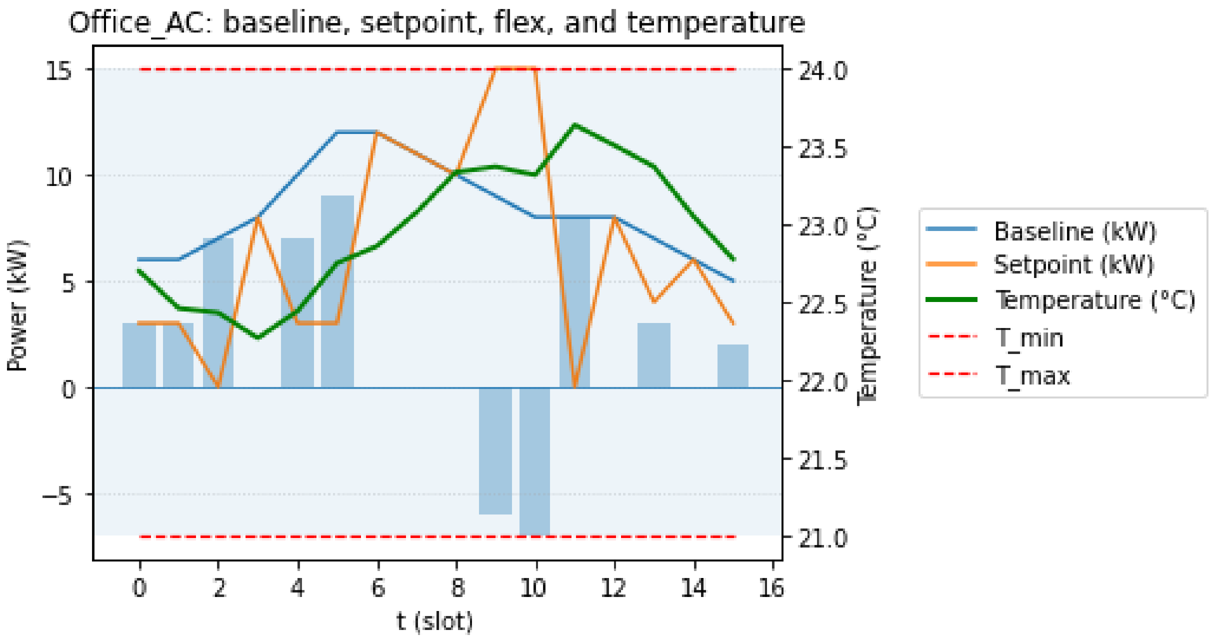

Figure 11: HVAC baseline, setpoint, flexibility, and indoor temperature

When upward flexibility is required, HVAC power is curtailed, causing a rise in indoor temperature. When downward flexibility is required, additional cooling is applied, leading to a temperature decrease. Throughout the entire period, the indoor temperature remains within the comfort band, demonstrating that the one-step thermal preview successfully restricts feasible activations to prevent comfort violations.

Figure 12 shows the SRMCs calculated dynamically for each asset and interval, separated into upward and downward flexibility directions. These SRMCs integrate energy market prices with the predicted states of each asset, determining the relative attractiveness of activation.

In the upward flexibility subplot, three assets are shown: the EV charging site, the BESS, and the HVAC unit. The EV site and BESS exhibit similarly shaped SRMC curves, as both combine the immediate energy price effect with a forward-looking opportunity term that reflects the cost of recharging later. The EV site maintains slightly lower SRMC values across the horizon, primarily because its forecast uncertainty component is smaller than the BESS degradation term. The HVAC SRMC shows stronger variation, rising as the indoor temperature approaches the upper comfort limit and the comfort penalty increases. In the early slots, the algorithm relies more on HVAC and EV curtailment due to their lower costs, but as the required flexibility volume is high, it fills the remaining portion with BESS activation. Toward the end of the horizon, when the BESS SRMC drops below that of the HVAC, the algorithm again prioritizes BESS for upward flexibility provision.

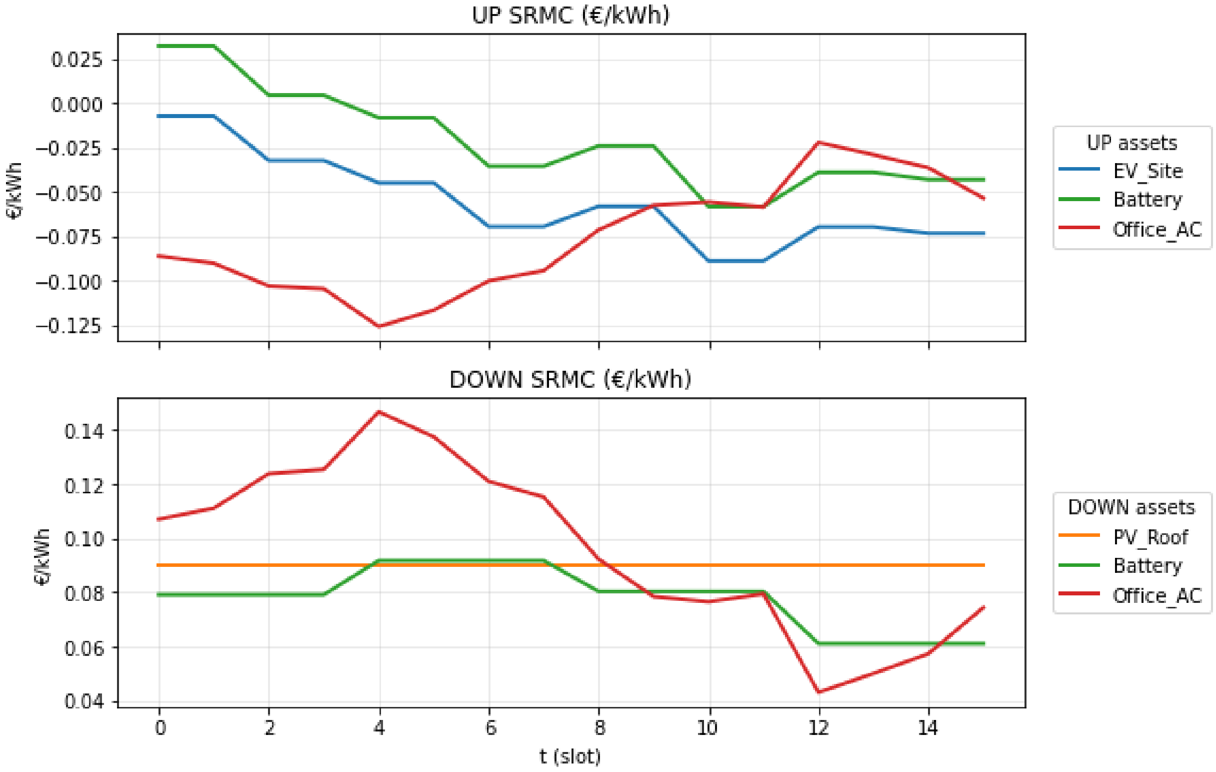

Figure 12: SRMC values for upward and downward flexibility by asset

In the downward flexibility subplot, the PV system maintains a constant SRMC equal to the feed-in tariff, since each curtailed kilowatt-hour represents a fixed opportunity loss. The BESS and HVAC curves, however, show more dynamic behavior. For the BESS, SRMC variations are primarily driven by the combined effect of the instantaneous energy term and the future opportunity term. The BESS SRMC remains mostly below those of PV and HVAC, crossing the PV level between time steps 4–7, when short-term price dynamics temporarily reduce its relative attractiveness. The HVAC SRMC, meanwhile, declines around step 7 as additional cooling becomes beneficial for restoring comfort, making downward activation more appealing.

This simplified test case demonstrates the end-to-end functioning of the optimal selection algorithm on a manageable example where results can be easily interpreted. The behavior of each asset type and the resulting dispatches align with the expected economic and technical logic: SRMC-driven prioritization, state-dependent feasibility, and comfort constraint enforcement.

An important observation from the test is the strong influence of the SRMC parameters (especially those defining degradation and forecast uncertainty terms) on how assets are prioritized. In practice, these parameters would need to be carefully calibrated and continuously updated by the aggregator or FSP to reflect real operating conditions, asset reliability, and market context. For instance, regularly adjusting the forecast uncertainty term based on recent performance would naturally steer the algorithm toward the most reliable and predictable resources, improving overall delivery quality.

In future work, the algorithm will be evaluated on larger and more complex portfolios that include a greater variety of assets. Special attention will be given to parameter sensitivity, computational scalability, and execution time, to assess how the algorithm performs as problem dimensionality grows. Evaluating the trade-off between solution quality and computational effort will be essential to determine its suitability for real-time deployment. These tests will ensure that the algorithm remains both robust and practical for real-world flexibility management and market participation.

4 Conclusions

This document presents the work performed during the second phase of **Task 4.4: Optimal Selection of Available Flexibility**, focusing on the development, implementation, and demonstration of algorithms that support aggregators and FSPs in optimizing the use of distributed flexibility resources. Building upon the forecasting and flexibility modelling work from the first phase, the activities described in this report complete the methodological framework by introducing the decision-making tools that enable practical utilization of flexibility in real or simulated market environments.

The first part of the work focused on the **Optimal Market Selection Algorithm**, which evaluates the participation potential of distributed flexibility resources across different market environments. The algorithm quantifies the economic benefits of participating in various market products, including day-ahead, intraday, and ancillary service markets, while considering technical limitations and regulatory requirements. The results obtained using real data from AVANTCAR's EV charging stations confirm that strategic market participation, particularly the combination of day-ahead optimization with ancillary services, can enhance aggregator revenues while maintaining operational feasibility.

As a complementary component of this framework, a **tariff-optimization sub-algorithm** was developed in collaboration with Amibit. This module demonstrates how similar optimization principles can be applied at the household level through the Reduxi HEMS platform to reduce electricity costs and avoid power-exceedance penalties under the new Slovenian dynamic network-tariff system.

In the second part the **Optimal Selection Algorithm** is presented, which addresses the internal activation process of an aggregator once a flexibility request has been received. Implemented as a **rolling-horizon genetic algorithm**, it identifies the least-cost feasible combination of resources that can deliver the requested flexibility while ensuring compliance with all technical and operational constraints. The algorithm integrates detailed models of asset behavior and cost through short-run marginal cost (SRMC) formulation, allowing dynamic and cost-based coordination of heterogeneous flexibility resources. A demonstration case with a simplified portfolio confirmed that the algorithm operates as expected, efficiently balancing technical feasibility, cost efficiency, and adaptability.

Together, these developments form a comprehensive framework for the optimal management of distributed flexibility—from identifying profitable market opportunities to determining how individual assets should be activated and how household-level flexibility can be leveraged to deliver value to both users and the grid.

5 References and acronyms

5.1 References

[1] Republika Slovenija, «Pravno-informacijski sistem RS, Akt o metodologiji za obračunavanje omrežnine za elektrooperaterje,» [En línea]. Available: https://pisrs.si/pregledPredpisa?id=AKT_1266.

[2] SciPy, «linprog,» [En línea]. Available: https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.linprog.html.

[3] «Reduxi AI Electricity,» [En línea]. Available: https://www.reduxi.eu/?_gl=1%2A7l18xe%2A_up%2AMQ..%2A_ga%2ANjM1OTgxNDYwLjE3NjE3NjcxNTE.%2A_ga_J91VNMJVTF%2AczE3NjE3NjcxNTAkbzEkZzAkdDE3NjE3NjcxNTAkajYwJGwwJGgw.

[4] European Union, «Directive (EU) 2019/944,» 2019. [En línea]. Available: https://eur-lex.europa.eu/eli/dir/2019/944/oj/eng.

[5] V. Rasouli, A. Gomes y C. H. Antunes, «A Genetic Algorithm Approach for Aggregation of Residential Electricity Prosumers' Flexibility,» *Applications of Evolutionary Computation: 28th European Conference, EvoApplications 2025, Held as Part of EvoStar 2025, Trieste, Italy, April 23–25, 2025, Proceedings, Part I,* 2025.

[6] P. Huang, T. Xu y Y. Sun, «A genetic algorithm based dynamic pricing for improving bi-directional interactions with reduced power imbalance,» *Energy and Buildings,* vol. 199, pp. 275-286, 2019.

[7] E. González-Romera, E. Romero-Cadaval, C. Roncero-Clemente, M.-I. Milanés-Montero, F. Barrero-González y A.-A. Alvi, «A Genetic Algorithm for Residential Virtual Power Plants with Electric Vehicle Management Providing Ancillary Services,» *Electronics,* vol. 12, 2023.

[8] M. Sarstedt y L. Hofmann, «Monetarization of the Feasible Operation Region of Active Distribution Grids Based on a Cost-Optimal Flexibility Disaggregation,» *IEEE Access,* vol. 10, 2022.

[9] G. G. Soma, G. M. Tina y S. Conti, «Optimal Bidding Strategies for the Participation of Aggregators in Energy Flexibility Markets,» *Energies,* vol. 18, nº 11, 2025.

[10] C. Calvillo, A. Sánchez-Miralles, J. Villar y F. Martín, «Optimal planning and operation of aggregated distributed energy resources with market participation,» *Applied Energy,* vol. 182, 2016.

[11] F. L. Müller, J. Szabó, O. Sundström y J. Lygeros, «Aggregation and Disaggregation of Energetic Flexibility From Distributed Energy Resources,» *IEEE Transactions on Smart Grid,* vol. 10, 2019.

[12] D. Danner, R. Huwa y H. De Meer,, «Multi-objective flexibility disaggregation to distributed energy management systems,» *SIGENERGY Energy Inform. Rev.,* vol. 2, 2022.

[13] S. C. S. &. K. Katoch, «A review on genetic algorithm: past, present, and future,» *Multimed Tools Appl,* vol. 80.

[14] Opentunity Grant Agreement. EC, 2022.

5.2 Acronyms

Table 9: Acronyms

Acronym	Meaning
aFRR	automatic Frequency Restoration Reserve
BESS	Battery Energy Storage System
DA	Day-Ahead
DER	Distributed Energy Resources
DNN	Deep Neural Network
DSO	Distribution System Operator
EV	Electric Vehicle
FSP	Flexibility Service Provider
GA	Genetic Algorithm
HEMS	Home Energy Management System
HP	Heat Pump
HVAC	Heating, Ventilation and Air-Conditioning
ID	Intraday
mFRR	manual Frequency Restoration Reserve
PV	Photovoltaic
SRMC	Short Run Marginal Cost
TSO	Transmission System Operator