UNIVERSITA' DEGLI STUDI DI PARMA

Dottorato di ricerca in Ingegneria Industriale

Ciclo XXVIII

DEVELOPMENT AND APPLICATION OF METHODOLOGIES FOR THE OPTIMIZATION OF ENERGY SYSTEMS

SVILUPPO ED APPLICAZIONE DI METODOLOGIE PER L'OTTIMIZZAZIONE DI SISTEMI ENERGETICI

Coordinatore:

Chiar.mo Prof. Ing. Marco Spiga

Tutor:

Chiar.mo Prof. Ing. Agostino Gambarotta

Dottorando:

Ing. Nicola Pompini

To Francesca.

ACKNOWLEDGMENTS

Many people helped me and inspired me during my whole experience as a Ph.D. student and I am grateful to all of them. First of all, I would like to thank my advisor, Prof. Dr. Agostino Gambarotta, for his continuous support, for the great opportunities that he gave me, and for having been a guide to me throughout these years. Next I wish to thank Dr. Marcello Canova, for the great pleasure and the profound professional enrichment that I attained by working with him, and for his kindness and friendship. Further, I am thankful to Prof. Paolo Casoli, for the confidence he expressed in me since he was my master thesis's advisor and mentor in my student career.

I want to deeply thank all the people from the Industrial Engineering Department for the great working environment they contributed to create. I would particularly like to express my gratitude to Dr. Mirko Morini, for his many valuable advices and for his passion for the engineering work, which has been the source of many stimulating conversations, Luca Riccò, for having been a precious colleague and friend, and Marco Crialesi, for all the technical (and non) inspiring discussions that we had.

The time I spent at the Center for Automotive Research in Columbus as a visiting scholar was extremely enjoyable and inspiring, I feel privileged for the occasion I had and I would like to thank all the many people who made that possible, especially my colleagues Eric Lott and Junqiang Zhou, for the fruitful discussions we had during the time we worked on the project, and for the many other pleasant moments that we shared.

I would like to express my overwhelming gratitude to my parents, for their unlimited support and their encouragement during these years, and for always being supportive of my choices. Finally, my deepest thanks go to Francesca, for her love and great patience, for inspiring me with her talent, and for all the great and hard times we have gone through together. Without her, this experience would not have meant the same.

ABSTRACT

Fossil fuels consumption reduction and the development of energy saving technologies are becoming a central topic for both the industry and the academic word, due to the drastic effects of anthropogenic emissions on the environment. As an increasing number of measures and regulations are being issued to cope with such issue, the development of low emission technologies is guiding the research in many different industrial sectors. While the realization of clean carbon free energy sources is regarded as one of the most promising solution in the long term, it is widely accepted that for a short-medium horizon an efficient and effective integration of such technologies is highly infeasible, due to technical limitations and the magnitude of the energy share, actually provided by traditional fossil sources, that these alternative technologies should cover. The optimization of the energy production and management processes, associated to the development of energy consumption reduction technologies, is on the other hand regarded as an adequate solution that can be more easily exploited to promptly cope with the problem.

The objective of this thesis is to investigate, develop, and apply a set of numerical tools for the optimal design and management of energy processes, which can be exploited to achieve a minimization of the fuel consumptions and an increase of the energetic efficiencies. The devised methodologies rely on a model based approach, which takes advantage of the predictions ability deriving from a mathematical representation of the examined system, to provide a framework for the estimation and optimization of the system behavior under realistic operating conditions. When optimizing the systems, particular emphasis has been given to the necessity of deriving a proper management strategy, accounting for the dynamic properties of the examined plant, which is necessary to achieve the best performance during the effective operating phase.

Throughout the thesis, the problem of energy efficiency optimization is investigated with reference to three different technological applications. In the first one, a multi-source plant for the fulfillment of the energetic demand of a building is considered. As the plant exploits multiple different technologies for the provision of the

required electric and thermal power, it is necessary to derive a proper scheduling policy, determining how the loads have to be effectively divided between the different sources, in order to obtain the maximal plant efficiency. Based on a simplified plant model, the problem is efficiently solved by applying the deterministic Dynamic Programming algorithm, and the results are compared to those attained by the adoption of a simpler rule-based policy, proving the advantages deriving from the adoption of an optimal control strategy.

In the second application, the design of a hybrid solution for energy recovery from a hydraulic excavator is investigated. As different plant technological layouts may be conceived and the additional components introduced require to be properly sized, a methodology to evaluate the benchmark potentiality of each different solution needs to be derived. The comparison between the different layouts is based on the predicted performance of the machine during a standardized digging duty cycle, which are estimated with the help of a detailed plant model. As the introduction of energy recovery devices introduces additional degrees of freedom to the system, it is necessary to derive the optimal management strategies for such devices in order to derive the maximum attainable performance from each layout solution. This task is again carried out with the help of the deterministic Dynamic Programming algorithm, which exploits a control oriented simplified model of the plant, designed for the sake of the optimization. Once the best achievable performance for each design solution is obtained, it is possible to carry out a fair comparison between the available alternatives.

In the third and final application, an Organic Rankine Cycle plant for the Waste Heat Recovery from the exhausts from a light vehicle application is investigated. The ORC plant has the potential to deliver considerable increases in the overall vehicle efficiency but it requires the development of a complex control algorithm, which must be able to comply with the variability of the energy source for the process. Moreover, the maximization of fuel economy must be carried out while keeping the plant in safe operating conditions. A robust and efficient system model, based on the Moving Boundary Methodology, has been developed to simulate the performance of the plant and to account for the effect of the phase changing in the used fluid. This model has been subsequently used to design a Model Predictive Controller which estimates the optimal control inputs for the system, in order to achieve the desired performance. An original optimization algorithm, based on the Particle Swarm Optimization, has been conceived to solve the connected nonlinear dynamic optimization problem. The results obtained from the adoption of the devised controller are compared to those

that can be reached with a classic PI based controller, showing again the advantages, from an energetic efficiency point of view, deriving from the adoption of an optimal control strategy.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

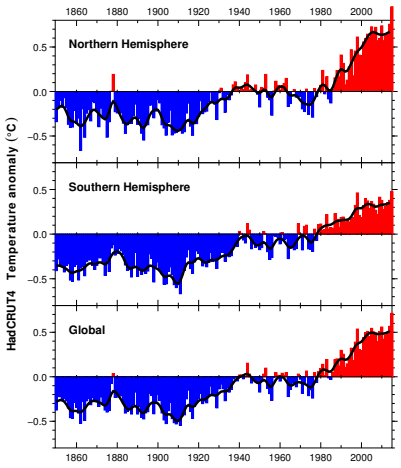
CHAPTER 1
INTRODUCTION

Today the world is at a critical juncture regarding the ongoing effects of global climate changes. The rapidity of the current modifications in climatic conditions represents a peculiar aspect of today's situation and has raised many concerns regarding the long term effects that it could have on the global ecosystem. Numerous mutations that can be attributed to climate change have already been detected in biotic and abiotic environment, from hydrology, coastal processes, marine and freshwater biological system, to agriculture, forestry, and terrestrial biological systems, based on more than 30 years of studies (see [6] for a comprehensive review). Even more drastic effects are expected to happen in the near future due to the known time lags in ecosystem response to climate changes, implying greater risks of overwhelming impacts during the twenty-first century [4].

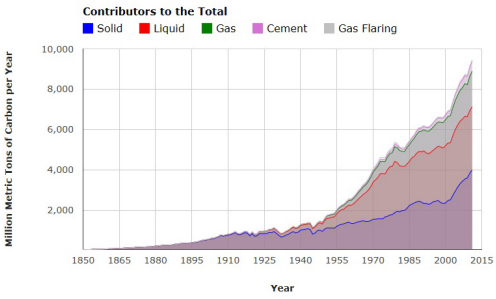
1.1 Global warming and the role of anthropogenic emissions

If the trend in average global temperature is analyzed since the beginning of the available records (see Figure 1.1a), it is possible to detect two relatively rapid and steady warming periods, the first from about 1910 through the early 1940s, the second starting by the mid-1970s and still ongoing. Just to provide a numerical reference for this trend, the data collected showed that the year 2014 has been the warmest one in the entire global record history, with an average temperature 0.56 °C above the 1961-1990 reference period mean [9]. Furthermore, the coldest year of the 21st century, 2008, was still warmer than all the years in the 20th century, with the only exception of 1998. While these changes have been rather common in the past, the effects on the environment were always mitigated thanks to the retention of a quasi-equilibrium condition between ecosystem distribution and climate, a process possible due to the much longer periods (in the order of geological timescales) associated to such drastic changes [4].

While many different effects actively contribute to determine the global temperature, it became widely accepted in the scientific community that the increase of greenhouse gases concentration in the atmosphere, a process related to many human activities (from mobility to industrial processes and residential needs), is one of the dominant causes of climate changes. To this regard, the trends in fossil carbon related emissions are depicted in Figure 1.1b, showing the substantial increase happened during the last 50 years. While the greenhouse effect has been known for about 300 years, it was only during the 1980s that scientific evidences began to establish a correlation between anthropogenic related greenhouse gases (GHG) emissions and global warming. Since then, numerous contribution have been submitted on the subject and today this position is widely accepted, with the Intergovernmental Panel on Climate Change (IPCC) stating in its last report that "Human influence on the climate system is clear. This is evident from the increasing greenhouse gas concentrations in the atmosphere, positive radiative forcing, observed warming, and understanding of the climate system." [5].



(a) Annual temperatures anomalies, 1850-2014 [9]



(b) CO_2 global anthropogenic emission [8]

Figure 1.1: Global climate changes and GHG emissions

Starting from the first claims of correlation between human activities and climate changes, many steps have been taken from world governments to address the problem of global warming and, even if the initial political measures where rather feeble,

the increasing awareness of the profoundly dangerous effects of climate mutations is driving the most recent political actions towards more compelling resolutions. Since the ratification in 1992 of the United Nations Framework Convention on Climate Change (UNFCCC), many steps were taken to address the problem of dangerous anthropogenic interferences over the climate system; while the treaty itself did not enforce any legal bindings on GHG emissions, it still provided the framework for the development of specific international protocols and today, with 196 ratifiers, it represents the largest international climate policy venue. Since 1995, the parties of the convention began to met annually at the so-called Conference of Parties (COP) and in 1997, with the adoption of the "Kyoto Protocol", a first legal binding committing the ratifiers governments to the development of a greenhouse gases emission reduction was signed. More recently, as part of the package of decisions taken at COP16 in Cancun, the Parties formally recognised what already emerged during COP15 works, stating that urgent measures to hold the increase of average global temperature below 2 °C (relative to pre-industrial levels) must be taken, for the sake of limiting the dreadful effects on the environment, and that a deep cut in GHG emissions is required to achieve that. Regarding this issue, the IPCC has estimated that, in order to preserve a 50% chance of maintaining global warming below the 2 °C limit, the world can support a maximum "emission budget" of $\sim 3000\ Gt$ (gigatonnes) of CO_2 [7] for the rest of the 21st century. With an estimated 2000 Gt already emitted before 2014, a carbon budget of $\sim$ 1000 Gt is left from the start of 2014 onwards, a number underlining the great importance of the forthcoming measures to effectively enforce the success of such a resolution.

1.2 Current energy scenario

A quick survey of the current energy scenario is given here, to provide a reference framework for the various pledges and measures that can be taken to reduce the impact of GHG emissions. Focusing on global primary energy consumption, a net growth has been registered for all kind of fuels in 2014, reaching records levels for all sources except that nuclear. Even if in the last few years the growth in global primary energy consumption was below the past decades average, with a registered +0.9 % in 2014 vs. a 10 years average of +2.1 % [2], the emerging economies still determined a massive growth in global energy consumption, with a +2.6 % increase from China and a +7.1 % from India. While the rate of the intensification from developing countries has been weaker than in the past, it is still amply adequate to compensate the reductions registered in the other more developed countries, like EU, registering

a fall of -3.9 % in its primary consumption, or Japan, with -3.0%. Furthermore, as this reductions were partly correlated to the economical crisis affecting western economies since 2008, the effects of the ongoing apparent economic recovery on first world energy consumption still needs to be assessed.

Focusing on prime energy sources, data from the British Petroleum statistical review of June 2015 [2] show that global oil consumption is still growing (+0.8 %), even if a little below its recent historical average, while the growth of natural gas has been well below its 10 year average (0.4 % vs 2.4 %), with declines in both OECD and emerging economies (with a declining record of -11.6 % in EU). Apart from the exception of India (+ 11.1 %), coal consumption experienced its weakest grew outside OECD countries, with just a low +1.1 %, mainly due to the flattening in Chinese consumption. Nuclear power grew by an above average +1.8 %, while global hydroelectric remained below its 2.0 % average, with the the only exception of China (+15.7 %). Finally, despite the lower fossil-fuel prices, the increase trend in renewable energy consumption continued, with global investments of $270 billion and an increase of +12 %, bringing renewable sources to account for a 6.0 % of the global power generation. Figure 1.2 depicts the evolution of overall energy consumption from the 1990s, based on the different primary sources.

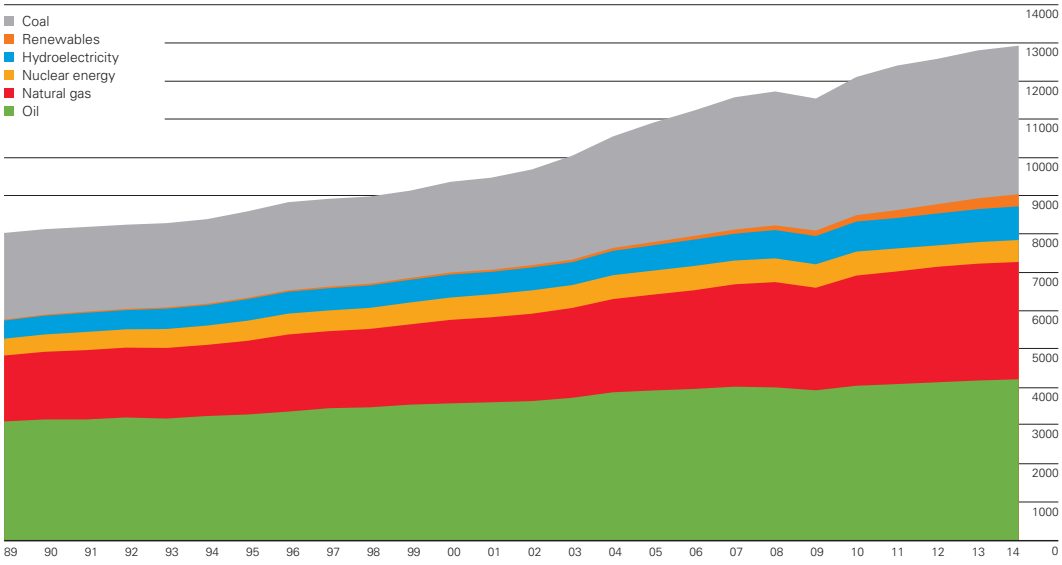


Figure 1.2: World primary energy consumption [Mtoe] [2]

As today, it is evident that fossil fuels still represent the largest contributors for the satisfaction of primary energy demand, covering the 80 % of total share, and accounting for 90 % of the energy related GHG emissions. A somewhat promising trend may however be noticed if looking at the stall of energy-related CO_2 emissions in 2014. Even with a growth in world economy of about +3 %, the total GHG emissions stalled in fact to a value of 32.2 Gt, unchanged from the previous year. This decoupling between the economic growth (with a corresponding increase in energy consumption) and the overall GHG emissions, represents an unusual event, as in the last 40 years the reductions in global emissions has always been tied to economic crisis. This effect has been justified by the increased deployment of renewable sources and the enhanced efforts to increase energy production and distribution efficiencies [3]. Another good indicator of this trend may be identified in the global energy intensity factor, defined as the amount of energy required to produce a unit of Gross domestic Product (GDP), which showed a decrease of -2.3 % relative to 2013, as a result of the emissions reduction policies adopted all over the world.

Based on IPCC 2014 data [7], the primary energy consumption is distributed as follows between the different economic sectors:

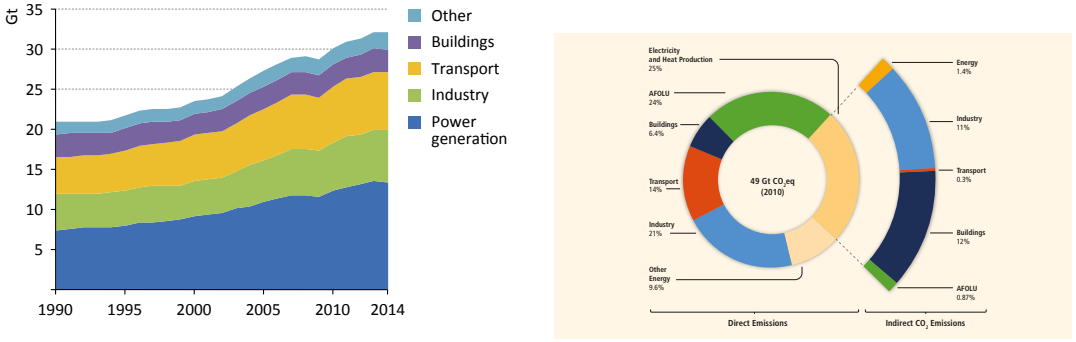
- **Transport**: The transport sector accounted for 27 % of the final energy use, generating 6.7 Gt_{CO_2eq} (gigatonnes of equivalent CO_2) of direct GHG emissions in 2010. The baseline CO_2 emissions from the vehicle sector are estimated to approximately double by 2050, depending on the urban growth and the investment in public transport system;

- **Building sector**: Residential and commercial energy consumption accounted for about 32 % of the overall energy use in 2010, with 8.8 Gt_{CO_2eq} generated considering both direct and indirect emissions. An estimated baseline increase of 50-150 % is expected by mid-century, resulting from improvements in wealth, changes in lifestyle, and the urbanisation process;

- **Industry**: The industrial processes accounted for 28 % of the global energy use, generating 13 Gt_{CO_2eq} of direct and indirect emissions, and represented the 30 % of global GHG emissions, surpassing both the building and transport end-use sectors; unless energy efficiency improvements are adopted, the baseline emissions associated to industrial processes are also expected to grow in the range of 50-150 % by 2050;

- **AFOLU sector**: The Agriculture, Forestry and Other Land Use accounted for about a quarter of the net anthropogenic GHG emissions ($\sim$ 10-12 Gt_{CO_2eq}),

mainly due to deforestation and agricultural practices. While uncertainties on AFOLU emissions estimations are higher than those for other sources, the annual baseline CO_2 production is expected to experience a net decline, with net emissions potentially reduced to less than half the level of 2010 by the year 2050, thanks to increased afforestation policies.



(a) GHG emissions by sector trends [3]

(b) GHG anthropogenic emissions detail (2010) [7]

Figure 1.3: Greenhouse gas emissions by sectors

1.3 Pledges for energy efficiency improvement and emission reduction technologies

From a primary energy production point of view, the increase of renewable, carbon free energy solutions alone is not expected to be sufficient to achieve the emission reduction targets. While great growths are expected from sources like wind power, solar photovoltaic, geothermal, and biomasses, the increase of global energy demand and the centrality of flexible and cheap fossil-based technologies in today energy production will prevent a significant switch towards renewable sources. Figure 1.4 illustrates this principle based on the predictions from the U.S. Energy Information Administration (EIA) [1] about primary energy consumption for the following 25 years.

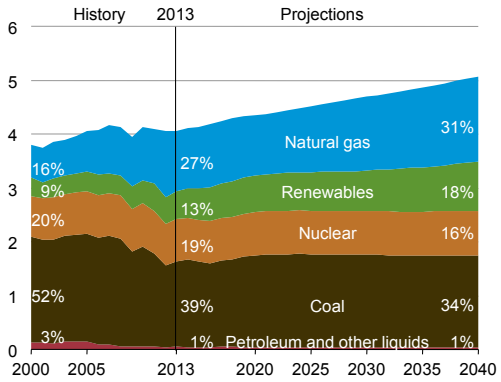


Figure 1.4: U.S. predicted primary energy consumption by source [quadrillion Btu] [1]

To additionally reduce the GHG related environmental impact from the different economic sectors, further measurements must therefore be taken beside the adoption of renewable sources. As an example of such measures, the Intergovernmental Panel on Climate Change issued the following guidelines in its 2014 report [7]:

- **Transport**: A 40 % below the baseline predicted growth in transport related emissions can be achieved if technical and behavioural mitigation measures are taken for all transportation modes and new urban infrastructures are developed. The development of new technologies is expected to yield and increase of vehicles efficiencies and performance of about 30-50 %. Moreover, the switch towards low-carbon fuels is expected to grow over time, with an increase in methane based fuels vehicles, the adoption of electricity from low-carbon sources to drive electric trains, buses, and road vehicle applications, and the commercialization of liquid and gaseous biofuels. Integrated urban planning and transit oriented development may also lead to considerable reduction in transport related emissions;

- **Buildings**: The recent advances in technologies and know-how may provide an opportunity to stabilize the impact of the building sector over global GHG emissions. The development of portfolios of energy efficiency policies and their implementation proves to be the most environmental and cost-effective solution for the reduction of emissions;

- **Industry**: The energy consumption from the industrial sector may be reduced

by about 25 % if adopting wide-scale upgrades, replacing the actual low efficiency sources with the best available technologies. The improvement of material use efficiency and recycling policies are also expected to lead to a consistent reduction in overall demand. Energy recovery from waste heat sources to reduce the fossil fuels demand can lead to additional significant emission contractions;

- **AFOLU**: The agricultural and forestry sector plays a central role for the sake of sustainable development. The most cost-effective solutions available are afforestation and sustainable forest management strategies. Moreover, the development of bioenergy technologies can contribute to a reduction of the GHG emissions from the other sectors, by providing low-carbon alternatives to fossil-based fuels.

To derive a realistic estimation of the intended world governments commitments towards the reduction of GHG emissions, it is possible to refer to the pledges submitted prior the United Nations Climate Change Conference (COP21, France, Nov. 2015), known as the Intended Nationally Determined Contributions INDC, which indicates the efforts that these countries are willing to put into the energy-related CO_2 abatement policies. Based on estimation from the International Energy Agency [3], while the adoption of such policies will have a positive impact in slowing the growth of energy related GHG emissions, these efforts will still be inadequate to meet the target of 50 % probability of lower than 2 °C temperature increases by 2100. See Figure 1.5 for more details, where the INDC scenario is compared with a so-called "450 scenario", based on the concept that, in order to reach the 2 °C target, the concentration of CO_2 in the atmosphere must not exceed the limit of 450 ppm.

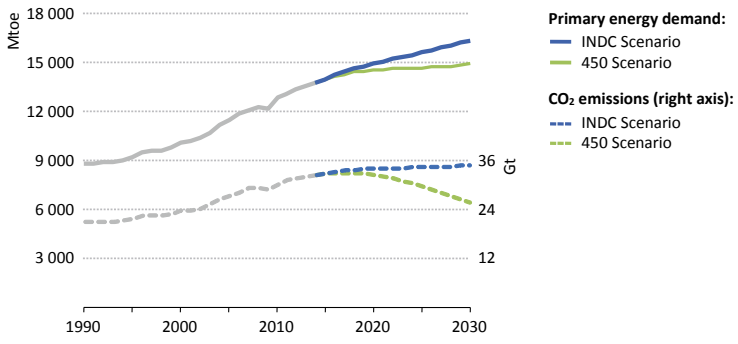


Figure 1.5: Comparison between the INDC scenario and the 2 °C target scenario [3]

Based on the INDC scenario, no peak for CO_2 emissions is in sight, and, following the IPCC estimations, the world would exhaust its "carbon budget" within the year 2040. As the estimated global GHG emissions by 2020 based on the Cancún Pledges and the INDC commitments are not consistent with long-term mitigation trajectories, the IPCC suggested the introduction of a series of additional efforts before 2030. One of the first objective they propose is to increase the speed of the already happening decarbonization process in electricity generation, with a combination of renewable energies, nuclear, and carbon capture and storage (CCS) technologies. Furthermore, the International Energy Agency proposed a number of feasible near-term strategies to achieve a peak in GHG emissions by 2020, still maintaining the same level of economic growth and development prospected in the INDC scenario. Among these policies (which estimated impact is illustrated in Figure 1.6 under the name of "bridge scenario"), the largest contribution to global GHG abatement is supposed to come from a great increase in energy efficiency. This strategy requires the adoption of minimum energy performance standards (MEPS) for both the industry and buildings sectors, along with the introduction of more efficient measures to cope with the heating and cooling processes (heat pumps, heat recovery, etc.). Regarding the transport sector, the imposition of new stricter standards on global average fuel consumption is supposed to ease the vehicle related impact to GHG emissions.

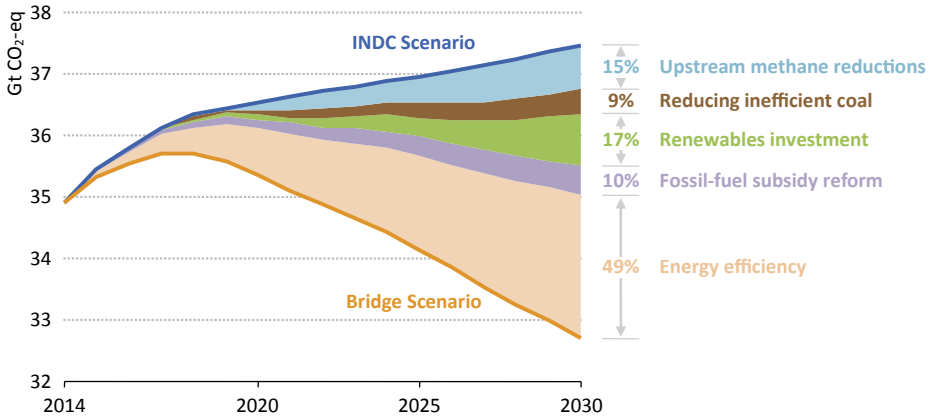


Figure 1.6: An estimated roadway to achieve peak CO_2 emissions in 2020 with near-term strategies [3]

1.4 Complex energy systems design and management

With the stricter standards that are being imposed worldwide on carbon emissions and the prospect of even greater forthcoming limitations, great changes are expected to happen in the global energy scenario, with effects spreading into all the sectors connected to GHG emissions, from electric power generation to the automotive world, from industrial production to the building and residential sector. Moreover, it is interesting to notice how many sources estimates that the largest contribution to GHG reduction in a medium horizon reference framework are expected to derive from the adoption of energy efficiency increase measures.

Although many technologies have been already developed to achieve greater energetic efficiencies or to provide "carbon-free" alternatives, their correct integration in the existing energy scenario still poses challenging issues. From an engineering point of view, these issues are related to the many phases of the implementation process, from the correct design of the applications to the development of effective management strategies able to optimize the energetic efficiency gains while maintaining unaffected (or even improved) the functionalities of the existing technologies.

For such reasons, tools and methodologies for an efficient design and management of complex energy systems have draw the attention of both the industrial and the academic worlds, defining the research direction.

1.5 Objective and structure of the thesis

Based on the previous considerations, the present thesis focuses on the application and development of numerical methods able to provide an engineering solution to the problem of efficiency optimization of complex energy systems, with the final objective of reducing their overall energy consumption. The adopted methodology relies on the development of physical based mathematic representations of the analyzed process, which are used to provide a reference framework for the evaluation of plant energetic performance. Once such models are obtained, the abstract energy optimization problem is converted into a mathematical one, identifying both the parameters that can be modified to achieve the desired performance and the restrictions that may limit such choices. Finally, the proper numerical optimization techniques are applied to obtain a solution of the conceived problem.

In the first part of the thesis, the topic of model based numerical optimization is introduced. Different types of plant models and representation are described, based on the level of detail required and the objectives of the analysis. Particular emphasis

is given to physic based, control oriented models, as they provide the primary tool for a comprehensive optimization of an energetic plant. Various numerical optimization techniques are then presented and described, providing insights on their role in both the design and regulation phases, and underlying the assumptions necessary to correctly describe the process. Three particular techniques are finally described in deeper details, as they will be subsequently exploited and applied to real applications.

In the second part of the thesis, the problem of optimal sizing and control of energetic systems is investigated based on three different applications:

- In the first one, a multi-source energy plant for the fulfillment of energetic demand in building is analyzed. This plant uses multiple sources, alternative to the classic fuel based ones, to provide thermal and electric power to the commercial structure. The effect that different scheduling policies (i.e. how to divide the energetic demand between the available sources) have on the primal energy consumption is estimated and the optimal plant control strategy is derived. While such a strategy cannot be implemented on-line to effectively drive the plant in its functioning, it offers many interesting insights on the process, proving the limitations of simpler rule-based strategies.

- In the second application, a hybrid solution for energy recovery in a hydraulic excavator is developed. As different plant layouts and component sizes are available for the design of similar technological solutions, a methodology to evaluate the benchmark potentiality of each different combination is derived, and the performance of the different solutions are compared on a fair common framework.

- In the third and final application, an Organic Rankine Cycle based Waste Heat Recovery system for light vehicles application is investigated. In this case, the problem is that of efficiently manage the heat recovery process, liable to the variability of the energy source, to maximize the fuel consumption reduction at the engine while keeping the plant in safe operating conditions.

For each of the applications, the particular model developed for representing the system will be described, an optimization problem will be conceived, and the suitable numerical optimization technique will be applied to maximize the efficiency of the plant.

Bibliography

[1] "Annual Energy Outlook 2015 with projections to 2040," U.S. Energy Information Administration (EIA), Tech. Rep., 2015. ix, 6, 7

[2] "BP Statistical Review of World Energy June 2015," British Petroleum, Tech. Rep., 2015. ix, 3, 4

[3] "Energy and Climate Change," International Energy Agency, Tech. Rep., 2015. ix, 5, 6, 8, 9

[4] W. Cramer, "Global Change Impacts on the Biosphere," in *Encyclopedia of Ecology*. Elsevier, 2008, pp. 1736–1741. 1

[5] IPCC, "Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change," Tech. Rep., 2013. 2

[6] IPCC 2007, "2007: Assessment of observed changes and responses in natural and managed systems. Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Chang," Tech. Rep., 2007. 1

[7] IPCC 2014, "Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change," Tech. Rep., 2014. 3, 5, 6, 7

[8] G. Marland, T. Boden, and R. J. Andres, "Global, Regional, and National CO2 Emissions," Oak Ridge National Laboratory, U.S. Department of Energy, Tech. Rep., 2003. 2

[9] C. P. Morice, J. J. Kennedy, N. A. Rayner, and P. D. Jones, "Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set," *Journal of Geophysical Research*, vol. 117, no. D8, p. D08101, Apr. 2012. 1, 2

CHAPTER 2
MODELLING TECHNIQUES FOR PROCESSES SIMULATION

While many of the questions connected to the design and optimization of a physical process may be theoretically addressed by means experimentation, such a procedure may result inappropriate or even impossible to carry out, either because too expensive (in terms of time and money), or because too dangerous, or simply because impossible, as the system to analyze does not exist yet [6]. It is to overcome these limitations that models are conceived, to derive a replica of the examined system and use it to perform simulations, an inexpensive and safe alternative to experiments, which can be used to predict how the real system would have behaved and act accordingly in the real world. Among the various modelling methodologies conceivable, the most important when dealing with technological problems is that of mathematical modelling, which consists in translating the real world problem into mathematical ones, solving such mathematical problems, and interpreting the results in the ordinary laymen's language [4]. While many different typologies of mathematical models may be developed, the common element they all share is their *predictive* ability, which allows to make quantitative estimations (whether deterministic or probabilistic), that can be exploited to anticipate the results form yet-to-be-conducted experiments or to predict a set of events in the real world [2, 3].

In the present chapter, a brief introduction to the field of mathematical modelling is depicted, focusing on the techniques and representations more suitable for the modelling and simulation of complex energetic processes.

2.1 Classification of mathematical models

To realize a mathematical simulation model, a certain number of simplifying hypothesis must be adopted, to define a schematic framework which will allow the reproduction of the physical phenomena of interest for the optimization of the process. Based on system theory, every process or component may be schematized following

an input/state/output representation, depicted in Figure 2.1, where the value of a certain number of observable (i.e. measurable) variables (*output*, y) is determined by a combination of external phenomena influencing the system (*inputs*, u) and the configuration (*state*, x) of the system itself.

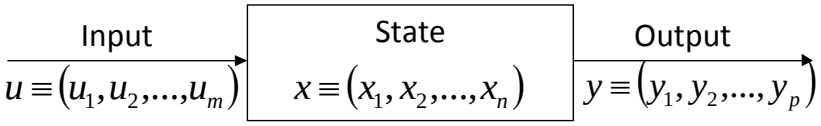


Figure 2.1: Generic input/state/output system representation

The mathematical correlations between input, states, and outputs may be formalized according to the generic state-space representation, Equation 2.1. This formulation states that the evolution of the state variables x (corresponding to the minimum set of parameters necessary to describe the configuration of the system) can be correlated to the exogenous inputs, u, and the actual configuration of the system, x, by means of a mathematical function f. Moreover, the observable effects of such actions may be correlated to system inputs and states by means of another mathematical correlation g.

$$\begin{cases} \dot{x}(t) = f\big(x(t), u(t), t\big) \\ y(t) = g\big(x(t), u(t), t\big) \end{cases} \tag{2.1}$$

Based on the nature of the mathematical correlations f and g, two important distinctions can be made when categorizing theoretical models [1]:

- *Black box models*: The correlation between the various variables are entirely based on empirical representations of the processes, without neither considering the underlying physical or chemical functioning principles, or developing a conceptualized representation of the process. Maps and lookup tables, experimental correlations, transfer functions, and neural networks are just a limited example of models that are based on this approach. The terminology refers to

the fact that the user does not have any insight on the real physical phenomena determining the behaviour of the system but can only empirically deduce a correlation by experimentally observing the output response of the system to some known inputs;

- *White box models*: In this case, on the other hand, the correlations between the variables are entirely based on known physical equations which describe the behaviour of the system. The model is therefore obtained entirely in a deductive analytical way, based on natural principles.

Based on the adopted approach, two procedure can be used to obtain a model [5]: if the black-box approach is followed, an *identification* procedure must be carried out, adjusting the internal parameters of the model to reduce the error between the model predictions and some known experimental data; if a white-box approach is instead followed, *modelling* requires the assembling of the considered fundamental laws to reproduce the behaviour of the analyzed system. While a model based on the white-box approach can be used to simulate the same physical system for many operating regimes, a black-box one is reliable only at the specific operating conditions at which it has been calibrated. Moreover, a white-box model can be used to derive the effect of some parameter modification on the process or even to simulate a system not yet realized in practice. Despite their obvious advantages however, white-box models are often too complex to conceive, as the physical phenomena determining even simple effects may be far from trivial, and the black-box approach is usually more practical.

Despite however this rigid distinction, in practice these two methods often complement each other, by defining the so-called *grey-box* approach, in which some simple physical correlations are used to describe the essential dynamics and behaviours of the system, while empirical correlations are used to describe the phenomena which does not require any particular insight, thus combining the advantages from both the modelling approaches.

Based on the simplifications adopted when conceiving a representation of the real process and the corresponding mathematical formulation, theoretical models may also be distinguished, according to [5], based on the classes and criteria defined in Table 2.1.

Table 2.1: Model classification

	Model Classification	
mutual dependence of variables	linear	nonlinear
parameters change with time	time invariant	time variant
dependence on spatial coordinates	lumped	distributed
randomness of variables	deterministic	stochastic
state variable change with time	static	dynamic

Mutual dependence of variables A process is described by a linear model if the equations describing the mutual dependence of variables (f and g) can be expressed in the form:

$$\mathcal{L}\phi = 0 \tag{2.2}$$

where ϕ is the vector of variables (states x, state derivatives $\dot{x}$, inputs u, and outputs y) and $\mathcal{L}$ is a linear operator satisfying:

$$\mathcal{L}(\phi_{\mathbf{1}} + \phi_{\mathbf{2}}) = \mathcal{L}\phi_{\mathbf{1}} + \mathcal{L}\phi_{\mathbf{2}} \tag{2.3}$$

If this property is not satisfied, the model is nonlinear.

While the equation describing most of the physical systems are essentially nonlinear, it is often advisable to obtain a simplified linearized approximation of the process if the effect of such nonlinearities is negligible in the examined range of operating condition, in order to facilitate the analysis of the system and the implementation of numerical optimization techniques. The linearization procedure on the other hand, while justified from a computational point of view, must always be applied cautiously, keeping in mind its intrinsic limitations and the correlated loss of accuracy in the model predictive ability.

Parameter change with time For time invariant processes, the form of the state update and output Equations 2.1 does not depend on the time parameter. While all the basic physical law are time invariant, in a simplified representation of

the system, some phenomena (e.g. wearing, ageing, breakdown, etc.) can be much more easily described assuming that some parameters in the system equations are subject to some sort of variation over time. Usually, such effects are neglected if the changes are slower than the characteristic duration of the process.

Dependence on spatial coordinates If time t is the only independent variable describing the system state changes over time, the model is formulated by means of ordinary differential equations (ODEs). If on the other hand the set of independent variables contains some of the spatial coordinates or any other kind of generalized coordinates, the dynamics of the system are described by means of partial derivative equations (PDEs). Except for the very few cases where analytical solutions are available, the standard procedure for dealing with such models requires a discretization over the additional coordinates, so that the PDEs can be effectively reduced to a system of ODEs. Moreover, even if many of the variables are not evenly spatially distributed, it is customary to considered the problem *lumped* in space, if the effect of space distribution is not relevant to the problem and is negligible enough, in order to avoid any unnecessary complication.

Randomness of variables In deterministic processes, the relationship between input, states, and output variables, even if not unique [5], is always completely determined. In stochastic processes on the contrary, the influence of random variables greatly affects the behavior of the system. Again, even if all the physical process are inherently stochastic, the decision to include such information depends on the influence that such aspect has on the predictive ability of the model.

State variable change with time A last important feature to characterize models and processes relies on the distinction between dynamic and static ones. A dynamic model accounts for the time-dependent changes in the state of the system, while static ones are time invariant. From a mathematical point of view, this means that, for a static model, the time derivatives of all process variables are always equal to zero ($f\big(x(t), u(t), t\big) \equiv 0$) and that the process can be entirely described by algebraic equations. This of course does not mean that no change is detected in the states or the outputs over time, but it means that such a change is a direct immediate consequence of a variation in the value of the inputs. While all the physical processes are inherently dynamic, it is always advisable, for the sake of simplicity, to neglect those dynamics which are either too fast or too low when compared to the ones

relevant for the description of the dynamics, assuming that such processes always happen as a transition between different steady-state conditions.

2.2 Physical system modelling process

While many different theoretical approaches may be followed when conceiving a black-box system model, the development of a white-box physic based model follows a more standardized procedure, which is essentially based on the principle of breaking down the properties of the entire system into a set of sub-systems whose behaviours are known [6]. The different phases necessary for the process of identification and mathematical representation of a physical system may be reassumed as follows [5]:

1. **Problem definition**: the objective and purposes of the model are defined according to the different constraints (accuracy, simplicity, computational cost, etc.), thus defining the level of detail required to describe the process and the relevant phenomena to be accounted;

2. **Process analysis**: the boundaries separating the process from its environment are determined, isolating the inputs and the outputs to the system, obtaining the outer shell of the system model as depicted in Figure 2.1. Next, the entire process is divided into its simper and elementary sub-processes and parts;

3. **Assembly of conservation equations**: when developing physical mathematical models, it is usually possible to formulate some general properties of the system by means of *balance* or *conservation principles* [2]. An example of such conservation laws are for example mass, energy, momentum, charge, etc. conservation equations. A general formulation of these balance laws may stated as follows:

$$q_{in}(t) + q_{out}(t) \pm q_{produced,v}(t) = \frac{dQ_v(t)}{dt} \tag{2.4}$$

where v represent the specific control volume over which the balance is evaluated, $Q(t)$ is the physical quantity monitored, $q_{in}(t)$ and $q_{out}(t)$ are the flow rates of $Q(t)$ entering or leaving the sub-system boundary, and $q_{produced}$ is the rate of production (or destruction) of $Q(t)$ inside the control volume. A simplifying hypothesis underlying such formulation is that the quantity Q is evenly distributed inside the system boundary, allowing for a lumped parameter approximation.

If the spatial distribution of Q cannot be neglected, the conservation law can be stated in the more general form:

$$\frac{\partial \Theta}{\partial t} + \nabla \cdot \mathbf{q} = 0 \tag{2.5}$$

where $\Theta(\mathbf{x}, t)$ is the density of Q and $\mathbf{q}(\mathbf{x}, t)$ is its flux.

From a system theory point of view, the physical quantities Q may easily be interpreted as the states of the system, while Equations 2.4 and 2.5 serve as the basis for deriving state update function f;

4. **Assembly of physical/phenomenological equations of state**: to interconnect the different subsystems representations (carried out by means of the conservation laws) and to derive a correlation between state update equations and input/outputs of the system, additional physical equations are usually necessary, to relate the process variables by means of empirical correlations describing different properties and phenomena (e.g. ideal gas law, Bernoulli equation, heat transfer process, diffusion, etc.);

5. **Solution of the mathematical model**: analytical solutions (very rarely) or rather numerical approximations are evaluated by means of a calculator; if the model is not lumped in space, the PDEs describing the conservation Equation 2.5 are usually converted to a system of ODEs, by discretizing the entire control volume in a series of smaller volumes, dv, and applying the lumped conservation Equation 2.4 assuming that the properties are evenly distributed in dv.

6. **Testing**: the results from the simulation are compared to those intuitively expected from the system, to check inconsistencies or unexpected model behaviors and fix them;

7. **Calibration**: the eventual unknown parameters of the model are estimated and tuned based on experimental data collected from the real physical process. If the whole process is too complex to examine at once, the singular subsystems are individually calibrated;

8. **Validation**: experimental data, different from those used for the calibration, are used to establish the satisfactory degree of model actual process predictions;

9. **Applying the model**: the model is finally used to perform the desired task it was built for.

Bibliography

[1] E. O. Doebelin, *System Dynamics: Modeling, Analysis, Simulation, Design.* New York, New York, USA: Marcel Dekker Inc., 2009, no. 4. 14

[2] C. L. Dym, *Principles of mathematical modeling*, 2nd ed. Elsevier Academic Press, 2004. 13, 18

[3] S. Howison, *Practical applied mathematics: modelling, analysis, approximation.* Cambridge, UK: Cambridge University Press, 2005. 13

[4] J. N. Kapur, *Mathematical Modelling.* New Dheli, India: New Age International Publishers, 1998. 13

[5] V. Kecman, *State-Space Models of Lumped and Distributed Systems*, ser. Lecture Notes in Control and Information Sciences, V. Kecman, Ed. Berlin: Springer-Verlag, 1988, vol. 112. 15, 17, 18

[6] L. Ljung and T. Glad, *Modeling of dynamic systems.* Englewood Cliffs, New Jersey, US: Prentice Hall, 1994. 13, 18

CHAPTER 3
NUMERICAL OPTIMIZATION: THEORY

After a mathematical representation of a system is developed, it is possible to translate the abstract optimization problem (e.g. maximize the profit of a process, minimize the energy consumptions of an engine, maximize the productivity of a plant, etc.) into a mathematical one if a proper *objective function*, representing a quantitative measure of the performance of the system, is identified. Such an objective function depends on a number of different characteristics of the system, which are referred to as *decision parameters*. The objective of the optimization procedure is to find the optimal combination of such parameters that leads to a maximization (or a minimization) of the objective function. Furthermore, when dealing with physical systems and real life problems, the choice of the decision parameters is usually subjected to a set of limitations, referred to as *constraints*, which represent the boundaries that are normally present when dealing with real process (e.g. control actuators saturation, limitations to ensure safety conditions, the narrowness of the available resources, etc.). The identifications of these parameters is the first step in the optimization problem and usually represents one of the most important parts of the entire procedure.

Once the proper optimization problem has been conceived, the correct algorithm must be chosen to find an effective solution solution. As many different sub-types of optimization problems may be formulated, there is not a singular procedure to obtain a solution, but rather a collection of methods which depends on the particular formulation given to the investigated problem. Many important distinctions may be made when characterizing different formulations of the optimization problem, but the first and most important one relies on the distinction between **static** and **dynamic** problems. In the former, the decision parameters are optimized for a specific instant of time. This is the case for example of deciding the proper sizing of a component or assingining the nominal operating point of a plant, as the decision parameter, once chosen, will not change as time progresses. A static optimization problem may also be conceived when the dynamics of the system are considerably faster or slower than

those of the parameter optimization, and the process can easily be assumed as quasi-static. In the latter case of dynamic optimization on the other hand, the decision parameters are actively changed as time progresses and their value is optimized for a given time interval. Rather than a fixed set of parameters, the optimization problem requires therefore the definition of a dynamically changing control law, able to effectively drive the progression of the process in such a way that a certain optimality criterion is met. For this reason, the problem is normally referred to as "optimal control", and forms a branch in the wider field of control theory.

In the present chapter, a reference framework for both static and dynamic optimization problems is provided, listing the possible variants of the two problems and a set of techniques that have been developed to deal with these different formulations.

3.1 Static optimization

A mathematical formulation for the generic static optimization problem may be developed as follows [26]:

$$\min_{x \in X} J(x) \quad \text{subject to} \begin{cases} h_i(x) = 0, & i = 1, 2, ..., m \\ g_j(x) \leq 0, & j = 1, 2, ..., n \end{cases} \tag{3.1}$$

where the problem is that of finding the right combination of values in the vector of *decision parameters* $x \in X \subset \mathbb{R}^n$ which minimize the *objective function* $J(x)$, given that the sets of equality $(h_i(x))$ and inequality $(g_j(x))$ *constraints* are satisfied.

The general formulation presented above immediately highlights one of the first possible way to classify static optimization problem, namely **unconstrained** and **constrained** problems. In the former case, where no conditions of the form $h_i(x)$, $g_j(x)$ are imposed, the complexity of the optimization is greatly reduced, though the solution of the simplified problem is still far from trivial. A wide set of techniques has been developed to address the basic unconstrained problem and many of these serve as the basis to deal with the constrained case. Moreover, it is also worth being noted that, if a proper penalty term is introduced into the objective function, such as that the violation of the constraints will penalize the performance, the constrained problem may be easily reduced to an unconstrained one. While this technique is usually adopted to tackle the presence of constraints in a very simple way, the effectiveness of the solution greatly depends on a proper formulation of the penalty function, a procedure which is highly problem dependant. If other methodologies, based on an

analytical approach to the problem, are available and easy to implement, they are usually preferred due to their increased flexibility and effectiveness.

Another primary distinction in static optimization problems may be operated based on the **global** or **local** nature of the solution. If the objective function $J(x)$ is non-linear, multiple different local minima may be present; as many optimization algorithms seek only a local solution, i.e. a point where the objective function is smaller than in other feasible nearby points, there is no guarantee that the global solution may be located or even recognized [26]. Local optimization methods on the other hand are usually faster and easy to apply and for this reason are widely employed in optimization algorithms, with the introduction of some proper modifications able to enhance the possibility of finding a global optimum [11]. The concept of **convexity** plays in here a crucial role and is therefore briefly summarized; a set $S \in \mathbb{R}^n$ is defined convex if, for any two points $x_1, x_2 \in S$, $\alpha x_1 + (1-\alpha)x_2 \in S$ for all $\alpha \in [0,\ 1]$, that is, the straight segment connecting the two points entirely lies in S. For a function f defined over a convex domain S, the function is convex if, for any two points x_1, x_2 in S:

$$f\big(\alpha x_1 + (1-\alpha)x_2\big) \leq \alpha f(x_1) + (1-\alpha)f(x_2), \quad \forall \alpha \in [0,\ 1] \tag{3.2}$$

In the case that both the objective function and the feasible set are convex, any local solution of the optimization problem is also global [26]. Many practical optimization problems may be formulated from the start as convex optimization problems; if that's the case, it is relatively straightforward to solve them. If on the other hand the problem is non-convex, the use of convex methods still plays an important role in the solution phase as these methods may be used to find an approximate solution to the problem or to determine the lower bound to the optimal value of the non-convex problem [11].

The numerical solution of Problem 3.1 is based on the application of iterative algorithms. While in fact such problems may be theoretically solved both by the use of brute force methods or calculus of variations, the former would obviously involve an extremely heavy computational burden while the latter is only effective in a very restricted number of cases, as high non-linearities in the objective function and the necessity of an effective constraint handling easily make the analytical approach unfeasible. Numerical algorithms on the other hand rely on an iterative procedure where, starting from a (set of) initial candidate solution(s) to the problem, a sequence of improved estimation is generated, hopefully leading towards the solution of the problem.

Numerical algorithms may be classified into two large families of methods: **iterative methods** and **metaheuristic algorithms**. The former are based on a mathematical approach to the problem, usually relying on approximating the objective function and its first and second order derivatives (gradient and Hessian) to estimate the direction where to proceed. Metaheuristic algorithms on the other hand still relies on iterative procedures but do not make any assumption on the analytic structure of the objective function and constraints, neither try to find a direction of descent; they make fewer assumptions on the optimization problem to be solved and rely instead on sets of usually simple high level rules which will (supposedly) guide the evolution of the search towards the optimal solution.

3.1.1 Iterative algorithms

Iterative algorithms for unconstrained optimization follow two fundamental strategies and may be therefore distinguished in: **line search** or **trust region** methods [26]. The progression towards the optimal estimated point is conducted as follows:

- Line search: At each iteration k of the algorithm, a direction p_k is chosen and the search progresses from current point x_k towards the direction p_k based on the following update equation:

  $$x_{k+1} = x_k + \alpha_k p_k$$

  where α_k is the step length, which should ideally be equal to the minimizer of:

  $$\min_{\alpha>0} J(x_k + \alpha p_k)$$

  While the exact solution of the equation above will guarantee the maximum benefit from proceeding in direction p_k, the computation cost may be expensive and the procedure usually unnecessary. Line search methods only tries a limited number of trial step lengths until an approximate minimum is find, and then proceed to the next point. The search direction p_k is evaluated based on the gradient (steepest descent) or on a second order Taylor series approximations of the function (Newton or Quasi-Newton).

- Trust region: In this case, a simplified "model function" m_k, approximating the

behaviour of J near the current point x_k, is constructed. A candidate step p_k is then found by approximately solving the subproblem:

$$\min_p m_k(x_k + p)$$

The model function m_k is usually a quadratic function of the form:

$$m_k(x_k + p) = J_k + p^T \nabla J_k + \frac{1}{2} p^T B_k p$$

where ∇J_k and B_k are respectively the function gradient and its Hessian $\nabla^2 J_k$ or some approximations. As the model m_k may not be a good approximation of the function J, the search is restricted to a small region around x_k, hence the name trust region. This region is usually a ball defined as $\|p\|_2 \leq \Delta$, with Δ the trust region radius. Contrary to line the line search method, in the trust region procedure the evaluation of step length and step direction are made simultaneously.

Both methods relies on the knowledge of the derivatives of the objective function. While in a few cases the derivative could be analytically calculated by the user and provided to the algorithm, in the great majority of the problems numerical approximations are used instead, with the finite difference approach being the most used. If the number of derivative evaluations to be performed is excessive, Derivative-free Optimization (DFO) algorithms may be used instead (e.g. the conjugate-direction or Nelder-Mead simplex-reflection methods [26]).

If equality constraints ($h_i(x) = 0$) are added to the problem, the method of Lagrange multipliers is used to effectively reduced the constrained problem to an equivalent unconstrained one [6]; if both equality and inequality constraints ($g_j(x) \leq 0$) are present, Karush-Kuhn-Tucker (KKT) conditions serves as the basis for the development of many advanced techniques. KKT conditions (Equations 3.4) are deduced from an extension of the Lagrange multipliers method and state that, if a Lagrangian function is defined of the form:

$$L(x, \lambda, \mu) = J(x) - \sum_{i=1}^{m} \lambda_i h_i(x) - \sum_{j=1}^{n} \mu_j g_j(x) \tag{3.3}$$

where λ and μ are the Lagrange multiplier vectors, then, if x^* is a local solution of Problem 3.1, $J(x)$, $g(x)$, $h(x)$ are continuously differentiable, and linear independence constraint qualification (LICQ, see [26]) is satisfied, then there exist unique Lagrange multipliers λ^*, μ^* such as that the following conditions are satisfied:

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0, \tag{3.4a}$$

$$h_i = 0, \quad \forall i \in [1, 2, ..., m] \tag{3.4b}$$

$$g_j \leq 0, \quad \forall j \in [1, 2, ..., n] \tag{3.4c}$$

$$\mu_j^* \leq 0, \quad \forall j \in [1, 2, ..., n] \tag{3.4d}$$

$$\mu_j^* g_j(x^*) = 0, \quad \forall j \in [1, 2, ..., n] \tag{3.4e}$$

A great portion of the algorithms developed for constrained optimization may be effectively interpreted as methods for numerically solving the KKT system of equations [11].

Another key aspect in constrained optimization is represented by the concept of duality. If a so-called "Lagrange dual function" is associated to the "primal" optimization Problem 3.1 as:

$$d(\lambda, \mu) = \inf_{x \in X} L(x, \lambda, \mu) = \inf_{x \in X} \left(J(x) - \sum_{i=1}^{m} \lambda_i h_i(x) - \sum_{j=1}^{n} \mu_j g_j(x) \right) \tag{3.5}$$

then, the Lagrange dual function $d(\lambda, \mu)$ gives a lower bound to the optimal solution of the primal optimization problem. Moreover, the connected Lagrange dual problem (i.e. finding the maximum value of the Lagrangian dual function, which represents the maximum lower bound to the primal problem) is a convex optimization problem, an can be therefore solved with much simpler methods, even if the primal is non-convex [11]. Finally, if also the primal problem is convex, the solution of the Lagrange dual problem will correspond to the optimal value for the primal problem (strong duality condition [11]). Many iterative algorithms for convex optimization exploit such a property, forming the family of *primal/dual methods*.

The existing iterative algorithms for constrained optimization may be categorized based on the form of the objective function and the constraints; the most basic problem is that of *Linear Programming* (LP), where both the objective function and the constraints are linear, and the resulting optimization problem is convex. In this case, the simplex method (belonging to the wider class of "active set" methods) represent one of the oldest numerical methods ever developed. As the solution of the LP problem, if existing, is forced to lie on the boundary of the feasible set, the simplex algorithm proceeds in formulating reduced versions of the original problem, where only a subset of the original constraints is considered, and always as equality ones. Once this subproblem is solved, a different subset of constraints is evaluated

in the next iteration of the algorithm, thus moving towards the effective solution of the problem [26]. Another class of methods worth to be cited for the solution of LP problems is that of "interior-point" methods, where the inequality constraints are approximated by means of a properly designed barrier function able to "push" the tentative solution inside the feasible set. The equivalent problem is solved by means of the Newton method and the barrier penalization is iteratively relaxed until a boundary of the problem is met [11].

Another particular category of constrained optimization problems is represented by the *Quadratic Programming* (QP) ones. In this case, the objective function is quadratic and the constraints are linear; the resulting optimization problem is still convex. This kind of problem has also been studied for long, due to some of its particular characteristics that can be exploited for developing efficient algorithms for dealing with more complex situations. Several methods have been developed and implemented in commercial softwares for the solution of the QP problem, based on active-set, interior-point and gradient projection methods. Strong duality condition is also often use to increase the performance of the algorithms, developing the so-called "primal-dual" methods.

Finally, if both the constraints and the objective function are nonlinear, then the optimization problem is classified as a *Non-linear Programming* (NLP) problem and solved using a variety of different methods, based on the particular nature of the objective function and constraints (see [26] for more details). The use of *penalty functions* and *augmented Lagrangian* methods represents in this case an efficient way to replace the constrained original problem with a sequence of simplified subproblems. *Sequential Quadratic Programming* (SQP) represents one of the most effective methods for the solution of non-linear constrained optimization problems, based on the iterative solution of a simplified QP problem resulting from the quadratic approximations of the objective function and a linear approximations of the constraints. Interior-point methods have also proved to be as successful for non-linear optimization as for LP, and, together with SQP methods, they are currently considered among the most powerful algorithms for solving NLP [26].

While the iterative algorithms described in this section present many advantageous qualities such as high accuracy, robustness, proof of convergence, and rapid convergence rate, they may become ineffective if large non-linearities (e.g. discontinuities) are present in the objective functions and constraints. Furthermore, as the search space dimension increases, iterative algorithms require increasingly higher

computational efforts. Finally, while they are very effective in providing local solutions, they may get easily stuck in local optimizers if the problem is non-convex. Metaheuristic algorithms provide in this case a valuable alternative to overcome the difficulties arising from the application of classic iterative procedures.

3.1.2 Metaheuristic algorithms

Metaheuristic algorithms are based on a completely different approach than classical methods. The term is coined by the composition of two Greek words meaning "to find" (*heuristic*, form the verb $\epsilon\upsilon\rho\iota\sigma\kappa\epsilon\iota\nu$) and "beyond, in an upper level" (*meta*), suggesting that the progression of the algorithm relies on a set of high level procedures and rules that guide the evolution of the system. Rather than being determined by analytical considerations, these rules defines the specific actions to be taken by the algorithm (according to the response of the system with respect to changes in input variables) based on more abstract concepts, inspired for example by the observation of natural mechanisms, and does not require therefore any kind information about the analytic structure of the objective function (gradient, Hessian). The goal of the metaheuristic algorithms is to efficiently explore the search space in order to find (near) optimal solutions; moreover, they usually incorporates mechanism to avoid being trapped in confined regions. Being based on an abstract description of the problem, they may be applied indiscriminately to a wide set of different cases. A classification of metaheuristic algorithms may be found in [29] and is here reported:

- **Nature inspired and nonnature inspired:** A large portion of the metaheuristic algorithms are inspired by natural phenomena, such as biology (evolutionary algorithms, artificial immune systems), social sciences (ant colony, bees colony, particle swarm) and physics (simulated annealing);
- **Memory usage and memoryless methods:** Some methods, as simulated annealing, do not use any information that may be extracted dynamically during the search, while others rely on short-term and long-term memories (e.g. tabu search);
- **Deterministic and stochastic:** In deterministic methods the update is made using deterministic processes, while in stochastic methods some random rules are applied during the search;
- **Population-based and single-based search:** Single-solution algorithms, as local search and simulated annealing, rely on the manipulation and evolution of

a single candidate solution as the algorithm progresses while population-based exploit a large group of candidates at each iteration. Single-solution methods intensify the search in local regions, while population based provide a better diversification in the entire search space;

- **Iterative and greedy**: In iterative algorithms, representing the majority of metaheuristics, a complete solution (or population) is present from the start and its evolution is controlled by some search operators; greedy algorithms on the other hand starts from an empty solution and add a decision variable at each step, until the complete solution is obtained.

Some of the most used metaheuristic algorithms are here presented, referring to the distinction between single-based and population-based approach:

- **Single-based alhorithms**

  1. *Local search:* One of the oldest and most simple metaheuristic methods [1], is based on replacing the current solution with a neighbour one presenting an improved objective function. Neighbour solutions may be evaluated deterministically or stochastically and the selection of the best neighbour is chosen based on different approaches (best improvements, first improvement found, random selection among improving neighbours, etc.). While the method proved to be very easy to apply and able to find fairly good solutions in a small amount of time, it presents the high disadvantage that it converges toward local optima. Many alternatives algorithms have originated from local search, implementing modifications that avoid getting stuck at local optima.
  2. *Simulated annealing:* Originating from the work of Kirkpatrick et al. [22] and Černý [31], the methods deals with the problem of local minima by accepting movements that may degrade the current solution of the problem. The algorithm receive its name as it is inspired by the physical process of formation of strong crystalline structures in metals. From an initial solution, moves that improve the objective function are always accepted, but there is also a probability of choosing a trajectory that degrades the cost. This probability is inversely proportional to the amount of performance degradation and proportional to a control parameter, named temperature. At high "temperatures" the probability of accepting worse moves is high and the movement is almost random, while for low "temperatures" the

algorithm is equivalent to local search. Cooling schedules, which define how temperature changes as the algorithm progresses, are a key aspect for the correct functioning of the algorithm.

3. *Tabu search:* Proposed by Glover in 1986 [19], it is similar to the steepest descent version of the local search, but it does accept non-improving solutions if all the neighbours of a point are non-improving so that, even if a local minima is met, a new current solution is chosen anyway. As this policy may generate cycles, tabu search keeps in memory the most recent trajectories and discards the points that have already been visited (hence the name). Variations of the algorithm are characterized by the way the memory of past position is kept (short, medium and long term memory) and the criterion to eventually accept tabu moves.

4. *Iterated local search:* As local search effectiveness depends on the initial point, the accuracy of the algorithm may be increased by selecting multiple starting points or by perturbing the solution of the local search algorithm and rerun the same algorithm from the new perturbed initial point.

5. *Guided local search:* The basic principle of this method is the introduction of a dynamic changing function which is penalized if the algorithm gets trapped to a local optima [32].

- **Population-based algorithms**

1. *Evolutionary algorithms:* Among the most famous metaheuristic algorithms, evolutionary algorithms are inspired by the principle of natural evolution. Many variants have been developed based on the same concept as genetic algorithms [20], evolution strategies, and genetic programming [23]. These methods are based on the evolution of a population of candidate solutions to the optimization problem; at each step of the algorithm, individuals are selected from the population to generate new offsprings which will inherit parents' characteristics based on crossover and mutation operators. The probability of each individual in the population to be selected for the reproduction is based on its fitness, i.e. the value of the objective function $J(x)$. Reproduction and fitness evaluation are repeated in an iterative manner until stopping criteria are met. Evolutionary algorithms have been observed to perform very effectively in many different applications and a great amount of literature production have been written on the argument (see for example [9, 14, 29, 33]).

2. *Swarm intelligence:* These methods are inspired by the cooperation and collective behavior of social animal species, such as ants, bees, fish, and birds. In swarm intelligence methods, each candidate solution moves through the search space based on simple strategies that implement an indirect communication between different individuals, thus generating a collective behavior similar to that observed in nature. Ant colony algorithm for example [16] mimics the pheromone releasing procedure and evaporation process that guides ant colony in the solution of shortest past decision problems. In particle swarm optimization [21] the movements of singular individual in the population are affected not only by each particle knowledge (personal fitness) but also by the knowledge shared from other individuals (global fitness), a process which leads to the emersion of coordinate behaviors. Bee colony algorithm [30] is inspired by the features that characterize the interactions inside honeybee colonies such as nectar exploration, mating, division of labor, etc.

Even if metaheuristic algorithms does not offer any guarantee to find a global or even bounded solution, they outperform the other methods when the search space and the dimension of the problem are large, and when the objective function is highly non-linear or presents large discontinuities. Even if the global optimality is not guaranteed, metaheuristic algorithms either account for special strategies for escaping local minima or, if some stochastic update mechanism is present, this usually has the intrinsic ability of encouraging the exploration of wider search spaces, thus preventing the solution to be stuck at local optimizers [29].

3.2 Dynamic optimization

If the plant is subject to the action of a set of time varying parameters u which affects both the dynamic evolution of the system state and the value of some performance parameters that has to be optimized, the problem belongs to the category of dynamic optimization. This is usually the case of determining the optimal trajectory of some control parameter used to actively regulate the plant. For such a problem, the development of an efficient control strategy relies on a mathematical representation of the system based on the state-space formulation depicted in Chapter 2.1 and here reported for the sake of clarity:

$$\dot{x}(t) = f(x(t), u(t), t), \quad x(t_0) = x_0 \tag{3.6}$$

where $x \in X \subset \mathbb{R}^n$ are the states of the system, i.e. the combination of parameters necessary to identify a generic configuration of the system itself, $u \in U \subset \mathbb{R}^m$ are the control inputs whose trajectories are to be optimized, and $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is the correlation determining the evolution of the system based on its current state and the applied controls. It is worth to highlight that, even if detailed plant models may be developed, for the sake of deriving an effective solution to the dynamic optimization problem, it is better to just consider the dynamics of the system (i.e. the states) who have a direct effect on the performance index to be optimized, and time scales consistent with those of the control inputs to be optimized. If in fact the dynamics of the plant are too fast or too slow (if compared to those of the controller), a quasi-static approximation of the system allows to obtain results very close to those attained from the implementation of an optimal dynamic strategy, a the price of the solution of a much more simple static optimization problem.

If on the other hand the dynamic aspects of the system are central to the optimization procedure, a mathematical representation of the plant performance is expressed by means of the following cost functional $J(x_0, u(t))$:

$$J(x_0, u(t)) = g\big(x(t_f)\big) + \int_{t_0}^{t_f} h\big(x(\tau), u(\tau), \tau\big)\ d\tau \tag{3.7}$$

where the objective function J is optimized over a time interval $[t_0; t_f]$ based the values of a terminal cost g, associated to the final obtained configuration of the system x_{t_f}, and the value of a stage cost h, associated to the control input trajectory $u(t)$ and the induced evolution of the system $x(t)$. Notice how, together with the control inputs trajectory, the initial configuration of the plant x_0 plays a central role in determining the value of the objective function.

In real applications, it is also necessary to account for the various physical limitations of the plant. Because of that, the optimization problem is usually subject to a series of constraints on both inputs and states of the form:

- $u(t) \in U \subset \mathbb{R}^m$: input constraints (frequently expressed in the form $u_{min} \leq u \leq u_{max}$), representing physical limitations on the control values, due for example to actuators saturation;
- $\dot{u}(t) \in \dot{U} \subset \mathbb{R}^m$: input rate of change constraints, again usually associated to control physical limitations;

- $x \in X(t) \subset \mathbb{R}^n$: state constraints, restricting the number of possible configurations allowed to the system during its evolution, e.g. safety limitations for some physical parameters;
- $x(t_f) \in X_f \subset \mathbb{R}^n$: terminal constraints, used if the system should be in a desired predetermined configuration at the end of the optimization horizon.

To determine the optimal control trajectory u^*, which maximizes the performance of the system, while accounting for the various physical limitations and the dynamic evolution of the plant, the following optimization problem, usually referred to as optimal control, can be defined:

$$u^*(t) = \arg\min_{u(t)} J(x_0, u(t)) \tag{3.8a}$$

$$\text{subject to} \quad \dot{x}(t) = f\big(x(t), u(t), t\big), \quad x(t_0) = x_0 \tag{3.8b}$$

$$x(t) \in X \quad \forall t \in [t_0; t_f] \tag{3.8c}$$

$$u(t) \in U \quad \forall t \in [t_0; t_f] \tag{3.8d}$$

$$\dot{u}(t) \in \dot{U} \quad \forall t \in [t_0; t_f] \tag{3.8e}$$

$$x(t_f) \in X_f \tag{3.8f}$$

3.2.1 Optimality conditions

Analytical approaches to the solution of the unconstrained version of optimal control problem 3.8a are based on calculus of variations. Starting from the work of Bellman on dynamic programming and its principle of optimality [7], a sufficient condition for optimality may be obtained deriving Hamilton-Jacobi-Bellman equation (see [2] for more details). From Equation 3.7 it is possible to obtain the function:

$$J^*(x(t), t) = \min_{u_{[t,t_f]}} J(x(t), u(t), t) = \min_{u_{[t,t_f]}} \left[g\big(x(t_f)\big) + \int_t^{t_f} h\big(x(\tau), u(\tau), \tau\big)\, d\tau \right] \tag{3.9}$$

which defines the optimal value of the objective function J, depending on the generic initial condition $x(t)$ at time t. For a $t_1 \in [t, t_f]$, the control trajectory $u_{[t,t_f]}$ is the concatenation of $u_{[t,t_1]}$ and $u_{[t_1,t_f]}$, and the minimization over $u_{[t,t_f]}$ is equivalent to minimizing over $u_{[t,t_1]}$ and $u_{[t_1,t_f]}$:

$$J^*(x(t),t) = \min_{u_{[t,t_1]}} \min_{u_{[t_1,t_f]}} \left[g\big(x(t_f)\big) + \int_t^{t_1} h\big(x(\tau),u(\tau),\tau\big)\ d\tau + \int_{t_1}^{t_f} h\big(x(\tau),u(\tau),\tau\big)\ d\tau \right] \tag{3.10}$$

With the first integral independent from $u_{[t_1,t_f]}$ and the second being an optimal performance index itself, it is possible rewrite $J^*(x(t),t)$ as:

$$J^*(x(t),t) = \min_{u_{[t,t_1]}} \left[\int_t^{t_1} h\big(x(\tau),u(\tau),\tau\big)\ d\tau + J^*(x(t_1),t_1) \right] \tag{3.11}$$

Equation 3.11 is a mathematical expression of Bellman's principle of optimality [17], and it formalizes the somehow self-evident knowledge that the "tail" of an optimal trajectory is still optimal. A classic example [7] to better illustrate the process is that of an auto travelling from a point A to a point B. If the fastest route between the two points passes through a third point C, the principle of optimality states the basic knowledge that the portion of the optimal route from A to B starting from C is also the fastest route from C to B. This seemingly trivial concept will play a central role for the derivation of the Dynamic Programming algorithm, described in Chapter 4.2.

Furthermore, by applying Taylor's theorem to expand the right side of 3.11 and rearranging the equation, it is possible to obtain the following equation:

$$\frac{\partial J^*}{\partial t} = -\min_{u(t)} \left[h\big(x(t),u(t),t\big) + \frac{\partial {J^*}^T}{\partial x}\ f(x(t),u(t),t) \right] \tag{3.12}$$

The partial differential equation above is known as Hamilton-Jacobi-Bellman equation and plays a central role in optimal control theory by providing sufficient conditions for the optimal performance index and determining an optimal control trajectory [2].

Another approach based on calculus of variations for the unconstrained optimal control problem 3.8a derives from the work of Pontryagin, who basically extended the method of Lagrange multipliers to the case of optimal control problems (see [18] for deeper details). If a costate vector λ is introduced, it is possible to define a Hamiltonian function $H : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \times [t_0, t_f] \to \mathbb{R}$ of the form:

$$H\big(x(t), u(t), \lambda(t), t\big) = h(x(t), u(t), t) + \lambda^T(t) f(x(t), u(t), t) \tag{3.13}$$

where the state equation is incorporated to the stage cost as in static optimization an equality constraint is incorporated into the objective function by means of a Lagrange multiplier (see Equation 3.3). Pontryagin's minimum principle requires that, for the control u^* to be optimal, the following conditions should be satisfied:

$$\dot{x}^*(t) = \nabla_\lambda H = f(x^*(t), u^*(t), t) \tag{3.14a}$$

$$x^*(t_0) = x_0 \tag{3.14b}$$

$$\dot{\lambda}^*(t) = -\nabla_x H = -\nabla_x h(x^*(t), u^*(t), t) - \left[\frac{\partial f}{\partial x}(x^*(t), u^*(t), t)\right]^T \lambda^*(t) \tag{3.14c}$$

$$\lambda^*(t_f) = \nabla_x g(x^*(t_f), t_f) \tag{3.14d}$$

$$u^*(t) = \arg\min_u H\big(x^*(t), u(t), \lambda^*(t), t\big) \tag{3.14e}$$

The equations above provide a set of necessary conditions for an optimum of the objective function and may be used to solve several optimal control problems [18]; moreover, they also represent an useful benchmark for checking the extremality of solutions found by numerical methods [4].

3.2.2 Algorithms

Many methods have been developed to solve the optimal control problem 3.8a. For a simplified problem formulation, where the system is linear (or can be linearized near an equilibrium point [3]) and the cost is a quadratic form of inputs and states:

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{3.15}$$

$$J(u(t)) = \frac{1}{2} x(t_f)^T Q_f x(t_f) + \frac{1}{2} \int_{t_0}^{t_f} \big(x(t)^T Q x(t) + u(t)^T R u(t)\big)\, dt \tag{3.16}$$

many solution based on the theory of linear algebra have been obtained. For an unconstrained, infinite time horizon problem, Hamilton-Jacobi-Bellman equation may be used to derive the so-called Linear Quadratic Regulator (LQR) controller (see for example [2, 25]). In this simple case, a direct solution to the problem is available in the form of a closed feedback relationship between the input u and the state x of

the system. The optimal trajectory can be therefore expressed as $u^*(t) = -Kx(t)$, with K a matrix resulting from the solution of a Riccati equation containing both the system matrixes A and B, and the weights Q_f, Q, R from the objective function. Furthermore, if linear constraints on states and inputs are presents, linear Model Predictive Control (MPC) techniques are able to reduce the solution of 3.8a, for a finite time horizon, to the solution of a Quadratic Programming problem [10, 27]. More detail about MPC will be given is Section 4.3 of the following chapter.

If on the other hand the system or the objective functions are non-linear, numerical solutions of Problem 3.8a may be attained based on Bellman principle of optimality and dynamic programming (DP) [7]. In this case the optimal policy is determined by back induction, proceeding backward in time from the final state and deriving the optimal trajectory in a piecewise method by successive applications of optimality principle. More details about the DP approach and the presentation of a discrete DP algorithm for deterministic problems are postponed to the next chapter, in Section 4.2.

Pontryagin's minimum principle and conditions for optimality served as the guideline for the first numerical methods developed for the solution of the dynamic optimization problem, the so-called *indirect methods*; in this case, the optimization algorithms attempts to directly solve the optimal control necessary conditions 3.14 [8]. To obtain solutions from these necessary conditions, methods based on the special structure of the conditions themselves may be used, e.g. the so-called gradient methods (see e.g. [12, 13]). The solution can be alternatively obtained by numerically deriving the optimal control u^* from $\frac{\partial H}{\partial u} = 0$ (Equation 3.14e) using Newton's method, and solving the corresponding general boundary value problems (Equations 3.14c, 3.14d) by means of a multiple shooting method (see [28]) or a collocation method (see [15]). It is to be noted that, as it is necessary to explicitly derive the adjoint equation 3.14c and boundary equation 3.14d the user must have a deep insight into the physical and mathematical nature of the optimization problem. Furthermore, the indirect technique is rather complex and the introduction of the costate adjoint variable λ reduces the robustness of the algorithm as the numerical solution for the adjoint equations can be very ill-conditionated [8].

To overcome the intrinsic difficulties of the indirect method, the modern approach on numerical solution of the dynamic optimization problem focused on the so-called *direct methods*. Following this approaches, the optimal control problem is transformed into a nonlinear programming one. This can be done with the so-called direct shooting methods, relying on a parametrization of the control inputs u (e.g. piecewise

linear, piecewise constant, etc.) and the use of explicit numerical integration methods to evaluate the evolution of the system, using for example Euler, Runge-Kutta,or Hermite-Simpson methods (see [24] for an example). With the reformulation of the optimal control problem as a nonlinear programming (NLP) one, where the decision parameters are the discrete values of the control input trajectory deriving from the parametrization, it is possible to apply one of the methods for static optimization presented in the previous section, such as sequential quadratic programming SQP [5, 8]. The advantage of the direct method approach is that the user does not have to be concerned with the formulation of the state adjoint equations 3.14c. Moreover, even if the size of the optimization problem is usually increased when adopting a direct method, NLP tools can usually exploit the intrinsic sparsity of the direct formulation to guarantee faster solution than those achievable if exploiting the theoretically more simple boundary value problem from the indirect formulation.

Bibliography

[1] E. Aarts and J. K. Lenstra, *Local Search in Combinatorial Optimization*. New York, US: John Wiley & Sons, Inc., 1997. 29

[2] B. D. O. Anderson and J. B. Moore, *Optimal control: linear quadratic methods*. Prentice-Hall International, Inc., 1990. 33, 34, 35

[3] P. J. Antsaklis and A. N. Michel, *A Linear Systems Primer*. Boston, MA: Birkhäuser Boston, 2007. 35

[4] V. M. Becerra, *Optimal control*, vol. 3 ed., 2008. 35

[5] V. Becerra, "Solving optimal control problems with state constraints using nonlinear programming and simulation tools," *Education, IEEE Transactions on*, vol. 47, no. 3, pp. 377–384, 2004. 37

[6] D. Bertsekas, *Nonlinear programming*, 2nd ed. Belmont, MA: Athena Scientific, 1993. 25

[7] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 3rd ed. Athena Scientific, Belmont, Massachusetts, 2005, vol. I. 33, 34, 36, 51

[8] J. T. Betts, *Practical Methods for Optimal Control Using Nonlinear Programming*. Philadelphia: Society for Industrial and Applied Mathematics, 2001. 36, 37, 62, 64

[9] C. Blum and a. Roli, "Metaheuristics in combinatorial optimization: overview and conceptual comparison," *ACM Computing Surveys*, vol. 35, pp. 189–213, 2003. 30

[10] F. Borrelli, A. Bemporad, and M. Morari, *Predictive Control for Linear and Hybrid Systems*. Cambridge University Press, 2014. ix, 36, 60, 61

[11] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2010, vol. 25, no. 3. 23, 26, 27

[12] A. Bryson and Y. Ho, *Applied optimal control: optimization, estimation and control*. New York: Emisphere Publishing Corporation, 1975. 36

[13] F. L. Chernousko and A. A. Lyubushin, "Method of successive approximations for solution of optimal control problems," *Optimal Control Applications and Methods*, vol. 3, no. 2, pp. 101–114, Oct. 2007. 36

[14] D. D'Ambrosio, W. Spataro, R. Rongo, and G. Iovine, "2.7 Genetic Algorithms, Optimization, and Evolutionary Modeling," in *Treatise on Geomorphology*. Elsevier, 2013, vol. 2, pp. 74–97. 30

[15] E. D. Dickmanns and K. Well, "Approximate Solution of Optimal Control Problems Using Third Order Hermite Polynomial Functions," in *Lecture Notes in Computational Science*. Heidelberg: Springer, 1975, pp. 158–166. 36

[16] M. Dorigo and T. Stützle, "The Ant Colony Optimization Metaheuristic: Algorithms, Applications, and Advances," in *Handbook of Metaheuristics*. Boston: Kluwer Academic Publishers, pp. 250–285. 31

[17] S. E. Dreyfus, *Dynamic Programming and the Calculus of Variations*. New York: Academic Press Inc., 1965. 34

[18] H. P. Geering, "Optimal control with engineering applications," 2007. 34, 35

[19] F. Glover, "Tabu Search - Part I," *ORSA journal on Computing*, vol. 2 1, no. 3, pp. 4–32, 1989. 30

[20] J. H. Holland, "Outline for a Logical Theory of Adaptive Systems," *Journal of the ACM*, vol. 9, no. 3, pp. 297–314, Jul. 1962. 30

[21] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Neural Networks, 1995. Proceedings., IEEE International Conference on*, vol. 4, pp. 1942–1948 vol.4, 1995. 31, 41

[22] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by Simulated Annealing," *Science*, vol. 220, no. 4598, pp. 671–680, May 1983. 29

[23] J. R. Koza, *Genetic programming: on the programming of computers by means of natural selection*. Cambridge, MA, USA: MIT Press, 1992. 30

[24] D. Kraft, "On Converting Optimal Control Problems into Nonlinear Programming Problems," in *Computational Mathematical Programming*. Berlin: Springer Berlin Heidelberg, 1985, pp. 261–280. 37

[25] H. Kwakernaak and R. Silvan, "Linear Optimal Control Systems," 1972. 35

[26] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York, US: Springer, 1999. 22, 23, 24, 25, 27

[27] J. B. Rawlings and D. Q. Mayne, *Model Predictive Control: Theory and Design*. Nob Hill Publishing, LLC, 2012. 36, 62, 63, 174

[28] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, 2nd ed. Springer, 1983. 36

[29] E. G. Talbi, *Metaheuristics: From Design to Implementation*. New Jersey, US: John Wiley & Sons, Inc, 2009. 28, 30, 31

[30] D. Teodorović, P. Lucic, G. Markovic, and M. Dell'Orco, "Bee Colony Optimization: Principles and Applications," *8th Seminar on Neural Network Applications in Electrical Engineering*, vol. 00, pp. 151–156, 2006. 31

[31] V. Černý, "Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm," *Journal of Optimization Theory and Applications*, vol. 45, no. 1, pp. 41–51, Jan. 1985. 29

[32] C. Voudouris and E. P. K. Tsang, "Guided Local Search," in *Handbook of Metaheuristics*. Boston: Kluwer Academic Publishers, pp. 185–218. 30

[33] X. Yu and M. Gen, *Introduction to Evolutionary Algorithms*, ser. Decision Engineering. London: Springer London, 2010. 30

CHAPTER 4
NUMERICAL OPTIMIZATION: SELECTED TECHNIQUES

In the present chapter a detailed description is given of three different techniques for numerical optimization which will be largely used in the last chapters of the thesis, when applied to the solution of energy optimization problems. The first methodology, Particle Swarm Optimization (PSO), is a metaheuristic technique for the solution of static optimization problems. The second technique presented is an algorithm for the Dynamic Programming based numerical solution of a deterministic optimal control problem. Finally, a technique for constrained optimal control, namely Model Predictive Control, is introduced and an original algorithm is developed, based on the implementation of the PSO metaheuristic method, for the solution of nonlinear control problems.

4.1 Particle Swarm Optimization

The Particle Swarm Optimization algorithm is a metaheuristic population based method for static optimization first proposed by Kennedy and Eberhart [26]. The algorithm is inspired by the work of Reynolds [37] who studied the movement of bird flocks for visual computer simulation. Having observed that the flock appears to be under central control, he theorized that the unpredictable complex social behavior of the birds may arise from local processes based on a restricted set of simple rules. Kennedy and Eberhart developed an algorithm based on the same assumption, where the evolution of a population representing potential solutions to the optimization problem is guided by some few simple rules which produces a complex and effective social behavior. The derived Particle Swarm Optimization (PSO) method is a gradient free stochastic algorithm which closely resembles other algorithms belonging to the Evolutionary Algorithm class.

Numerous modifications have been proposed since the method was first presented mainly to improve the rate of convergence of the algorithm, increase the diversity in

the population and better perform with objective functions presenting multiple local minima, and deal with the presence of constraints. PSO algorithm has also been successfully applied to the solution of Multi-Objective Optimization problems [51], due to its ability (shared with other population based methods) to explore simultaneously multiple candidate solutions. Since its development, the PSO algorithm has been effectively applied to a vast number of different problems; its first and more fruitful applications involved neural network training, with the algorithm showing significantly better performances when compared to Gradient Descent algorithms [39]. Over the course of years, the algorithm spread to different fields, with some of the first applications outside the neural network training found in the work of Fukuyama and Yoshida [53], who applied a modified PSO to calculate the correct response to changes in the load on an electric grid, or Wegley [50] who applied the PSO method to optimize pump operations in water distribution systems. In the recent years, the algorithm became a common tool even for solving energy optimization problem (e.g. the optimal sizing of distributed energy resources in microgrid [32], the optimal design and management of a cogeneration system with energy storage [43], the management of a hybrid electric powertrain system [16], etc.). In the following section, the PSO algorithm will be presented along with its main modifications, introduced to increase the robustness of the method and to guarantee constraint handling ability.

4.1.1 The basic PSO algorithm

In the PSO algorithm, a population (swarm) of particles represents a set of potential solution to the static optimization Problem 3.1, here reported for the sake of clarity:

$$\min_{x \in X} J(x) \quad \text{subject to} \begin{cases} h_i(x) = 0, & i = 1, 2, ..., m \\ g_j(x) \leq 0, & j = 1, 2, ..., n \end{cases} \tag{4.1}$$

The optimization parameters x are assigned to the position of a generic particle i of the swarm. The movement of the particles inside the search space is then guided by the algorithm which updates each particle position and velocity based on the following equation, inspired by the perceived swarm behaviour:

$$v^i_{j+1} = w v^i_j + r_1 c_1 (p^i_{best} - x^i_j) + r_2 c_2 (g_{best} - x^i_j) \tag{4.2a}$$

$$x^i_{j+1} = x^i_j + v^i_{j+1}, \qquad \forall i \in S,\ j = 1, 2, ... \tag{4.2b}$$

where

S represents the s particles swarm ($S = \{1, 2, 3, ..., s\}$);

v is the particle velocity;
x is the particle position;
w is an inertia weight;
p_{best} is the particle personal best position;
g_{best} is the swarm global best position;
c_1 is the so called cognitive parameter;
c_2 is the so called social parameter;
r_1, r_2 are stochastic factors in the range $[0, 1]$;
i is the particle number;
j is the iteration number of the algorithm.

The first of the two update equations (4.2a) is the core of the PSO algorithm; it describes how the evolution of a particle trajectory is affected by its movement and the information coming from the rest of the swarm. It is composed of three different terms, namely:

1. an *inertia term* v_j^i, accounting for each particle propensity to follow its previous trajectory. This term is responsible for the tendency of continuing to explore the search space, preventing a premature convergence of the algorithm.

2. a *personal cognitive parameter* p_{best}^i, which serves as a "nostalgia" factor, attracting each particles towards the position where it performed best.

3. a *social cognitive parameter* g_{best}, which links all the particles together, attracting all the particles towards the global best position experienced by the swarm in its entirety.

At each step of the algorithm, the positions of the particles are updated based on Equation 4.2b and, after evaluating to cost function $J(x^i)$ associated to each particle, every individual best position p_{best}^i is eventually updated together with the global swarm best position g_{best}. The velocities can at this point be updated again and the algorithm progresses with the particles hopefully converging towards the global optimal solution of Problem 4.1. Figure 4.1 depicts a graphical illustration of the position and velocity update procedure for a set of particles moving in a 3-D search space.

The values of the resulting velocity updates v_{j+1}^i are usually clamped inside a range $[-v_{max}, v_{max}]$ to prevent the particles from straining out of the search space. If bounds of the type $[x_{min}, x_{max}]$ are present, the value of v_{max} is usually set equal to $v_{max} = k \cdot 0.5(x_{max} - x_{min})$, with $0.1 \leq k \leq 1.0$.

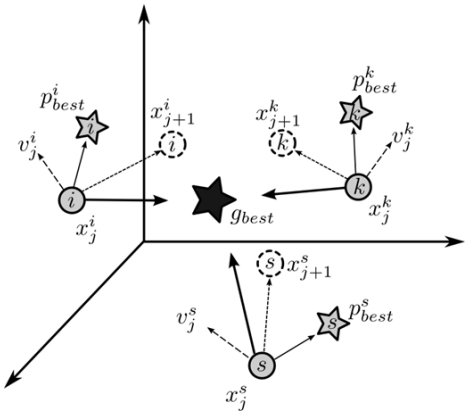


Figure 4.1: Illustration of position and velocity update for particles $\{i, j, k\} \in M$ at j-th algorithm iteration

The coordinates of each particle are randomly initialized at the start of the algorithm inside the search space. As many of the pseudo-random generators available lead to low-order correlations when used to generate random vectors, the initial positions can be distributed uniformly through the search space using sub-random sequences like Sobols sequence or a Latin hypercube distribution [10].

The PSO algorithm retains many of the peculiar characteristics of Evolutionary Algorithm: it relies on a population of individuals representing potential solutions to the optimization problem, and it uses the information of each generation of the algorithm to guide the evolution of the upcoming generation (offspring). While however the velocity update equation may resemble the arithmetic crossover operator found in Genetic Algorithms or a mutation operator whose strength relies on the distance from the "parents", a better way to understand the PSO algorithm is to portray it as a process of adaptation rather than a replacing process (death and birth) as it happens in Genetic Algorithms [27]. This feature represents the great difference that subsist between PSO (and other Swarm Methods) and Evolutionary Algorithms, with the first retaining information about position *and* velocity between each generation, contrary to the seconds which only keep track of position.

4.1.2 PSO algorithm improvements

Many improvements have been proposed to enhance the performances of the basic PSO algorithm. The first techniques focused on improving the rate of convergence of the algorithm, and usually involved changes in the terms of the velocity update

Equation 4.2a, without modifying the structure of the algorithm. One of the first approaches focused on the introduction and calibration of the inertia weight factor w with the work of Shi and Eberhart [41] who concluded that the best results leading to fast convergence were achieved with $0.8 \leq w \leq 1.2$. Lower values would prevent the swarm to stray out from its initial space search, while higher coefficients, even if allowing the exploration of wider areas, prevented a fast convergence of the algorithm. Another study focused on the determination of an optimal inertia weight coefficient was performed by Shi and Eberhart [42] who focused on a variable inertia weight and found a significant improvement in the performances with a weight linearly decreasing from 0.9 to 0.4 during the first 1500 simulations. Clerc [17] investigated the introduction of a *constriction factor* to ensure convergence. He proposed the modification of the velocity upgrade equation as follows:

$$v_{j+1}^i = \chi\left(v_j^i + r_1 c_1 (p_{best}^i - x_j^i) + r_2 c_2 (g_{best} - x_j^i)\right) \tag{4.3}$$

where

$$\chi = \frac{2}{\left|2 - \phi - \sqrt{\phi^2 - 4\phi}\right|} \tag{4.4}$$

and $\phi = c_1 + c_2,\ \phi > 4$. With Clerc formulation, the introduction of the constriction factor χ avoids the definition of velocity clamping terms $[-v_{max}, v_{max}]$. The author performed tests over a wide range of functions and proved the effectiveness of the proposed modification. Angeline [5] on the other hand introduced some concept of selection from the Evolutionary Algorithms with the purpose of increasing the focus of the algorithm over specific region of the search space that delivered promising solutions in the recent past. However, the solution gave contradictory results based on the test function adopted. This was related to the fact that, while the selection procedure improves the local search abilities of the PSO, it also hampers its global search effectiveness, forcing the swarm towards the first discovered local minima.

To address the problem of multiple local minima, many improvements have also been proposed. These approaches usually relies on the usage of a local best l_{best} social cognitive parameter, which reduces the iteration between a particle and the rest of the swarm. As an example, Suganthan [44] proposed a spatial location based method where at each iteration the distance of each particle from every other one is evaluated. Given the maximum distance d_{max}^k between two particles at the k-th iteration. The evaluation of the ratio $d_{i,l}^k = {||x_i^k - x_l^k||}/{d_{max}^k}$ is first performed, corresponding to the adimensional distance of a generic particle i from another particle j. Then a threshold, named $frac^k$, is defined and evaluated as follows:

$$frac^k = \frac{3\ k + 0.6\ k_{max}}{k_{max}} \tag{4.5}$$

with k the current iteration number and k_{max} the maximum iteration number allowed. With the above parameters, Suganthan suggested to define the set $\mathcal{N}_i^k$ of neighbours of the particle i as:

$$\mathcal{N}_i^k = \{x_l\} \mid d_{i,l}^k < frac^k, \quad l \in M \tag{4.6}$$

With these assumptions, the term g_{best} is replaced in velocity update Equation 4.2a by the term l_{best}^i defined as:

$$l_{best}^i = x_m \in \mathcal{N}_i^k \mid J(x_m) = \min\{f(x_l)\}, \quad \forall x_l \in \mathcal{N}_i^k \tag{4.7}$$

The ratio proposed by Suganthan gradually increases over the number of iterations, until, as it reaches 1, the local set $\mathcal{N}_i^k$ will coincide with the whole swarm and l_{best}^i will equal g_{best}. The adoption of local social parameters reduces premature convergence of the algorithm and maintains multiple attractors.

Constraint Handling As many other nature-inspired stochastic optimization methods, the PSO algorithm is formulated as an unconstrained optimizer. One of the consequences is that the basic PSO formulation does not have any built-in method that restricts the movement of the particles within the search space. To deal with this issue, repair algorithms are usually the most adopted technique as they can be easily adjusted to the PSO problem formulation. The repair method approach is based on the idea to map each infeasible particle $x \in \mathcal{S}$ back to a feasible one $z \in \mathcal{F}$ (with $\mathcal{S}, \mathcal{F}$ the set of infeasible and feasible solutions) by means of a *repair function* $f_{repair} : \mathcal{S} \to \mathcal{F}$. The repair approach is particularly suitable to deal with box constraints of the form $\mathcal{F} = [lb_1, ub_1] \times [lb_2, ub_2] \times \cdots \times [lb_n, ub_n]$, a condition rather common in PSO formulation. Figure 4.2 depicts a series of repair methods for handling particles violating the position constraints and mapping them back inside the feasible region, as illustrated in [24].

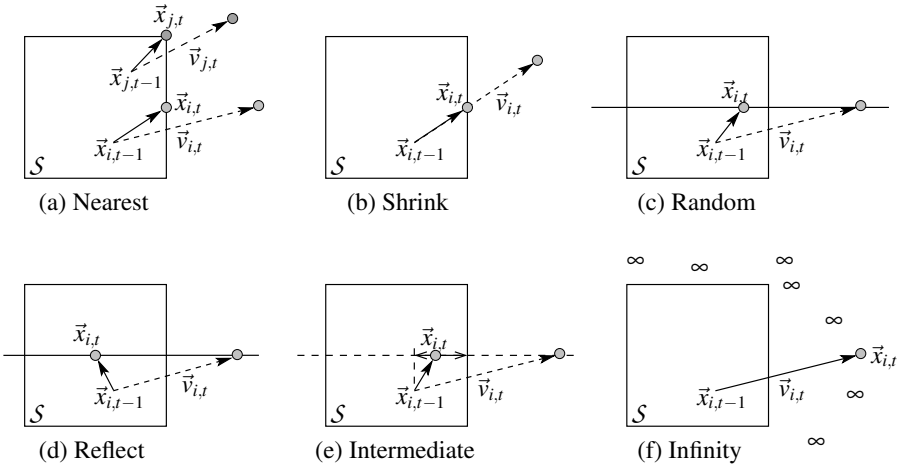


Figure 4.2: Various position handling strategies for PSO [24]

- *Nearest (a):* The infeasible particle is mapped back to the nearest boundary;
- *Shrink (b):* The particle is set to the intersection point of $x^i_{j+1} - x^i_j$ with the boundary;
- *Random (c):* Each infeasible particle is mapped back to a random position inside the feasible space;
- *Reflect (d):* The infeasible particle is reflected by the boundary;
- *Intermediate (f):* If the particle x^i_{j+1} is infeasible, it is set back to an intermediate value between x^i_j and the boundary;
- *Infinity (g):* The particle is allowed to stray out the feasible region but an infinite objective function is associated to all the unfeasible positions.

Other repair algorithms approaches do not try to restrict the movement of particles inside the boundary but focuses on techniques that will help to steer the particles back to the feasible domain. Robinson et al. [38] proposed for example three different approaches to deal with particles which escaped the boundaries:

1. *Absorbing walls:* The velocity of a particle escaping the feasible space is zeroed, as if the crossing of the boundary absorbs the particle energy;
2. *Reflecting walls:* The velocity of a particle escaping the feasible space is reversed and the particle is pushed back inside the feasible region;

3. *Invisible walls:* Particles are allowed to escape the feasible space but their fitness is not evaluated.

The three methods are observed to produce similar results, but the *Invisible walls* method has the additional advantage of reducing computation time as fitness evaluation is usually the most time consuming phase of the algorithm.

Other type of constraints (different from the simple box constraints) are dealt with following some of the well known methods for constrained optimization. A large literature of constraint handling techniques, developed for Evolutionary and Swarm Intelligence Algorithms, may be found in [30]; some of the most common and simple techniques are presented here. One of the oldest and most popular method relies on the introduction of a penalty function of the form:

$$p(x) = \sum_{i=1}^{m} r_i \cdot |g_i(x)| + \sum_{j=1}^{p} c_j \cdot |h_j(x)| \tag{4.8}$$

where g_i and h_j are respectively the equality and inequality constraints, and r_i, c_j are positive constants named "penalty factors". The penalty function is added to the original objective function $J(x)$ and the constrained problem is transformed into an unconstrained one. Interior and exterior penalty functions are commonly used in mathematical programming, with the first causing the penalty function to move towards infinity as the point moves towards the boundary, and the second reducing the penalty as the particle approaches the feasibility region from outside. Only external penalty functions are used in stochastic algorithms due to the reasonable assumption that the first generation of the algorithm may only contain infeasible solutions. Although their implementation is quite simple, the tuning of the penalty parameters is highly problem dependent and the adoption of alternative methods is usually preferred.

A method closely resembling the penalty function idea relies on the separation of the objective and constraints into a case function of the form:

$$J'(x) = \begin{cases} J(x) & \text{if feasible} \\ k\left(1 + \sum_{i=1}^{m} r_i \cdot |g_i(x)| + \sum_{j=1}^{p} c_j \cdot |h_j(x)|\right) & \text{otherwise} \end{cases} \tag{4.9}$$

with k tuned so that a feasible solution will always have a better fitness value if compared to an infeasible solution, and the fitness value of an infeasible solution is based only on the accumulated constraints violation. This method, even if in the basic formulation has been found to generate an important diversity loss, still provides the

basis of many more advanced methods for constraint handling in population based stochastic algorithms [30].

Many constraint handling approaches have also been tailored specifically to the PSO formulation; for example Sedlaczek and Eberhard [40] proposed an Augmented Lagrange Particle Swarm Optimization (ALPSO) based on the Lagrange multipliers approach to constrained optimization, while Vandeerplaats [47] suggested to ignore the inertia term in the velocity update equation for infeasible particles, to more quickly steer the particle back to the feasible region.

4.2 Dynamic Programming

The Dynamic Programming (DP) algorithm is a very powerful numerical method for dealing with optimal control problems (see Chapter 3.2). The term "Dynamic Programming" originates from R. Bellman who described the process he developed for solving problems in a recursive form, by nesting smaller decision problems inside larger ones. The concept of DP was firstly formulated between 1948 and 1952, as an attempt of solving real world application of game theory inside the RAND Corporation (e.g. the determination of optimal use for guided missiles against enemy targets [7]). While such problems may be easily formulated as typical multistage decision processes, the techniques available at the time (mainly based on classic calculus of variations) could not be applied easily, and the numerical alternatives were computationally prohibitive. With the introduction of the principle of optimality, Bellman discovered a technique able to effectively circumvent the dimensional explosion for multistage process optimization. This approach is based on the observation that, citing Bellman, "An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision". Thanks to this property it is possible to formulate a recursive method for the solution of multistage problems which significantly decreases the amount of sub-problems to be solved in an iterative formulation. Shortly after its development, the DP algorithm was successfully applied to optimal control theory, by representing the control process as a multistage process and thus overcoming the limitations of calculus of variations. A mathematical representation of Bellman's optimality principle is depicted in Chapter 3.2 with Equation 3.11.

Ideally, the Dynamic Programming method would guarantee global optimality of the solution, regardless of the type of problem addressed [19]. However, if stochastic phenomena are present, the algorithm can only provide a solution based on the expectation of the evolution of the system. In this case, only sub-optimal controllers

may be developed, whose efficacy depends on the ability of correctly representing the effect of random influences. A rather useful application of the DP algorithm in these circumstances consists in its application to solve the corresponding deterministic problem, based on the non-physical assumption that all the future disturbances are known. Even if this procedure returns a non-causal controller (i.e. a controller which actions are based on the knowledge of future events), it provides a benchmark to which all the causal sub-optimal controllers may be compared to, and it indicates the best performances that the system could potentially achieve. The knowledge of the non-causal optimal control policy may also be exploited to develop rule-based algorithms, by properly analyzing the actions taken by the controller with respect to the evolution of the random variables (see for example [33], where the authors derive a real-time strategy for the optimization of a HEV, based on the results from the deterministic DP solution of the problem).

Today the Dynamic Programming algorithm is widely used in numerical multi-stage optimization processes. Many examples may be found in the field energy management, e.g. with the optimal control of hybrid vehicles [6, 34], the optimization of CHCP systems operation strategy [21], the scheduling of residential energy [52], etc.

The following section presents the mathematical formalization of the basic DP problem for a generic optimal control problem. An algorithm for the numerical implementation of the DP methodology, developed by the ETH, is presented. The algorithm relies on a discretization of the state space of the system and is based on a deterministic formulation of the optimal control problem. The discretization procedure gives rise to many numerical problems, related to both algorithm accuracy and computational time. Some modifications for increasing the accuracy of the basic algorithm are therefore illustrated along with other modifications which attempt to reduce the computational effort by sacrificing either global optimality or flexibility.

4.2.1 The basic DP problem

The Dynamic Programming (DP) paradigm may be used to address the solution of the optimal control problem:

$$\min_{u(t)} J(x_0, u(t)) \tag{4.10a}$$

$$s.t. \quad \dot{x}(t) = f(x(t), u(t), w(t), t) \tag{4.10b}$$
$$x(0) = x_0 \tag{4.10c}$$
$$x(t_f) \in [x_{f,min},\ x_{f,max}] \tag{4.10d}$$
$$x(t) \in X(t) \subset \mathbb{R}^n \tag{4.10e}$$
$$u(t) \in U(t) \subset \mathbb{R}^m \tag{4.10f}$$

with an objective function of the form:

$$J(x_0, u(t)) = G(x(t_f)) + \int_0^{t_f} H(x(t), u(t), t)dt \tag{4.11}$$

The problem has a fixed final time and a partially constrained final state. Furthermore, the problem is characterized by time-variant input and state constraints. A first step to implement Bellman's DP algorithm [12] requires that the continuous time model 4.10b is discretized with a sample time $T = {}^{t_f}/_{N}$:

$$x_{k+1} = f_k(x_k, u_k, w_k), \qquad k = 0,\ 1,\ ...,\ N-1 \tag{4.12}$$

where

k is the discrete time index;
x_k is the state of the system, a vector of the space X_k ;
u_k is the decision variable to be selected at time k, a vector of the space U_k;
w_k is a random variable representing external disturbances;
N is the horizon of the optimization, i.e. the number of time samples.

With the control u_k constrained to a subset $U_k(x_k) \subset U_k$, depending on current state x_k, it is possible to define the class of *admissible* control laws (policies) $\pi = \{\mu_0, \mu_1, ..., \mu_{N-1}\}$, with μ_k mapping state x_k into controls $u_k = \mu_k(x_k)$ such that $\mu_k(x_k) \in U_k(x_k), \quad \forall x_k \in X_k$. Because of the presence of the random variable w_k, the cost function J cannot be meaningfully expressed, so the general optimization problem is formulated based on the expected cost, given the initial state x_0 and the admissible policy π, defined as:

$$J_\pi(x_0) = E\left\{ g_N(x_N) + \sum_{k=1}^{N-1} g_k(x_k, \mu_k(x_k), w_k) \right\} \tag{4.13}$$

where E symbolize the expectation operator with respect to the joint distribution

of the random variables involved. The solution of the optimal control Problem 4.10 requires the evaluation of the optimal policy π^* that minimize the cost:

$$\pi^* = arg \min_{\pi \in \Pi} J_\pi(x_0) \tag{4.14}$$

with Π the set of admissible policies. The Dynamic Programming solution of Problem 4.10 exploits Bellman *principle of optimality* 3.11 [8] which may be reassumed into the intuitive assertion that, if a policy $\pi^* = \{\mu_0^*, \mu_1^*, ..., \mu_{N-1}^*\}$ is optimal in the sense of Equation 4.14, then the truncated policy $\{\mu_i^*, \mu_{i+1}^*, ..., \mu_{N-1}^*\}$ is optimal for the "tail" subproblem of minimizing $J_\pi(x_i)$. This implies that an optimal control policy may be constructed in a piecemeal fashion, starting from solving subproblem 4.10 for $x_i = x_{N-1}$ and proceeding sequentially backwards in time extending the "tail" subproblem to the previous step until the initial condition x_0 is reached. Ideally, if an analytical solution of the "tail" subproblem can be found, the DP algorithm would provide a close-form expression for the cost J_k or the optimal policy π^*; this is the case for example of LQR or linear MPC controllers. Unfortunately, in many cases it is not possible to obtain an analytical solution and numerical methods must be applied. Even if in these cases the application of the DP algorithm is quite time-consuming, it still can be used as the basis for more practical sub-optimal approaches.

4.2.2 Deterministic DP algorithm

An algorithm providing a numerical solution to the deterministic formulation of Problem 4.10 has been developed by Sundström and Guzzella [46]; in this case the trajectory $w(t)$ of the future external disturbances is assumed to be perfectly known. This solution is obviously not causal but it provides an useful benchmark against which a causal controller can be compared. The basic formulation of the algorithm is based on the following formulation of the discretized cost:

$$\begin{aligned} J_\pi(x_0) =& g_N(x_N) + \phi_N(x_N) + \\ &+ \sum_{k=0}^{N-1} h_k(x_k, \mu_k(x_k), w_k) + \phi_k(x_k) \end{aligned} \tag{4.15}$$

where g_N and h_k are respectively the final and stage costs, while ϕ is a penalty function added to enforce state constraints. As the solution of the DP problem requires the knowledge of the optimal trajectory for every x_k in the state space, to proceed with the mathematical implementation of the algorithm, for continuous problems ($x \in X \subset \mathbb{R}^n$), the state space is spatially discretized. The notation x_k^i

indicates the state variable x at the node of time-index k and space-index i. The DP solution of the problem is obtained by evaluating backward in time the optimal "cost-to-go" function $\mathcal{J}_k(x^i)$ (related to the tail subproblem) and its associated optimal control policy π_k^i based on the following algorithm:

1. For the final step N, the cost-to-go function is:

$$\mathcal{J}_N(x^i) = g_N(x^i) + \phi_N(x^i) \tag{4.16}$$

2. For the intermediate step k, the cost-to-go function is:

$$\begin{aligned}\mathcal{J}_k(x^i) = \min_{u_k \in U_k} \{h_k(x^i, u_k, w_k) + \phi_k(x^i) + ... \\ + \mathcal{J}_{k+1}(F_k(x^i, u_k, w_k))\}\end{aligned} \tag{4.17}$$

As the updated state $F_k(x^i, u_k, w_k)$ is continuous in the state space while $\mathcal{J}_{k+1}$ is known only at the grid points x^i, the term $\mathcal{J}_{k+1}(F_k)$ is interpolated; usually a linear interpolation is adopted, as the computational cost of the operation is negligible if compared to the cost of the function evaluation. The algorithm estimates an optimal map, assigning to each point x_k^i an optimal sub-tail trajectory π_k^i. When applying this map to the forward simulation of the problem, an additional interpolation of the control signal is performed, as the actual state will not coincide with a point in the grid. Sundström and Guzzella [46] applied the developed DP algorithm to the solution of the well-known Lotka-Volterra fishery problem and to an hybrid-electric vehicle application, with the task of minimizing the total fuel mass consumption by properly controlling a torque split factor u, obtaining good performances of the algorithm in both cases.

4.2.3 DP modifications for accuracy increase and reduced computational cost

The Dynamic Programming algorithm developed by Sundström and Guzzella relies on the discretization of the state space to obtain a solution in the case that an analytical one cannot be assigned continuously to each state $x_k \in X_k \subset \mathbb{R}^n$. This discretization obviously comports a reduction in the algorithm accuracy which can be avoided only by selecting a densely gridded space map; on the other hand, the number of function evaluation and the computation time increase significantly with the increase of the

grid resolution. The number of function evaluation for the basic DP algorithm is equal to:

$$N_{feval}^{DP} = \prod_{i=1}^{n} N_{x,i} \cdot \prod_{j=1}^{m} N_{u,j} \cdot N \tag{4.18}$$

where n is the number of states, $N_{x,i}$ is the number of grid points for the generic state x_i, m is the number of inputs, $N_{u,j}$ is the number of grid points for the generic input u_j, and N is the number of samples. This highlights what is probably one of the greatest difficulties associated with the implementation of the DP algorithm, the so-called "curse of dimensionality". As the number of states increases, in fact, not only the computational time but also the amount of memory required to store the "sub-tail" solutions will rapidly increase, limiting the possibilities of successfully implementing the procedure.

Many techniques have been proposed to deal with the curse of dimensionality. A great amount is based on the introduction of approximated function structures for the evaluation of the cost function, generating the family of so-called Adaptive Dynamic Programming (A-DP) algorithms; an extensive review of these methods can be found in [48]. As an example of an A-DP application, Bertsekas and Tsitsiklis [11] developed the core for a neuro-dynamic programming methodology, based on a stochastic approximation methods for the evaluation of the cost function relying on neural network architectures. Other authors proposed modifications of the gridding procedure, as Luus [28] for example, who developed a method, called Iterative DP, to dynamically change the grid size based on the information gained about the candidate best trajectories.

While these methods efficiently tackle the curse of dimensionality, they sacrifice either the global optimality property of the algorithm or its flexibility. A much simpler way to address the problem of computational cost is that of reducing the number of grid points for state discretization. Even if this solution may appear trivial, it has the advantage of maintaining the most important property of the basic DP formulation, i.e. the guarantee of global optimality of the solution. Some proper countermeasures to preserve the accuracy of the algorithm must however be taken.

DP accuracy improvement techniques As Sundström and al. observed [45], in real life problem, due to the presence of final state constraints in the formulation of the problem, the optimal state trajectory will generally be located at the boundary between feasible and unfeasible regions. When a more sparse state-space gridding is

adopted, the greatest accuracy loss derives from the interpolation of the "cost-to-go" function for points which are near this boundary.

If, as in the basic DP formulation, an infinite cost $\phi_k(x^i \notin X_k) \to \infty$ is assigned to infeasible points, it may occur that the infeasibility penalization would extend even to points which are actually located inside the feasible region, with the practical effect of enlarging the extension of the estimated unfeasible subspace. To better understand this concept, it is possible to refer to Figure 4.3.

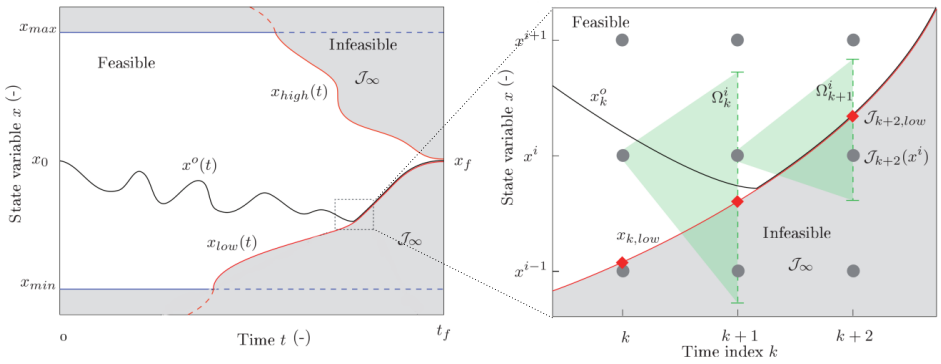


Figure 4.3: Cost-to-go interpolation near boundary line [45]

Let Ω_k^i define the set of reachable states from point x^i at time step k over one step:

$$\Omega_k^i = \{x \mid x = F_k(x^i, u_k, w_k) \; \forall u_k \in U_k\} \tag{4.19}$$

Then, focusing on point $x_{k+1}^i \in X_{k+1}$, it is possible to see that the entire set Ω_{k+1}^i of reachable states lies between the feasible grid point x_{k+2}^{i+1} and the infeasible point x_{k+2}^i (with has an associated infinite cost-to-go $\mathcal{J}_{k+2}(x^i) \to \infty$ due to the presence of the infinite penalty factor $\phi_{k+2}(x^i)$). Even if the cost associated to x_{k+2}^{i+1} is finite and part of the reachable set Ω_{k+1}^i is feasible, the cost-to-go $\mathcal{J}_{k+1}(x^i)$ associated to x_{k+1}^i is infinite as it is based on the interpolation between a finite and an infinite cost-to-go. This first interpolation error produces even more serious accuracy losses as the algorithm progresses. When evaluating the cost-to-go function from point x_k^i in fact, even if a portion of Ω_k^i lies entirely between two feasible solutions, x_{k+1}^i

and x_{k+1}^{i+1} (with associated null penalty terms $\phi_k(x_{k+1}^{i+1}) = \phi_k(x_{k+1}^i) = 0$), an infinite cost is nevertheless assigned to the trajectories starting from state state x_k^i, due to the erroneous infinite cost $\mathcal{J}_{k+1}(x^i)$ associated to the tail sub-problem from x_{k+1}^i evaluated in the previous iteration. If the analytic solution to the problem implies that the optimal state trajectory must pass close or even overlap the boundary between the feasible and unfeasible region, the infinite penalty factor approach would greatly affect the accuracy of the solution provided by the DP algorithm.

One of the first solution proposed was that of applying a great but finite penalty to the unfeasible states; the choice of a good penalty factor would however introduce a problem-dependent calibration parameter, thus weakening the flexibility of the algorithm. On the other hand, if the exact boundary line between backward reachable and unreachable states is properly evaluated, this additional knowledge may be used to effectively increase the interpolation accuracy. Sundström et al. [45] followed this approach when proposing a modification of the algorithm, called the *boundary line* method.

Boundary line DP The algorithm applies only to one-dimensional problems and is composed of two stages:

1. *Computation of the boundary line:* the boundary lines between feasible and unfeasible state points are evaluated before running the DP algorithm. Because the problem is one dimensional, only two boundary lines (lower and upper) are present. If $x_{f_{min}}$ is the minimal final state, the lower boundary line is evaluated by solving backward in time the following problem:

$$\begin{aligned} & \min_{x_{k,low}, u_k} x_{k,low} \\ s.t. \quad & x_{k+1,low} = F_k(x_{k,low}, u_k, w_k) \\ & u_k \in U_k \\ & x_{k,low} \in X_k \end{aligned}$$

   with $x_{N,low} = x_{f_{min}}$. The cost-to-go $\mathcal{J}_{k,low}$ and the minimizer $u_{k,low}$ are also stored. It is to be noted that in this case, the lower boundary line is continuous in the state variable and only discretized in time. The same procedure is applied to evaluate the upper boundary line;

2. *Interpolation near the boundary line:* if during the DP algorithm the set Ω_k^i contains the boundary line, the interpolation is performed between the exact boundary line point and the nearest feasible state.

The authors tested the method on the Lotka-Volterra fishery problem and the hybrid-electric vehicle application and found that it improved the efficiency of the algorithm by providing the same accuracy of the basic DP algorithm even if a much smaller state-space resolution was adopted, with the computational cost reduced by 13 times.

Level-set DP Elbert et al. [19] ulteriorly adapted the DP algorithm to comply with final state constraint for n-dimensional problems. One of the main problems in dealing with backward reachability for $n \geq 2$ is that the backward reachable space might generally be non-convex, even if the terminal target set is convex (see Figure 4.4).

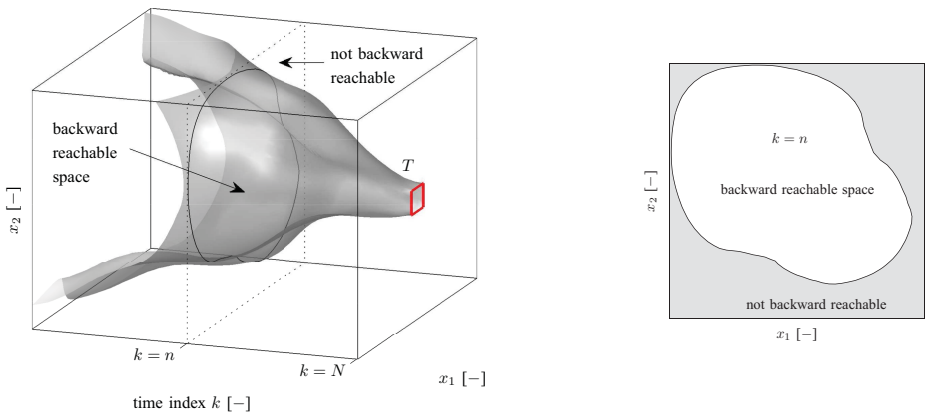


Figure 4.4: Backward reachable state for a generic two states system [19]

To tackle the problem of representing the boundary of the backward-reachable subspace X_k, the authors introduce a level-set function $\mathcal{I} : X \subset \mathbb{R}^n \to \mathbb{R}$ such as that $X_k = \{x \in X \mid \mathcal{I}(x) \leq 0\}$; this function assumes a negative value for feasible states and a positive value for unfeasible one (see Figure 4.5).

The use of a level-set function allows for an easy representation of the non-convex regions and can be applied to a system with any number of states. While the evaluation of $\mathcal{I}(x)$ is performed only for the nodes of the discretized state-space, feasibility of points outside the grid can be easily evaluated through interpolation. If the level-set function $\mathcal{I}_N(x)$ is defined based on the target set X_N such as that $\mathcal{I}_N(x^i)$ is equal

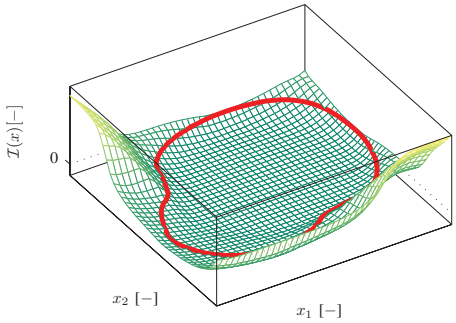


Figure 4.5: Example of a level set function $\mathcal{I}(x)$, with x feasible if in the lower region [19]

to the distance of the generic grid point x^i from the boundary (positive if outside and negative if inside), the backward progression of the level-set function may be estimated with the DP algorithm for $k = N - 1$ to 0:

$$\mathcal{I}_k(x^i) = \min_{u_k \in U_k} \left\{ \mathcal{I}_{k+1}(F_k(x^i, u_k, w_k)) \right\} \tag{4.21}$$

With this procedure, it is possible to define the set of feasible control signals $U_k^F(x^i) = \left\{ u_k \in U_k \mid \mathcal{I}_{k+1}\Big(F_k(x^i, u_k, w_k)\Big) \leq 0 \right\}$ which leads the evolution of the system from a generic point x_k^i to a feasible updated state subspace. The evaluation of the original "cost-to-go" function $\mathcal{J}_k(x^i)$, Equation 4.17, is performed only based on the feasible set $U_k^F(x^i)$ rather than on the entire set U_k. If the feasible set is empty, i.e. $U_k^F(x^i) = \emptyset$, the optimal trajectory is that which minimizes the level-set function. The algorithm was effectively implemented and the tests performed by the authors over a simple 2-D optimal control problem showed that the level-set DP algorithm performs consistently better when compared to the basic formulation. With the increase in the number of states, the benefits deriving from an effective evaluation of the exact boundary are even more evident, with the level-set algorithm achieving the same accuracy performance of the basic DP formulation while using a grid of points 334 times smaller.

4.3 Model Predictive Control

Model Predictive Control is a dynamic optimization strategy that seeks to exploit the knowledge of the process, coming from a suitable developed plant model, to predict

the future evolution of the system and optimize the control signal accordingly. The method was first introduced during the 1980s to improve the control performance for oil and chemical industrial processes, in an attempt to tackle the multivariable constrained formulation of the problem typical of those applications. The method has spread since then to various different industrial applications thanks to its flexibility and robustness (see [31] for a short historical review of the method). In the present section, a short literature review on Model Predictive Control is presented, with some brief consideration about the implementation of the algorithm for both the linear and nonlinear cases and the introduction of the so-called "economic MPC" problem formulation. In the second part of the section, an original method for the solution of the economic nonlinear MPC problem formulation is presented, relying on the application of the PSO algorithm for the numerical solution of the parameterized constrained optimal control problem.

4.3.1 MPC theory

The Model Predictive Control approach to the dynamic optimization problem is based on the estimation of an open loop optimal control policy which will theoretically minimize the desired objective function over a finite future horizon. To obtain such policy, the measurements acquired from the plant are used to initialize a model of the same process which is subsequently used to estimate the future evolution of the state of the system and the objective function based on the generic control trajectory that can be applied during the finite optimization horizon, usually referred to as the *prediction horizon*. The evaluation of the optimal control policy, given the prediction from the model, is a mathematical optimization problem and many different strategies can be adopted to solve it, based on the formulation of the estimation model and the objective function.

Once such an optimal policy is obtained, rather implementing it in its entirety for the entire duration of the prediction horizon, only its initial part is actually implemented, until a new measurement from the plant is available. Once the real updated values of the systems parameters are known, the model is re-initialized and a new optimization procedure is conceived; this procedure is then repeated in a recursive iterative manner, hence the alternative nomenclature of Receding Horizon Control (RHC). Figure 4.6 depicts a graphical representation of the iterative optimization procedure of a sampled system, where a piecewise constant input trajectory $u(t)$ is optimized at time t based on the predicted output from the system model $y(t+k|t)$ over a prediction horizon composed of N_P steps of the system sample time. Notice

that the control trajectory is optimized only over the first N_C steps, with $N_C \leq N_P$, and kept constant for the rest of the prediction horizon, a rather common approach in MPC as it allows to lessen the computational burden required for solving the optimization problem [49]. The computed optimal input trajectory is then applied only during the sampling interval $[t, t+1]$ and a new optimization problem is formulated once measurements from the new system state at time step $t+1$ are obtained.

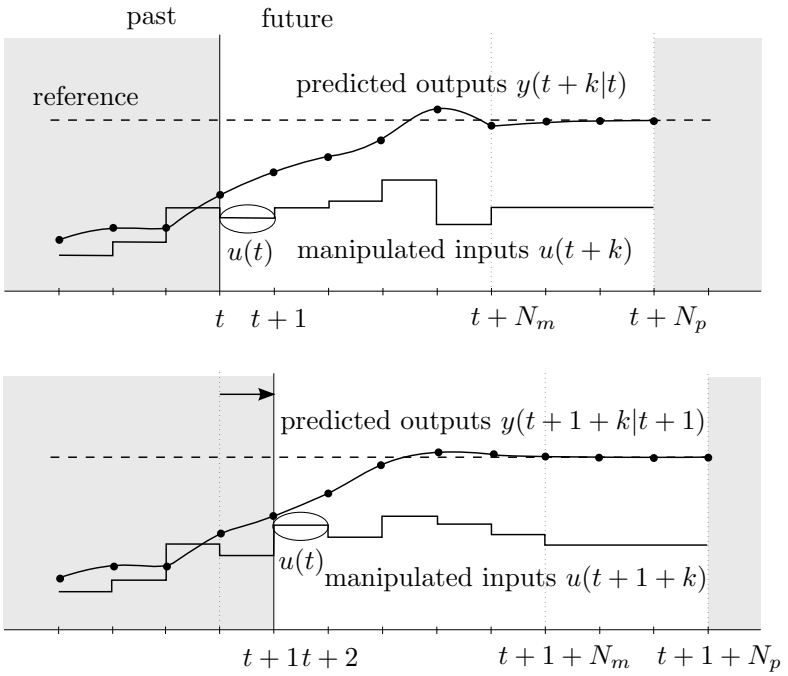


Figure 4.6: Schematic of Receding Horizon Control [14]

One of the great advantages deriving from the MPC formulation relies in the possibility of directly account for the system constraints during the optimization procedure. Contrary for example to the LQR approach which can only deal with unconstrained problem, or the adoption of PID controllers which requires the introduction of additional counter-measures for compensating the effect of control saturation (e.g. anti wind-up), in the MPC approach the constraints are directly embedded into the problem formulation. This avoids the necessity of any additional counter measure for compensating the physical constraints on control inputs or enforcing the constraints on system variables, which would complicate the structure of the controller and reduce the optimality of the solution.

Moreover, the iteration of the optimization procedure, typical of the Receding Horizon Control formulation, provides a feedback on the real evolution of the process, allowing the adoption of corrective measures to adapt the action of the control system to the effect of external disturbances and comply with the errors from process modelling.

Linear MPC If the system to be optimized can be described by a linear discrete time model of the form:

$$x(k+1) = Ax(k) + Bu(k), \qquad x(0) = x_0 \tag{4.22}$$

subject to a series of linear input and states constraints of the form:

$$Ex + Fu \leq Y \tag{4.23}$$

and a quadratic objective function is to be optimized, of the form:

$$J(x_0, u) = x(N_P)^T P_f x(N_P) + \sum_{i=0}^{N_P-1} x^T(i) Q x(i) + \sum_{i=1}^{N_C-1} u^T(i) R u(i) \tag{4.24}$$

then the solution of the MPC optimization may be reduced to a Quadratic Programming problem by applying Bellman optimality principle [14] and the techniques developed to address the QP static problem may be easily adapted the MPC dynamic optimization. While linearities of the plant model usually come from simplifications or linearizations of more complex models, the linear input output constraints may be used to effectively represent the common situation of box constraints of the form $u_{min} \leq u \leq u_{max}$, $x_{min} \leq x \leq x_{max}$, which can efficiently represent actuators saturation and operation constraints. Moreover, the quadratic formulation of the objective function is the basis for tracking, where the objective of the controller is to bring the system into a desired configuration rapidly and with some guaranteed stability performances. Hence the great interest that linear MPC still retains in the world of optimal control.

A critical feature of the basic QP formulation is that it requires an implicit online solution of the optimization problem, to be evaluated by the mathematical programming algorithm at each time step. Regarding this matter, it is worth being cited the work of Borrelli et al. [14] who proposed an alternative approach to derive an explicit piecewise linear correlation between the measure state x_0 and the optimal control

trajectory u^*. The method involves the off-line pre-computation of the correlation thanks to a multi-parametric programming approach is adopted. The coefficients for the piecewise linear correlation are then mapped based on the system state and stored in the form of a lookup table that can guarantee a fast and efficient on-line implementation.

A large literature is available on the subject of linear MPC, focusing mainly on conditions for ensuring the closed loop stability of the algorithm (see [9, 36]).

Nonlinear MPC In non-linear system and objective functions are accounted, the solution of the MPC problem requires much more complex approaches. Among the various solutions, the approaches relying on "direct methods" are by far the most used; in this case the solution of the dynamic optimization problem first relies on an approximation of the control trajectory $u(t)$ over the prediction horizon T_P, based on a time discretization of the trajectory, which allows the reduction of the number of parameters to be optimized. The estimation of the system evolution and the objective function is then performed applying numerical methods (e.g. Euler, Runge-Kutta, Hermite-Simpson, etc.) and the dynamic optimization is formulated as a Non-Linear Programming problem, thus exploiting the algorithms depicted in Chapter 3.1.1. Given the particular structure of the problem, some properties of the ordinary differential equations describing the approximated evolution of the system (e.g. sparsity) can be used to simplify the formulation of the NLP problem (see for example [13]).

Economic MPC Since its first implementation in the chemical field, MPC has been historically developed mainly as a method for process control, i.e. a control algorithm whose objective is that of guaranteeing that some output of the process are kept inside a specific range or that they converge rapidly towards a target point regardless of the presence of external disturbances (set-point tracking, noise rejection). In this case, the objective function in the optimal control problem 3.8a is usually formulated as a quadratic function of the error between current system state and desired conditions. The target for the design of an MPC controller are in this case stability and robustness, i.e. guaranteed performances even with model uncertainties. A large part of the theoretical literature on MPC algorithms focuses therefore on criteria for ensuring such desirable properties [15, 29, 36].

The high level objective for the optimization of a plant process on the other hand is always related to some "economic" parameters such as, for example, the net profitability of the process, the efficiency of the operation, the energetic efficiency, etc. In

most of the industrial applications however, this problem is not directly approached in its entire complexity but it is rather split into a number of different subproblems. The control algorithm is usually partitioned into two different layers, which separates the actual control phase from the information management and the decision making [22]. The higher layer of this structure, often referred to as Real-Time Optimization (RTO), acts as a supervisory technique and is responsible for the evaluation of the optimal steady state operating conditions for the plant which guarantees the maximization of the economic performance index. RTO methods are usually based on a non-linear steady-state model of the process, eventually comprehending some sub-routines for data validation or parameter estimation [23], and they require the solution of a static optimization problem. After the supervisor has evaluated the optimal target operating conditions, these values are passed to a lower level advanced controller which is responsible for the implementation of such targets and which also takes care of stability and rapid convergence. The objective function implemented into the low level controller is therefore unrelated to the economic performance and is rather aimed at improving the tracking and disturbance rejection abilities. MPC algorithms are normally employed in the low level part of this hierarchical optimization structure, due to their flexibility, robustness and their ability to effectively deal with plant constraints on both input and states.

For many problems, the majority of the operating time is effectively spent at near steady conditions and the hierarchical division of the control action is proved to be very effective. The paradigm of separating the control action into two different layers relies however on the assumption that process disturbances and transients may be neglected if the high level optimization is executed in time intervals long enough to allow the process to reach steady state [1]. If on the other hand the economic objective is influenced by unpredictable rapidly varying external disturbances, the basic assumption that transients can be neglected does not hold any more, as the system may never even reach steady-state conditions, and the resulting control policy can only be sub-optimal [3, 36]. Many solutions have been proposed to deal with the problem (see for example [20]); some of these, named "frequent RTO", involve that the set points for the regulatory layer are chosen in much shorter time intervals; others are based on the integration of steady-state optimization into the MPC with the so-called LP-MPC and QP-MPC two stage structures, where an upper MPC layer computes the target setpoints for both the controlled and manipulated variables, while the lower MPC layer effectively enforces the implementation of the solution, based on the information from both the upper MPC and the RTO layers. Still retaining

a hierarchical structure of the problem formulation, dynamic real time optimization (D-RTO) [25] requires that dynamic informations are added to the RTO formulation so that an optimal trajectory, rather than an optimal set-point, is fed to the lower regulator.

An alternative solution to the two layer structure involves the implementation of the economic information directly into the controller layer, thanks to a suitable modication of the objective function. Some basic approaches [54] are limited to the addition of some economic related terms to the tracking performance index, in the form for example of a terminal cost. While this method has been proven effective, it still requires the presence of an RTO layer for the generation of the reference set-point for the formulation of the tracking performance index. If however an additional step is taken, the RTO layer may be eliminated entirely by replacing the tracking cost with the economic stage cost function normally used in the upper layer [18]. This approach, known as "economic MPC", was originally developed by Rawlings et al. [35] and presents some peculiar features; for example, preserving the original economic cost in the objective formulation has the positive effect of slowing down convergence towards an equilibrium point if such a transient occurs in a region where the stage cost is better than at steady state, whereas the traditional tracking controller will simply try to move towards the steady state in shortest possible time. Furthermore, the optimal operating point for the system may not even be at steady-state conditions, with the system operating with a periodic/cyclic behaviour [4]. As the economic MPC problem is usually nonlinear in the objective and the constraints, numerical methods for the solution are usually based on the direct NLP approach (see Section 3.2). For more in depth details about stability and numerical implementation of the method, details can be found in [2].

4.3.2 Economic PSO-based non-linear MPC

In this thesis, an original method for the solution of the nonlinear optimization Problem 3.8a is developed. The algorithm is based on the economic MPC formulation, but it exploits a metaheuristic optimization method, rather than a NLP algorithm, for the numerical solution of the receding horizon optimization problem. This method follows a sequential direct approach [13] and employs the PSO algorithm for the optimization of a piecewise linear parametrization of the control input u.

The optimal control Problem 3.8a may be formulated as follows for the receding horizon framework:

$$\bar{u}^*(\cdot) = arg \min_{\bar{u}(\cdot)} J(x(t), \bar{u}(\cdot)) = E(\bar{x}(t+T_P)) + \int_t^{t+T_P} L(\bar{x}(\tau), \bar{u}(\tau)) d\tau \tag{4.25a}$$

$$\begin{aligned}
s.t. \quad & \dot{\bar{x}}(\tau) = f(\bar{x}(\tau), \bar{u}(\tau), \bar{w}(\tau)) && \bar{x}(t) = x(t) && (4.25b)\\
& \bar{y}(\tau) = h(\bar{x}(\tau), \bar{u}(\tau)) && && (4.25c)\\
& \bar{u}(\tau) \in U && \forall \tau \in [t, t+T_C] && (4.25d)\\
& \bar{u}(\tau) = \bar{u}(t+T_C) && \forall \tau \in [t+T_C, t+T_P] && (4.25e)\\
& \bar{y}(\tau) \in Y && \forall \tau \in [t, t+T_P] && (4.25f)\\
& \bar{x}(t+T_P) \in X_N && && (4.25g)
\end{aligned}$$

where $T_C \leq T_P$ are respectively the control and prediction horizons, $L(\cdot,\cdot) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is the stage cost, $E(\cdot) : \mathbb{R}^n \to \mathbb{R}$ is the terminal cost, and X_N is the terminal state constraint set. The cost function J is evaluated with respect to the predicted state and input trajectories $(\bar{x}(\tau), \bar{u}(\tau))$, evaluated with the aid of a plant model and distinguished from the real trajectories $(x(t), u(t))$. Solution of Problem 4.25 yields the open-loop optimal control law $\bar{u}^*(\cdot) : [t, t+T_P] \to \mathbb{R}^n$ to be applied by the control algorithm at time t. It is to be noted that, due to the presence of an external unknown disturbances (embedded in the term $w(t)$), the solution is not deterministic and the real evolution of the system will likely diverge from the prediction. Hence the necessity of iterating the prediction and optimization phases, as the system evolves, in a receding horizon mechanism to provide a feedback mechanism which is able to correct the effect of the unknown disturbances.

The problem is assumed to be subject to input/output box constraints of the form:

$$u(t) \in U := \{u \in \mathbb{R}^m | u_{min} \leq u \leq u_{max}\} \tag{4.26a}$$

$$\dot{u}(t) \in \dot{U} := \{\dot{u} \in \mathbb{R}^m | \dot{u}_{min} \leq \dot{u} \leq \dot{u}_{max}\} \tag{4.26b}$$

$$y(t) \in Y := \{y \in \mathbb{R}^p | y_{min} \leq y \leq y_{max}\} \tag{4.26c}$$

As Problem 4.25 in a continuous time representation requires the solution of a functional optimization problem, a discretization into a sampled-data structure is firstly carried out, allowing the conversion into a mathematical programming problem. The manipulated control inputs are parameterized as piecewise linear signals during the sampling intervals:

$$u(t) \equiv u(t_k) + \left(\frac{u(t_{k+1}) - u(t_k)}{T}\right)(t - t_k) \qquad \forall t \in [t_k, t_{k+1}),\ k \in \mathbb{N} \tag{4.27}$$

with $t_k := kT$ and T sampling period. At each sampling instance t_k, the solution of 4.25 yields the optimal discretized control sequence:

$$\bar{\mathbf{u}}^*(t_k) = \big[\underbrace{\bar{u}^*_{t_k|1}, \bar{u}^*_{t_k|2}, ..., \bar{u}^*_{t_k|N_C}}_{N_C}, \underbrace{\bar{u}^*_{t_k|N_C+1}, ..., \bar{u}^*_{t_k|N_P}}_{N_P - N_C}\big] \tag{4.28}$$

where N_C and N_P are the lengths of the control and prediction horizon such as that $T_C \leq T_P$, and $T_C = N_C T$, $T_P = N_P T$, while $\bar{u}_{t_k|l}$, with $l = \{1, ..., N_P\}$ represent the values of the control input vector at the l^{th} predicted step from time t_k. The first element of the optimal discrete control sequence $\bar{\mathbf{u}}^*(t_k)$ is applied to the system at each sample interval T:

$$u(t) := u_{t_k} + \left(\frac{\bar{u}^*_{t_k|1} - u_{t_k}}{T}\right)(t - t_k) \qquad t \in [t_k, t_{k+1}) \tag{4.29}$$

To determine the parameters necessary to define the optimal control strategy $\bar{\mathbf{u}}^*(t_k)$, the PSO algorithm depicted in Section 4.1 is exploited. To adapt the PSO method to the solution of the n-MPC Problem 4.25, each particle of the algorithm must describe a generic trajectory $\bar{\mathbf{u}}(t_k)$ of the control inputs during the prediction horizon. This is achieved by considering the first N_C terms of Equation 4.28 as the optimization variables (i.e. particle coordinates); consequently, the search space of each particle has dimensions $N_C \times m$, with m the number of control inputs.

Given a swarm of s particles, let the vector $\Delta\bar{\mathbf{u}}^i_j := [\Delta\bar{u}^i_{j,1}, ..., \Delta\bar{u}^i_{j,l}, ..., \Delta\bar{u}^i_{j,N_C}]$ and $\mathbf{v}^i_j := [v^i_{j,1}, ..., v^i_{j,l}, ..., v^i_{j,N_C}]$ denote the position and velocity of the i-th particle at the j-th iteration of the algorithm, with $i \in S = 1, 2, ..., s,\ j \in \mathbb{N},\ l \in \{1, 2, ..., N_C\}$. Then, the corresponding input trajectory over the prediction horizon is $\bar{\mathbf{u}}^i_j(t_k) = [\bar{u}^i_{j,t_k|1}, ..., \bar{u}^i_{j,t_k|l}, ..., \bar{u}^i_{j,t_k|N_P}]$, with $\bar{u}^i_{j,t_k|,l}$ equal to:

$$\bar{u}^i_{j,t_k|l} = \begin{cases} u(t_k) + \sum_{z=1}^{l} \Delta\bar{u}^i_{j,z} & \text{if } l \leq N_C \\ u(t_k) + \sum_{z=1}^{N_C} \Delta\bar{u}^i_{j,z} & \text{if } N_C < l \leq N_P \end{cases} \tag{4.30}$$

The Δ formulation defined above allows to easily account for input rate of charge constraints of the form 4.26b, by simply restricting the search space of the algorithm. The PSO particles coordinates represent therefore the N_C increments (or decrements)

of the control inputs at each of the sample times of the control horizon. Also, as the control horizon is usually shorter than the prediction one, the trajectory is assumed to stay constant during the last part of the prediction horizon (see Equation 4.25f).

At sample time t_k, after system state $x(t_k)$, disturbances $w(t_k)$, and inputs $u(t_k)$ are measured, it is possible to predict the evolution of the system, over the prediction horizon T_P, resulting from the application of a generic control sequence $\bar{u}(t) := \bar{\mathbf{u}}_j^i(t_k)$, by running the non-linear plant model (4.25c, 4.25d). To do so, the stochastic terms $w(t)$, representing the action of the external disturbances, are assumed to stay constant during the prediction horizon, i.e. $\bar{w}(t) \equiv w(t_k),\ \forall t \in [t_k, t_{k+1})$. This allows to associate to each particle a cost function $J_j^i(x(t_k), w(t_k), u(t_k), \Delta\bar{\mathbf{u}}_j^i)$.

Given the notation above for the swarm particles and the control trajectories, the following algorithm for the implementation of the PSO method for the n-MPC problem is defined:

Algorithm 1 PSO based n-MPC

Given the state $x(t_k)$ and the external disturbances $w(t_k)$ at the sampling time t_k:

1: Initialize model states $\bar{x}(t_k) := x(t_k)$, set the model disturbances as constant $\bar{w}(\tau) := w(t_k),\ \forall \tau \in [t_k, t_k + T_P]$, and randomly initialize positions and velocities for all s particles at iteration 0:

$$\Delta\bar{\mathbf{u}}_0^i = [\bar{u}_{0,1}, \bar{u}_{0,2}, ..., \bar{u}_{0,N_C}^i]$$
$$\mathbf{v}_0^i = \mathbf{0}, \qquad \forall i \in S$$

2: Evaluate cost function $J_0^i(x(t_k), w(t_k), u(t_k), \Delta\bar{\mathbf{u}}_0^i)$ for each particle $i \in S$ and assign personal best position and cost:

$$p_{best}^i = \Delta\bar{\mathbf{u}}_0^i, \qquad V_{p_{best}}^i = J_0^i(x(t_k), w(t_k), u(t_k), \Delta\bar{\mathbf{u}}_0^i), \qquad \forall i \in S$$

Find also the global best feasible particle i_g with the minimum associated cost function:

$$i_g := \arg\min_i \{J_0^i(x(t_k), w(t_k), u(t_k), \Delta\bar{\mathbf{u}}_0^i)\}, \quad i \in S$$
$$s.t. \quad \bar{\mathbf{u}}_0^i \quad \text{is feasible}$$

and determine the global best position and cost:

$$g_{best} = \Delta\bar{\mathbf{u}}_{t0}^{i_g}, \qquad V_{g_{best}} = J_0^{i_g}(x(t_k), w(t_k), u(t_k), \Delta\bar{\mathbf{u}}_0^{i_g})$$

3: Update velocity and position for each particle following PSO update Equations 4.2:

$$\mathbf{v}_{j+1}^{i} = \omega \mathbf{v}_{j}^{i} + r_1 c_1 (p_{best}^{i} - \Delta\bar{\mathbf{u}}_{j}^{i}) + r_2 c_2 (g_{best} - \Delta\bar{\mathbf{u}}_{j}^{i})$$
$$\Delta\bar{\mathbf{u}}_{j+1}^{i} = \Delta\bar{\mathbf{u}}_{j}^{i} + \mathbf{v}_{j+1}^{i} \qquad \forall i \in S$$

4: Evaluate the cost function for the next generation $J_{j+1}^{i}(x(t_k), w(t_k), u(t_k), \Delta\bar{\mathbf{u}}_{j+1}^{i})$ and perform the following conditional routine for updating each particle personal best:

if $J_{j+1}^{i}(x(t_k), w(t_k), u(t_k), \Delta\bar{\mathbf{u}}_{j+1}^{i}) < V_{p,best}^{i}$ **then**
$p_{best}^{i} = \Delta\bar{\mathbf{u}}_{tj+1}^{i}$
$V_{p_{best}}^{i} = J_{j+1}^{i}(x(t_k), w(t_k), u(t_k), \Delta\bar{\mathbf{u}}_{j+1}^{i}), \qquad \forall i \in M$
else
continue;
end if

Update the global best feasible particle i_g and perform the following routine:

if $J_{j+1}^{i_g}(x(t_k), w(t_k), u(t_k), \Delta\bar{\mathbf{u}}_{j+1}^{i_g}) < V_{g_{best}}$ **then**
$g_{best} = \Delta\bar{\mathbf{u}}_{tj+1}^{i_g}$
$V_{g_{best}} = J_{j+1}^{i_g}(x(t_k), w(t_k), u(t_k), \Delta\bar{\mathbf{u}}_{j+1}^{i_g})$
else
continue;
end if

5: If terminal criteria $j = j_{max}$ is met, continue to next step; otherwise $j = j + 1$ and move back to step 3

6: Stop and return the optimal control sequence $\Delta\bar{\mathbf{u}}^{*} = g_{best}$

The proposed algorithm can easily account for the input rate of change constraints (Equation 4.26b) by correctly defining the search space of each particle and the "absorbing wall" procedure is applied to particles straining out of the feasible search space as the algorithm progresses (see Section 4.1.2). Constraints on input saturation (Equation 4.26a) are easily addressed by pre-screening the control trajectories obtained from Equation 4.30 once the corresponding particle position is known. Finally, constraints on system outputs (Equation 4.26c) may be integrated into the objective function as soft constraints (with the introduction of a penalty factor) or kept as hard constraints.

Bibliography

[1] V. Adetola and M. Guay, "Integration of real-time optimization and model predictive control," *Journal of Process Control*, vol. 20, no. 2, pp. 125–133, Feb. 2010. 63

[2] R. Amrit, "Optimizing process economics in model predictive control," Ph.D. dissertation, University of Madison, Wisconsin, 2011. 64

[3] D. Angeli, "Economic Model Predictive Control," in *Encyclopedia of Systems and Control*. London: Springer London, 2013, pp. 1–9. 63

[4] D. Angeli, R. Amrit, and J. B. Rawlings, "On Average Performance and Stability of Economic Model Predictive Control," *IEEE Transactions on Automatic Control*, vol. 57, no. 7, pp. 1615–1626, Jul. 2012. 64

[5] P. Angeline, "Using selection to improve particle swarm optimization," in *1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence*. IEEE, 1998, pp. 84–89. 45

[6] M. Back, S. Terwen, and V. Krebs, "Predictive powertrain control for hybrid electrical vehicles," in *IFAC Symposium on Advances in Automotive Control*, Salerno, Italy, 2004, pp. 451–457. 50

[7] R. Bellman and E. Lee, "History and development of dynamic programming," *Control Systems Magazine, IEEE*, vol. 4, no. 4, pp. 24–28, 1984. 49

[8] R. E. Bellman, *Dynamic Programming*. Princeton N.J.: Princeton University Press, 1957. 52

[9] A. Bemporad and M. Morari, "Robust model predictive control: A survey," *Robustness in identification and control*, vol. 245, pp. 207–226, 1999. 62

[10] F. V. D. Bergh, "An analysis of particle swarm optimizers," Ph.D. dissertation, University of Pretoria, 2006. 44

[11] D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-Dynamic Programming*. Belmont, MA: Athena Scientific, 1996. 54

[12] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 3rd ed. Athena Scientific, Belmont, Massachusetts, 2005, vol. I. 33, 34, 36, 51

[13] J. T. Betts, *Practical Methods for Optimal Control Using Nonlinear Programming*. Philadelphia: Society for Industrial and Applied Mathematics, 2001. 36, 37, 62, 64

[14] F. Borrelli, A. Bemporad, and M. Morari, *Predictive Control for Linear and Hybrid Systems*. Cambridge University Press, 2014. ix, 36, 60, 61

[15] E. F. Camacho and C. Bordons, *Model Predictive control*, ser. Advanced Textbooks in Control and Signal Processing. London: Springer London, 2007. 62

[16] S. Y. Chen, Y. H. Hung, C. H. Wu, and S. T. Huang, "Optimal energy management of a hybrid electric powertrain system using improved particle swarm optimization," *Applied Energy*, vol. 160, pp. 132–145, Dec. 2015. 42

[17] M. Clerc, "The swarm and the queen: towards a deterministic and adaptive particle swarm optimization," in *Proceedings of the 1999 Congress on Evolutionary Computation-CEC99*. IEEE, 1999, pp. 1951–1957. 45

[18] A. David and J. B. Rawlings, "Receding Horizon Cost Optimization and Control for Nonlinear Plants," in *8th IFAC Symposium on Nonlinear Control Systems*, 2010, pp. 1217–1223. 64

[19] P. Elbert, S. Ebbesen, and L. Guzzella, "Implementation of Dynamic Programming for n-Dimensional Optimal Control Problems With Final State Constraints," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 3, pp. 924–931, 2012. ix, 49, 57, 58

[20] S. Engell, "Feedback control for optimal process operation," *Journal of Process Control*, vol. 17, no. 3, pp. 203–219, Mar. 2007. 63

[21] A. L. Facci, L. Andreassi, and S. Ubertini, "Optimization of CHCP (combined heat power and cooling) systems operation strategy using dynamic programming," *Energy*, vol. 66, pp. 387–400, 2014. 50, 75, 76

[22] W. Findeisen, F. N. Bailey, M. Bryds, K. Malinowski, P. Tatjewski, and A. Wozniak, *Control and Coordination in Hierarchical Systems*. New York, US: John Wiley & Sons, Inc., 1980. 63

[23] J. Forbes and T. Marlin, "Design cost: a systematic approach to technology selection for model-based real-time optimization systems," *Computers & Chemical Engineering*, vol. 20, no. 6-7, pp. 717–734, Jun. 1996. 63

[24] S. Helwig, "Particle Swarms for Constrained Optimization Partikelschw arme f ur Optimierungsprobleme mit Nebenbedingungen," Ph.D. dissertation, Technischen Fakult at der Universit at Erlangen-Nurnberg, 2010. ix, 46, 47

[25] J. V. Kadam and W. Marquardt, "Integration of Economical Optimization and Control for Intentionally Transient Process Operation," in *Assessment and Future Directions of Nonlinear Model Predictive Control*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 419–434. 64

[26] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Neural Networks, 1995. Proceedings., IEEE International Conference on*, vol. 4, pp. 1942–1948 vol.4, 1995. 31, 41

[27] J. Kennedy and R. C. Eberhart, *Swarm intelligence*. Morgan Kaufmann, 2001, vol. 5, no. 11. 44

[28] R. Luus, "On the application of iterative dynamic programming to singular optimal control problems," *IEEE Transactions on Automatic Control*, vol. 37, no. 11, pp. 1802–1806, 1992. 54

[29] J. Maciejowski, *Predictive control with constraints*. London, UK: Prentice Hall, 2002. 62

[30] E. Mezura-Montes and C. a. C. Coello, "Constraint-handling in nature-inspired numerical optimization: Past, present and future." *Swarm and Evolutionary Computation*, vol. 1, no. 4, pp. 173–194, 2011. 48, 49

[31] M. Morari and J. H. Lee, "Model predictive control: past, present and future," *Computers & Chemical Engineering*, vol. 23, no. 4-5, pp. 667–682, 1999. 59

[32] A. Navaeefard, S. M. M. Tafreshi, M. Barzegari, and A. J. Shahrood, "Optimal sizing of distributed energy resources in microgrid considering wind energy uncertainty with respect to reliability," in *2010 IEEE International Energy Conference*. IEEE, Dec. 2010, pp. 820–825. 42

[33] M. Neuman, H. Sandberg, and B. Wahlberg, "Rule-Based Control of Series HEVs Derived from Deterministic Dynamic Programming," *American Control Conference*, 2009. 50

[34] J. Pu and C. Yin, "Optimal control of fuel economy in parallel hybrid electric vehicles," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 221, no. 9, pp. 1097–1106, Jan. 2007. 50

[35] J. B. Rawlings and R. Amrit, "Optimizing Process Economic Performance Using Model Predictive Control," in *Nonlinear Model Predictive Control*. Springer Berlin Heidelberg, 2009, pp. 119–138. 64

[36] J. B. Rawlings and D. Q. Mayne, *Model Predictive Control: Theory and Design*. Nob Hill Publishing, LLC, 2012. 36, 62, 63, 174

[37] C. W. Reynolds, "Flocks, herds and schools: A distributed behavioral model," in *Proceedings of the 14th annual conference on Computer graphics and interactive techniques - SIGGRAPH '87*. New York, New York, USA: ACM Press, 1987, pp. 25–34. 41

[38] J. Robinson and Y. Rahmat-Samii, "Particle swarm optimization in electromagnetics," *Antennas and Propagation, IEEE Transactions on*, vol. 52, no. 2, pp. 397–407, 2004. 47

[39] J. Salerno, "Using the particle swarm optimization technique to train a recurrent neural model," in *Proceedings Ninth IEEE International Conference on Tools with Artificial Intelligence*. IEEE Comput. Soc, 1997, pp. 45–49. 42

[40] K. Sedlaczek and P. Eberhard, "Constrained Particle Swarm Optimization of Mechanical Systems," *Structural and Multidisciplinary Optimization*, vol. 32, no. 4, pp. 277–286, 2006. 49

[41] Y. Shi and R. Eberhart, "Empirical study of particle swarm optimization," in *Proceedings of the 1999 Congress on Evolutionary Computation*, 1999, pp. 1945–1950. 45

[42] Y. Shi and R. C. Eberhart, "Parameter selection in particle swarm optimization," in *Evolutionary Programming VII*. Springer Berlin Heidelberg, 1998, pp. 591–600. 45

[43] A. Stoppato, A. Benato, N. Destro, and A. Mirandola, "A model for the optimal design and management of a cogeneration system with energy storage," *Energy and Buildings*, Sep. 2015. 42

[44] P. Suganthan, "Particle swarm optimiser with neighbourhood operator," in *Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406)*. IEEE, 1999, pp. 1958–1962. 45

[45] O. Sundström, D. Ambühl, and L. Guzzella, "On Implementation of Dynamic Programming for Optimal Control Problems with Final State Constraints," *Oil & Gas Science and Technology Revue de lInstitut Français du Pétrole*, vol. 65, no. 1, pp. 91–102, Jan. 2010. ix, 54, 55, 56

[46] O. Sundström and L. Guzzella, "A generic dynamic programming Matlab function," in *18th IEEE Conference on Control Applications*, no. 7, 2009, pp. 1625–1630. 52, 53

[47] G. N. Vanderplaats, *Numerical Optimization Techniques for Engineering De sign.* Colorado Springs: Vanderplaats Research & Development, 2001. 49

[48] F. Y. Wang, H. Zhang, and D. Liu, "Adaptive Dynamic Programming: An Introduction," *IEEE Computational Intelligence Magazine*, vol. 4, no. May, pp. 39–47, 2009. 54

[49] L. Wang, *Model Predictive Control System Design and Implementation Using MATLAB*, ser. Advances in Industrial Control. London: Springer London, 2009. 60

[50] C. Wegley, M. Eusuff, and K. Lansey, "Determining Pump Operations using Particle Swarm Optimization," in *Building Partnerships.* Reston, VA: American Society of Civil Engineers, Sep. 2000, pp. 1–6. 42

[51] Xiaohui Hu and R. Eberhart, "Multiobjective optimization using dynamic neighborhood particle swarm optimization," in *Proceedings of the 2002 Congress on Evolutionary Computation*, vol. 2. IEEE, 2002, pp. 1677–1681. 42

[52] Y. Xu, D. Liu, and Q. Wei, "Action dependent heuristic dynamic programming based residential energy scheduling with home energy inter-exchange," *Energy Conversion and Management*, vol. 103, pp. 553–561, 2015. 50

[53] H. Yoshida, K. Kawata, Y. Fukuyama, S. Takayama, and Y. Nakanishi, "A particle swarm optimization for reactive power and voltage control considering voltage security assessment," *IEEE Transactions on Power Systems*, vol. 15, no. 4, pp. 1232–1239, 2001. 42

[54] A. Zanin, M. Tvrzská de Gouvêa, and D. Odloak, "Integrating real-time optimization into the model predictive controller of the FCC system," *Control Engineering Practice*, vol. 10, no. 8, pp. 819–831, Aug. 2002. 64

CHAPTER 5
OPTIMIZATION OF THE LOAD ALLOCATION IN A MULTI-SOURCE ENERGY APPLICATION FOR THE BUILDING SECTOR

The buildings sector accounts for approximately 40% of the entire global energy consumption and its energy demand is predicted to continuously grow in the forthcoming years. Following the pledge of limiting carbon dioxide emissions and reducing the energy consumption in this sector, different approaches have been separately followed, from the introduction of subsidiary renewable sources to a decrease of the demand resulting from buildings renovation. In this chapter, the potential of a multi-source solution for thermal and electric power generation in a building application is investigated, focusing on the issues regarding a correct integration of the different technologies and the development of optimal management procedures able to maximize the efficiency of the resulting complex energy system. A methodology for the evaluation of the benchmark efficiency of the plant is derived, based on the application of the Dynamic Programming method. The results are compared to those attained by the usage of a heuristic rule-based strategy, and exploited to identify the critical issues to address when designing a sub-optimal on line controller. Some final conclusions are drawn regarding the necessity of accounting from plant benchmark potentiality even during the design phase, and the possibility of extending the DP optimization procedure to define a model based causal controller for on-line implementation.

5.1 Introduction

Renewable energy technologies have become available to fulfill the power demand from residential and commercial buildings; solar thermal collectors are able to provide a clean alternative source for the generation of thermal power [11], while photovoltaic panels [17, 27] and wind turbines for mass [20] or micro [9] production can be exploited to meet the electric demand. While this technologies provide clean and cheap

sources, they are also heavily subject to stochastic external conditions which prevent an easy integration in the traditional power generation system. Decentralized storage systems have been developed to cope with the renewable sources intrinsic intermittent behaviour, thus increasing the systems flexibility, such as Compressed Air or Liquid Air Energy Storage (CAES, LAES) [12], hydroelectric energy storage (PHES) [16], battery and hydrogen based energy storages [10] for electric energy, and various active and passive thermal storage systems (see [23, 24] for a review). Also, with the introduction of the Smart Grid paradigm, the passive individual role of autonomous consumer/producers is surpassed with the development of an active participation strategy for the correct shaping of the comprehensive grid behavior [22].

Though remarkable efforts have been taken on the side of renewable energies, fossil fuel based technologies are supposed to remain the primary source for thermal and power generation, thanks to their flexibility and robustness. For this reason, the development of increased efficiency fuel-based technologies will still play a central role in the reduction of GHG emissions. On the side of thermal energy generation, even if the current condensing boilers have reached extremely high thermal conversion efficiencies [15], the traditional direct conversion of fuel energy into thermal power proved to be extremely ineffective from an energy conversion efficiency point og view, due to the high associated exergy destruction. Electric driven or absorbtion based heat pumps provide a much more efficient alternative for the thermal energy production, although their performance and the effective GHG potential reduction are highly influenced by the atmospheric conditions and the composition of the national prime energy mix [5]. Combined heat and power systems (CHP) represent another rapidly spreading technology which relies on the simultaneous production of electricity and thermal power. CHP power systems play a central role in distributed and micro generation [3, 7] as they yield higher energy conversion efficiencies (up to over 80%) if compared compared to the average 30-35% efficiency from conventional electricity generation systems [26], thus entailing a reduction in both the operating costs and the greenhouse gases emissions by providing a higher fuel utilization factor. Finally, hybrid multi-source systems have been developed to fulfill the energy demand by means of a combination of two or more of the previously mentioned technologies. Nosrat and Pearce for example [25] investigated the performance of a trigeneration hybrid photovoltaic/CHP system coupled with an absorption chiller for the fulfillment of the electric, domestic hot water, space heating, and space cooling demand in a residential application. The authors developed a model based schedule strategy for the management of the plant and simulated the application of the system for a

typical home in Vancouver, obtaining performance improvements of over 50 %. Xi et al. [28] conducted experimental studies on a solar-assisted ground source heat pump system for space heating and showed the effectiveness of the solution.

In such a complex scenario, involving a plurality of different energy sources and technologies and the possibility of shaping the actual production demand ratio by means of accumulation devices, it is necessary to define adequate methods to correctly configure and manage the hybrid systems, in order achieve the optimal exploitation of the different sources available. When dealing with a complex multi-energy system in fact, the daily operational management problem, based on the evaluation of the optimal load allocation strategy, requires a deep understanding of the comprehensive system behaviour and the iterations between the different components. As the same energy demand can be satisfied with a large combinations of the different sources, each one leading to different overall performances, the control algorithm must be able to exploit the knowledge of the system behaviour and dynamics to efficiently evaluate which one of the numerous available combination is to be implemented to maximize the efficiency of the system. Many solution for the resulting dynamic optimization problem are found in literature. based on either linear approximations of the plant processes [13, 21] or the usage of non-linear [6, 14] and mixed integer non-linear [19] programming tools.

In the present chapter, the problem of defining the correct scheduling policy for a multi-source solution for the fulfillment of the electric and thermal demand of a building is analyzed. A deterministic Dynamic Programming (DP) based control algorithm is used to tackle the corresponding optimal control problem. The DP algorithm is becoming an increasingly employed tool for the evaluation of optimal control strategies in hybrid power plant solutions; Marano et el. [18], for example, exploited the algorithm to derive the optimal management for a hybrid photovoltaic panels and wind turbines plant, able to store any excess of produced energy into a compressed air energy storage (CAES); furthermore, the authors showed that the resulting non-trivial control policy could lead to substantial reduction in operating and maintenance costs with respect to the conventional scenario. Bianchi et al. [4] compared the performance of a rule-based scheduling policy to those of a DP optimal strategy based on the management of a hybrid plant for the production of electric power from a wind farm coupled with a compound pump hydro storage (PHS) and gas turbines. Facci et al. [7] proposed a DP based methodology for determining the optimized management strategy of a grid connected, ICE based, trigeneration plant for the production of combined heat power and cooling, reporting a significant

economical advantage (from 8% up to more than 100%) deriving from the adoption of the DP optimal solution when compared to the results of a heuristic strategy.

The deterministic DP algorithm used in the present analysis does not yield a direct solution for the implementation of a causal control strategy, as it requires the perfect knowledge of future external conditions and energy demand profiles. The methodology is on the other hand adopted due to the intrinsic advantage of guaranteeing a certified global optimal solution to the dynamic optimization problem. This property can be exploited to derive a benchmark for the system performance against which it is possible to eventually compare other casual layouts. Moreover, if an efficient prediction model for the future conditions is developed, the procedure outlined in this chapter can be easily exploited to derive a causal plant controller.

5.2 Multi source energy plant for building applications

The investigation carried out in this chapter is based on the study of a multi-source energy plant, described in detail in [1], which is composed of a large number of different integrated subsystems, each contributing to the fulfilment of the thermal or electric demand of a commercial building. The nine different technologies adopted are:

1. Solar thermal collectors for heat generation (STH);
2. Solar photovoltaic panels for electricity generation (PV);
3. A combined heat and power micro gas turbine for the cogeneration of thermal and electrical power (CHP);
4. A reversible air source heat pump (ASHP) for both heat and cooling production;
5. A reversible ground source heat pump (GSHP) for both heat and cooling production;
6. An absorption chiller (ABS) for the cooling production;
7. An auxiliary condensing boiler (AB) for the integration of the heat production;
8. An auxiliary electric chiller for the integration of the cooling production (AC);
9. A thermal energy storage for hot water (STORAGE).

The multi-source plant is used to meet the commercial energy demand of a thirteen floors tower composed of a basement (used as warehouse and garage), a commercial area located at ground and first floor, and different locals for office use in the remaining 12 floors (see [2] for additional details). As the present analysis will only focus on the winter and mid-season period, the machines used to meet the cooling demand, i.e. the absorption chiller (ABS) and the electric chiller (AC), are neglected. Figure 5.1 depicts the schematic of the comprehensive plant, showing how the various technologies are integrated and highlighting the difference sources (both renewable and traditional) for the prime energy supply.

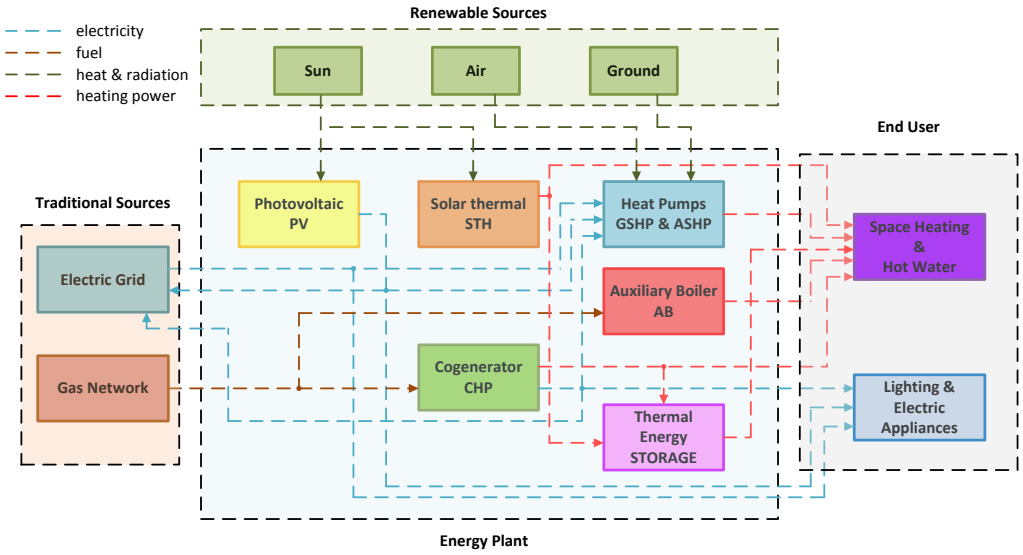


Figure 5.1: Schematic of the multi-source energy plant

The sizes of the different technologies are obtained based on the work of Barbieri et al. [1] who derived a methodology for minimizing the annual flux of primary energy. The procedure relies on a Genetic Algorithm based static optimizer for the evaluation of the optimal sizes of non conventional power sources, i.e. the nominal power, P_{nom}, of the cogenerator (CHP) and heat pumps (GSHP and ASHP), and the share of the total surface to be assigned respectively to the photovoltaic (PV) electricity production and the solar thermal heat (STH) production. Moreover, the nominal power of the auxiliary burner AB is chosen to guarantee that the boiler can potentially satisfy

the entire thermal demand by itself, while the size of the thermal storage is evaluated according to UNI 8477-2, which prescribes the allocation of 0.04 m^3 of tank volume for each m^2 of solar thermal surface, and 2 kWh of thermal capacity for each kW_{th} of thermal power produced by the cogenerator. Table 5.1 summarizes the sizes of the different technologies that are obtained by applying the aforementioned procedure to the present configuration of the building.

Table 5.1: Summary of size and parameters of the adopted technologies

Technology	**Parameter**	**Unit of Measurement**	**Value**
Photovoltaic PV	Surface	$[m^2]$	325
Solar Thermal Heating STH	Surface	$[m^2]$	3
Combined Heat and Power CHP	$P_{el,nom}$	$[kW_{el}]$	100
	$\eta_{el,nom}$	[-]	0.29
	$P_{th,nom}$	$[kW_{th}]$	195
	$\eta_{th,nom}$	[-]	0.57
Ground Source Heat Pump GSHP	$P_{th,nom}$	$[kW_{th}]$	40
	COP_{nom}	[-]	3.35
Ground Source Heat Pump ASHP	$P_{th,nom}$	$[kW_{th}]$	100
	COP_{nom}	[-]	2.80
Auxiliary Boiler AB	$P_{th,nom}$	$[kW_{th}]$	234
	$\eta_{th,nom}$	[-]	1.06
Thermal Storage STORAGE	E_{max}	$[kWh_{th}]$	393

Energy load profile and atmospheric conditions The building construction site is located in northern Italy, inside climatic zone E, where the winter and mid-season period runs from October 15th to April 15th. Thermal energy for heating and hot water generation and electric energy for lighting and appliances have been evaluated by Munari et al. [2] with the help of EdilClimaEC700®, a software for static energetic simulations. Figure 5.2 depicts the monthly energy demand (thermal

and electrical) for the winter and mid-season period; 212 MWh_{th} are cumulatively required for space heating and hot water production, while 212 MWh_{el} are required for the electric appliances.

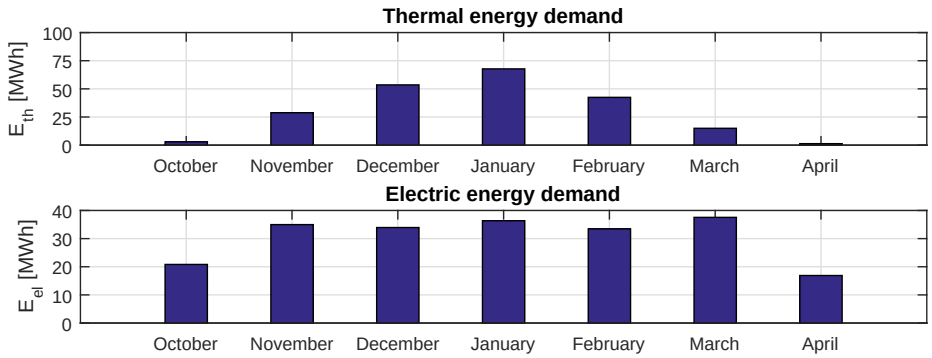


Figure 5.2: Monthly energy demand for the winter and mid-season period

Assuming that working days are from 8 a.m to 5 p.m., Monday to Saturday, and taking into account all the festivities, electric and thermal loads are distributed over the months with a hourly discretization by means of non-dimensional profiles which takes into account the type of the users [2]. The profiles for a mid-December working week are depicted in Figure 5.3. A hourly thermal power peak demand of 234 kW is registered in January, while the electric hourly power peak demand lies in December, with 96 kW.

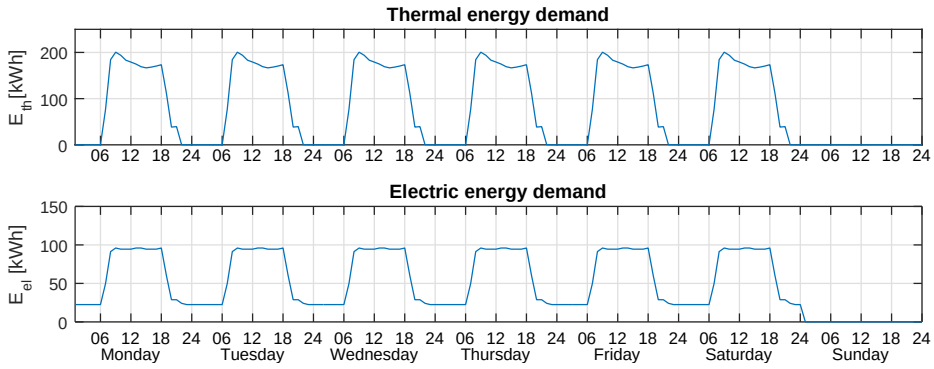


Figure 5.3: Weekly energy demand profile, mid-December

Still based on the building location, Munari et al. [2] derived the monthly average daily values for solar radiation and air temperature based on the normative UNI 10349. Figure 5.4 depicts the trend of these variables during the winter period.

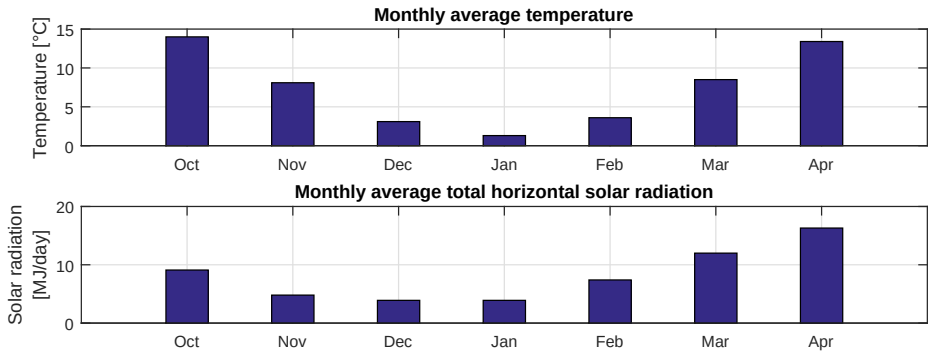


Figure 5.4: Monthly average daily values for solar radiation and air temperature for the winter period

As for thermal and electric loads, temperature and radiation values are distributed over the duration of the hours by means of proper non-dimensional profiles.

5.2.1 Models for plant components

The various subsystems which compose the multi source energy plant are simulated by means of a set of models developed by Barbieri et al. [1] and implemented in the MATLAB® environment. These models rely on a systemic approach for the simulation of the components and make use of a set of performance parameters related to external environmental condition and machines load, such as the electric and thermal efficiencies for the CHP, the coefficient of performance (COP) for the heat pumps, the radiant energy conversion efficiency for the solar panels, etc. A discrete time representation is adopted, with a sample time Δt of 1 *hour* defined accordingly to the discretization of energy loads and environmental conditions depicted in the previous section.

In the following paragraphs, each component sub-model is briefly described, focusing on the relationship between the produced energy E and the following parameters:

- *external disturbances* (w), representing the exogenous atmospheric parameters and the building thermal and electric loads;
- *control inputs* (u), representing the decision parameters from the control algorithm;
- *system states* (x), representing the set of time varying parameters needed to describe the dynamic evolution of the system.

Furthermore, the following notation is applied, with the subscript $_{th}$ used to refer to thermal power (energy), $_{el}$ denoting electric power (energy), and the index $k \in \mathbb{N}$ representing the generic time instant at which the model correlations are evaluated.

Photovolataic - PV The energy production from the photovoltaic system is not actively regulated by the control strategy, neither it is determined by any relevant dynamic behaviour ($u_{PV} = \emptyset,\ x_{PV} = \emptyset$). Its contribution is therefore uniquely determined by the external conditions, affecting the efficiency of the photovoltaic panel and the inverter. For the mono-crystalline silicon panel adopted in the investigated application, the panel efficiency is related to the difference between external temperature and the Nominal Operative Cell Temperature NOCT (see for example [8]).

The net electric energy production per hour can be therefore simply expressed as:

$$E_{el,PV}(k) = P_{el,PV}(w(k)) \cdot \Delta t \tag{5.1}$$

where the panel power $P_{el,PV}$ depends on the panel efficiency and the total amount of solar incident radiation on the photovoltaic surface. Electric energy produced by the photovoltaic system is either used to meet the building and heat pumps demand or fed back to the grid.

Solar Heating - STH Similarly to the photovoltaic panels situation, no device is introduced for an active regulation of the thermal energy produced by the solar heating system ($u_{STH} = \emptyset,\ x_{STH} = \emptyset$). The efficiency of the panel is related to air temperature and radiation level [2] and the net thermal energy production per hour may be expressed as:

$$E_{th,STH}(k) = P_{th,STH}(w(k)) \cdot \Delta t \tag{5.2}$$

The thermal energy produced by the solar heating system contributes to satisfy

the building thermal load. Any excess of produced energy is either stored to the thermal accumulator, if possible, or dissipated otherwise.

Cogenarator - CHP The Combined Heat and Power cogenerator (CHP) is composed of a micro gas turbine (MGT) that is fueled by the gas network. The machine produces electric energy (either used to meet the building and heat pumps demand or fed back to the grid) and uses the residual thermal energy to satisfy the building hot water and space heating demand. Any excess of produced thermal energy is stored (if possible) or wasted otherwise. It is possible to regulate the thermal energy production of the CHP by means of a control variable, u_{CHP}, which defines the fraction of the maximum thermal to effectively produce.

$$E_{th,CHP}(k) = u_{CHP}(k) \cdot P_{th,CHP,max}(w(k)) \cdot \Delta t \tag{5.3}$$

The minimum thermal load permitted for the CHP is assumed to be equal to 80% of the maximum; if $u_{CHP} \leq 0.8$ the machine is shut-off:

$$u_{CHP} \in \Big\{\{0\} \ \cup \ [0.8, 1]\Big\}$$

Furthermore, the maximum amount of thermal energy that can be produced is related to the external conditions $w(k)$.

Given the amount of thermal energy produced, $E_{th,CHP}$, it is possible to obtain the corresponding amount of electric energy, $E_{el,CHP}$, that is generated at the same time, by means of the two efficiencies, $\eta_{el,CHP}$ and $\eta_{th,CHP}$. The first one, $\eta_{el,CHP}$, indicates the thermodynamic efficiency for the production of electric energy from the fuel; the second one, $\eta_{th,CHP}$, indicates the efficiency of the conversion of the residual heat at the outlet of the MGT to thermal energy for the usage in the building. Both these efficiencies are related to the fraction of the maximum thermal energy produced, u_{CHP}, and the external conditions, $w(k)$. A more detailed description may be found in [1, 2]. With the assumptions above, the produced electric energy is:

$$E_{el,CHP}(k) = \frac{\eta_{el,CHP}(u_{CHP}(k), w(k))}{1 - \eta_{el,CHP}(u_{CHP}(k), w(k))} \cdot \frac{E_{th,CHP}(k)}{\eta_{th,CHP}(u_{CHP}(k), w(k))} \tag{5.4}$$

The amount of primary fuel energy, $E_{CHP,fuel}$, needed in order to generate the desired amount of thermal and electric energy is evaluated by means of the fuel-to-electricity efficiency, $\eta_{el,CHP}$, of the system and it also takes into account the additional fuel consumption that is required in the case that the machine has to

be turned on. The startup penalty term is assumed to correspond to the energy consumed by operating 10 minutes at nominal power; to detect the startup phase, a binary state variable $x_{CHP} \in \{0, 1\}$ is introduced, accounting for the ON-OFF condition of the system.

$$E_{CHP,fuel}(k) = \frac{E_{el,CHP}(k)}{\eta_{el,CHP}(u_{CHP}(k), w(k))} + \\ + E_{CHP,startup}(x_{CHP}(k), u_{CHP}(k), w(k)) \tag{5.5}$$

$$E_{CHP,startup}(k) = \begin{cases} \frac{P_{el,CHP,nom}(w(k))}{\eta_{el,CHP,nom}(w(k))} \cdot \frac{\Delta t}{6}, \text{ if } \begin{cases} x_{CHP}(k) = 0 \\ u_{CHP}(k) \neq 0 \end{cases} \\ 0, \text{ otherwise} \end{cases} \tag{5.6}$$

The state update of the system is defined by the following simple rule:

$$x_{CHP}(k+1) = \begin{cases} 1, & \text{if } u_{CHP}(k) \neq 0 \\ 0, & \text{if } u_{CHP}(k) = 0 \end{cases} \tag{5.7}$$

Ground Source and Air Source Heat Pumps - GSHP & ASHP The models for the two heat pumps are based on the same structure, and the notation *XSHP* is therefore adopted to indicate the generic component. The two heat pumps can contribute with a fraction u_{XSHP} of their maximal thermal power production to the fulfillment of the thermal demand of the building:

$$E_{th,XSHP}(k) = u_{XSHP}(k) \cdot P_{th,XSHP,max}(w(k)) \cdot \Delta t \tag{5.8}$$

The corresponding electric energy consumption is obtained by means of the evaluation of the Coefficient Of Performances (COP):

$$E_{el,XSHP}(k) = \frac{E_{th,CHP}(k)}{COP_{XSHP}(w(k))} \tag{5.9}$$

The nominal performance and the maximum thermal power that the machines can produce are affected by the temperature of the external heat exchanger and therefore related to the external conditions $w(k)$. Furthermore, accordingly to UNI/TS 11300, the COP is affected by load variations [2]. No dynamic behaviour is accounted in the heat pump model ($x_{XSHP} = \emptyset$) and any excess of thermal power produced is wasted. As with the CHP, even for the heat pumps there is a lower limit for the thermal

production under which the machines are turned off, in this case equal to 10% of the maximum machine power.

Auxiliary Boiler - AB An auxiliary boiler, powered by natural gas, is responsible for the production of the residual thermal energy, $E_{th,AB}$, eventually not fulfilled by the other systems. No control input is defined ($u_{AB} = \emptyset$) as the component load is determined by the overall energy balance of the building. Moreover, the dynamics of the system are assumed negligible ($x_{AB} = \emptyset$). The amount of fuel primary energy needed to generate the required amount of thermal energy is evaluated as follows:

$$E_{AB,fuel}(k) = \frac{E_{th,AB}(k)}{\eta_{th,AB}(E_{th,AB}(k))} \tag{5.10}$$

where the auxiliary boiler efficiency $\eta_{th,AB}$ is related uniquely to the load.

Thermal Energy Storage - STORAGE Both the STH system and the CHP cogenerator can send any excess of produced thermal energy to a hot water storage tank. The energy stored in the accumulator is an additional state of the system ($x_{STORAGE}(k) = E_{STORAGE}(k)$). With $E_{th,STH,unused}$ and $E_{th,CHP,unused}$ the amount of excess energy produced by the STH and the CHP, and $E_{th,STORAGE,max}$ the maximum amount of energy that can be stored into the tank (see Table 5.1), the amount of energy introduced into the accumulator at each hour is equal to:

$$\begin{aligned} E_{th,STORAGE,in}(k) = \min \big(& E_{th,STORAGE,max} - x_{STORAGE}(k) \ , \\ & E_{th,STH,unused}(k) + E_{th,CHP,unused}(k) \big) \end{aligned} \tag{5.11}$$

A fraction of the energy stored in the tank $\big(u_{STORAGE} \in [0,1]\big)$ can be used in the future to fulfill the thermal demand from the building:

$$E_{th,STORAGE,out}(k) = u_{STORAGE}(k) \cdot x_{STORAGE}(k) \tag{5.12}$$

From the heat storage energy balance equation, and assuming a constant dissipation proportional to the heat exchange area (and therefore to the amount of energy stored in the tank), the state update equation for the thermal storage can be formulated as follows:

$$x_{STORAGE}(k+1) = (1 - c_{diss}) \cdot (x_{STORAGE}(k) + E_{th,STORAGE,in}(k) + - E_{th,STORAGE,out}(k)) \tag{5.13}$$

where c_{diss} is a dissipation coefficient assumed equal to 0.5%.

5.2.2 Comprehensive plant model and energy balance

The combination of the thermal demand from the building and the production from the various technologies define the amount of energy that is required from the auxiliary burner. From energy conservation equations it follows that:

$$E_{th,AB}(k) = E_{th,demand}(k) - \Big(E_{th,STH}(k) + E_{th,CHP}(k) + E_{th,GSHP}(k) + + E_{th,ASHP}(k) + E_{th,STORAGE,out}(k) \Big) \tag{5.14}$$

If $E_{th,AB}(k)$ is lower than zero, there is an excess of energy production and the boiler is not activated. As mentioned before, this excess of thermal production is entirely wasted if it is coming from the heat pumps, while it can be stored in the thermal accumulator if it is coming from the STH or the CHP and the operation is feasible.

By applying energy conservation on the electric side, it is possible to derive the amount of electric energy requested from the grid as:

$$E_{el,request}(k) = E_{el,demand}(k) - E_{el,PV}(k) - E_{el,CHP}(k) + E_{el,GSHP}(k) + + E_{el,ASHP}(k) \tag{5.15}$$

If $E_{el,request}(k)$ is lower than zero, the electricity is sent to the grid and labeled as $E_{el,sent} = |E_{el,request}|$, otherwise is withdrawn and labeled as $E_{el,taken} = E_{el,request}$.

Finally, the total amount of fuel energy supplied by the gas network is simply:

$$E_{fuel}(k) = E_{CHP,fuel}(k) + E_{AB,fuel}(k) \tag{5.16}$$

The set of equation presented in the previous paragraph may be organized to

formulate the comprehensive energy production plant model into the discrete non-linear state space representation:

$$\begin{cases} x(k+1) = f\big(x(k), u(k), w(k)\big) \\ y(k) = g\big(x(k), u(k), w(k)\big) \end{cases} \quad \text{subject to: } u(k) \in U,\ x(k) \in X \tag{5.17}$$

with $x = [x_{CHP}, x_{STORAGE}]$, $u = [u_{CHP}, u_{GSHP}, u_{ASHP}, u_{STORAGE}]$ and $y = [E_{el,sent}$, $E_{el,taken}$,$E_{fuel}]$. The set of feasible inputs and states (U, X) are summarized in Table 5.2.

Table 5.2: Summary of multi-energy system plant constraints

Input constraints
$u_{CHP} \in \{0\} \cup [0.8, 1]$
$u_{GSHP} \in \{0\} \cup [0.1, 1]$
$u_{ASHP} \in \{0\} \cup [0.1, 1]$
$u_{STORAGE} \in [0, 1]$
State constraints
$x_{CHP} \in \{0, 1\}$
$x_{STORAGE} \in [0, E_{th,STORAGE,max}]$

5.3 Control strategy

The objective of the energy plant control strategy is that of minimizing the overall primary energy consumption (thus guaranteeing the maximum energetic efficiency) by choosing the right allocation of the thermal and electric demand between the different energy sources available. The problem can be formulated as an optimal control one if a suitable objective function is defined. The solution of such problem may be pursued with one of the methods outlined in Section 3.2. For the present application, given that an estimation of the external disturbances is available, the maximum attainable performances of the plant are evaluated by the application of the deterministic Dynamic Programming algorithm (see Section 4.2).

Moreover, a simple heuristic rule-based controller, developed by Barbieri et al. [1] and applied in previous studies on the plant [2], is briefly described. The performance of this simple strategy will be compared to those of the optimal DP algorithm to assess its efficiency.

5.3.1 Dynamic Programming optimal policy

Given a generic control policy $\pi = [u_{CHP}(1), u_{GSHP}(1), u_{ASHP}(1), u_{STORAGE}(1), ..., u_{CHP}(N-1), u_{GSHP}(N-1), u_{ASHP}(N-1), u_{STORAGE}(N-1)]$ defining the energy schedule for each of the N hours composing the winter and mid-season optimization horizon, it is possible to mathematically estimate the corresponding energetic efficiency by an objective function J defined as follows:

$$J(x_0, \pi, w(k)) = \sum_{k=0}^{N-1} \Bigg(f_{fuel} \cdot E_{fuel}(x(k), u(k), w(k)) + f_{el,taken} \cdot E_{el,taken}(x(k), u(k), w(k)) - f_{el,sent} \cdot E_{el,sent}(x(k), u(k), w(k)) \Bigg) \tag{5.18}$$

subject to:

$$x(k+1) = f\big(x(k), u(k), w(k)\big), \quad x(0) = x_0 \tag{5.19}$$

$$x(k) \in X \tag{5.20}$$

$$u(k) \in U \tag{5.21}$$

The objective function J represents the overall prime energy consumption of the plant, with the coefficients f indicating the energy conversion factors defined following the Italian regulation D.M. 4/08/2011 ($f_{fuel} = 1$, $f_{el,taken} = 2.53$, $f_{el,sent} = 2.35$).

If the external disturbances $w(k)$ are known in advance, the deterministic Dynamic Programming algorithm is able to provide the optimal control policy:

$$\pi^* = \arg\min_{\pi \in \Pi} J(x_0, \pi, w(k)) \tag{5.22}$$

where Π represent the set of admissible trajectories satisfying both the input and state constraints over the entire optimization horizon. Furthermore, the implementation of the algorithm requires a discretization of both the inputs and the states. An evenly spaced 10 points grid is used to represent the effect of the tank share $u_{STORAGE}$, while non-evenly spaced grids are adopted for the other inputs ($u_{CHP}, u_{GSHP}, u_{ASHP}$), with the first point of the grid being 0 and the remaining nine evenly distributed between

each machine minimum and maximum load. The state associated to the CHP is already discrete ($x_{CHP} \in \{0, 1\}$), while the state representing the energy stored in the tank ($x_{STORAGE}$) is described by means of an evenly spaced 10 elements grid.

To define the initial system condition x_0, it is assumed that the CHP is turned off and the tank is empty ($x_{CHP} = 0$, $x_{STORAGE} = 0$) at the start of the optimization horizon. Furthermore, no constraints are imposed on the final admissible states set.

5.3.2 Schedule based controller

The performances of the DP optimal control policy are compared to those of a heuristic rule based algorithm developed by Barbieri et al. [1] for the management of the plant and applied in previous studies on the system [2]. This algorithm defines a strategy for the fulfillment of the thermal load based on a list of switch-on priorities for the different available subsystems, which favors the adoption of renewable sources and the heat production from alternative mechanisms. The detailed switch-on priority list is reported below:

1. Thermal storage (STORAGE);
2. Renewable sources, photovoltaic and solar heating (PV & STH);
3. Combined heat and power (CHP);
4. Ground source heat pump (GSHP);
5. Air source heat pump (ASHP);
6. Auxiliary boiler (AB).

The heat storage is the first system which contributes to the fulfillment of the thermal demand at each hour and its thermal energy capacity is added to the amount of the thermal demand from the building, so that the other sources will eventually store the produced energy in the tank. The switch-on strategy for the other elements is depicted in further details in Figure 5.5; the *i-th* machine is activated if the residual energy demand is higher than the machine minimum energy production threshold. Moreover, if the energy demand is higher than the maximum production of the *i-th* machine, the subsystem will operate at maximum load and the thermal residual demand will be updated and allocated to the next machines in the list.

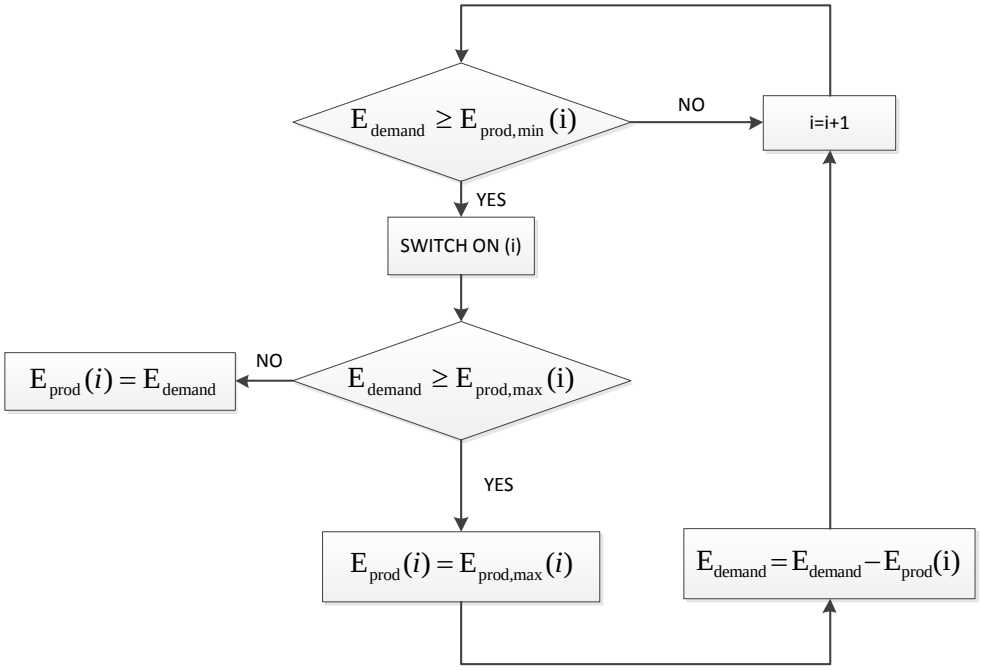


Figure 5.5: Flowchart for rule-based machine scheduling controller [1]

5.4 Policies comparison

The performance of the two policies described in the previous section are compared based on the prime energy fuel consumption over the entire winter mid-season horizon. As depicted in Figure 5.6, the optimal allocation strategy deriving from the Dynamic Programming algorithm is able to deliver a consistent increase in the system efficiency over the entire period, with a cumulative prime energy saving of 56.8 MWh, corresponding to a percentage reduction of 8.4 %.

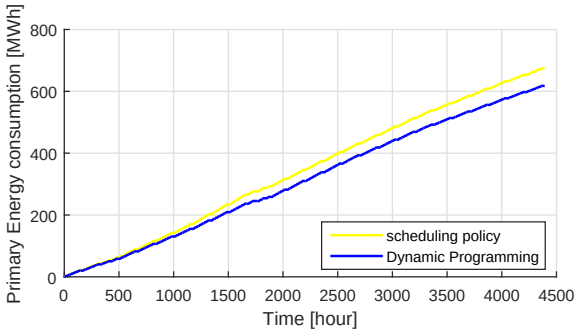


Figure 5.6: Primary energy overall consumption

More in-depth details may be gathered if referring to the single terms composing the overall prime energy consumption, as depicted in Figure 5.7. The adoption of the optimal DP policy involves a cumulative decrease of fuel energy consumption of 14.6 MWh, a reduction in the electric energy taken from the grid of 12.5 MWh_{el}, and an increase of 4.2 MWh_{el} in the amount of electric energy sent to the grid.

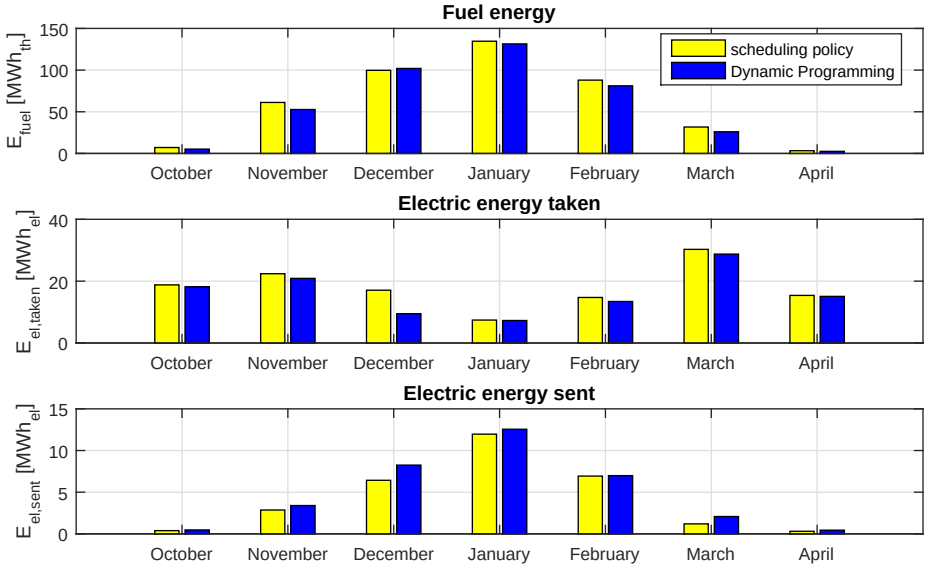
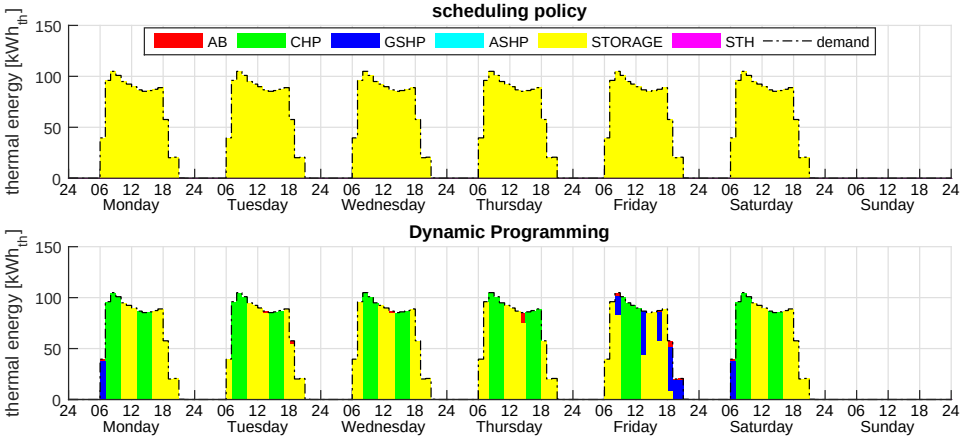


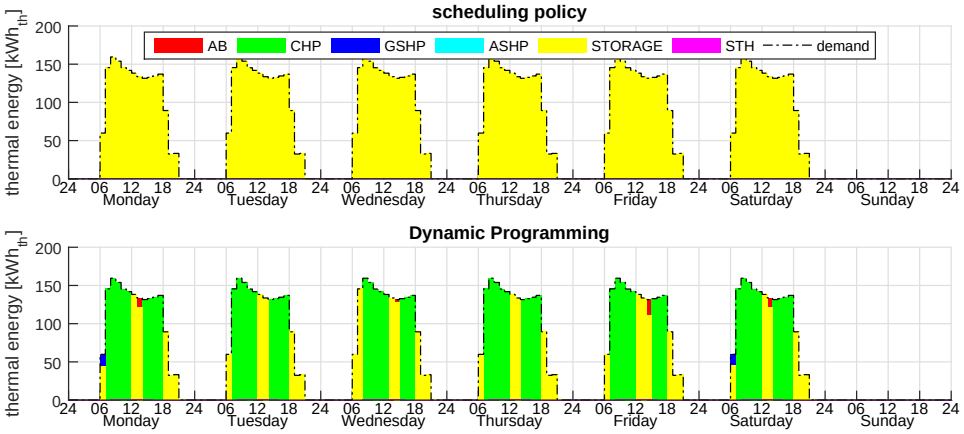
Figure 5.7: Month composition of primary energy consumption

Figure 5.8 depicts how the thermal demand is met by the two policies during two

standard weeks of the month of November and February. With the scheduling policy, the thermal storage is always preferred for the immediate satisfaction of the thermal demand; consequently the remaining sources are only used to refill the hot water tank. Moreover, due to the sizing of the various components and the schedule order, the STH and the CHP are always entirely responsible for the tank refilling, with neither the ASHP nor the GSHP being ever used during the winter and mid-season period.



(a) mid-November



(b) mid-February

Figure 5.8: CHP weekly load allocation example

With the DP policy on the other hand, the meeting of the thermal demand is usually attained by an alternative usage of the stored energy and the CHP. Moreover,

the GSHP and the Auxiliary Boiler are sometimes used to cover a residual demand. In this case the preference is given to use of the GSHP if the required amount of energy is greater than the minimum threshold from the machine.

As the CHP is the major source for the production of thermal energy in both the situations, additional investigations on the particular sub-system behaviour have been conducted. A very beneficial effect of the Dynamic Programming algorithm is noticeable when analysing the CHP net month thermal production and electric energy conversion efficiency, as depicted in Figure 5.9. While in fact the net thermal energy production from the CHP is nearly equal in the two situations, the different scheduling strategy provided by the Dynamic Programming algorithm allows the machine to operate at constant nominal load, thus maximizing the corresponding amount of electric energy produced from the combustion of the fuel.

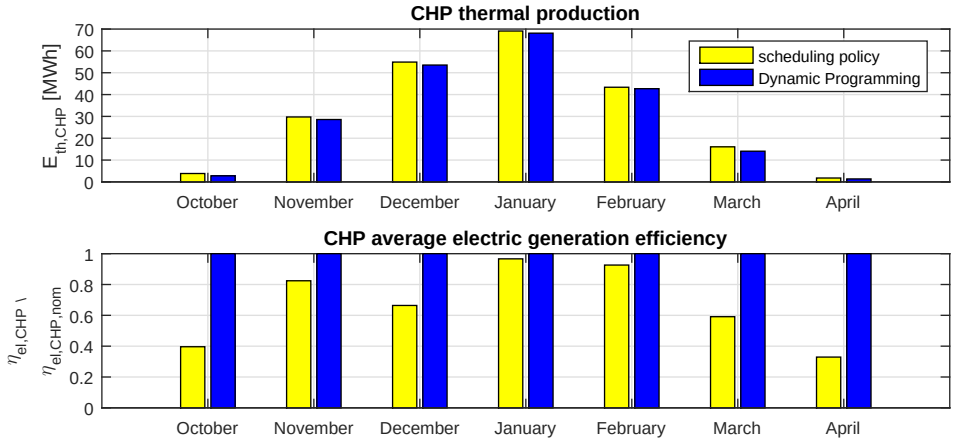
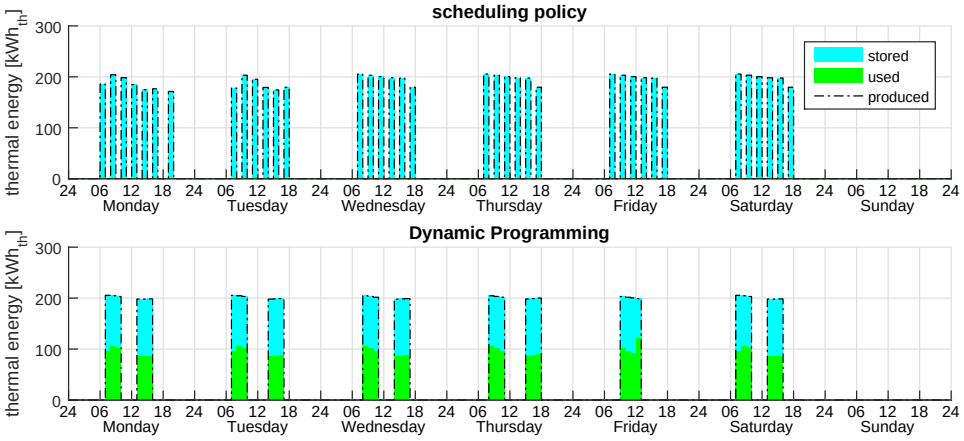


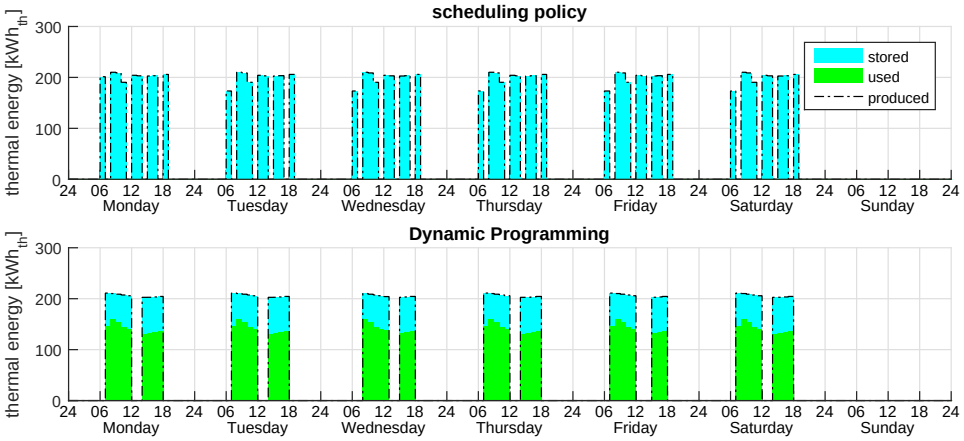
Figure 5.9: CHP monthly performance

The ability of the Dynamic Programming control policy to maintain the CHP at nominal load conditions may be even more deeply understood if referring to Figure 5.10, where the weekly CHP scheduling is depicted for two typical operating weeks from the months of November and February. In the Dynamic Programming case, the CHP covers the basic thermal while operating at nominal conditions, and sends any excess of produced energy to the accumulator. Once the accumulator is nearly full, the CHP is shut down and kept off as long as it is possible, while any future thermal

demand is covered with the help of the stored energy and some little contribution from the GSHP or the AB.



(a) mid-November



(b) mid-February

Figure 5.10: CHP weekly load allocation example

Finally, Figure 5.11 depicts the monthly energetic dissipations due to heat losses from the accumulator and the additional fuel consumption associated to the CHP startup phase. The DP optimal policy is able to reduce the heat losses by 4.2 MWh and, thanks to the alternation in the usage of the CHP and the accumulator, the number of CHP startup phases is reduced, entailing a reduction of 16.8 MWh in the fuel energy consumption.

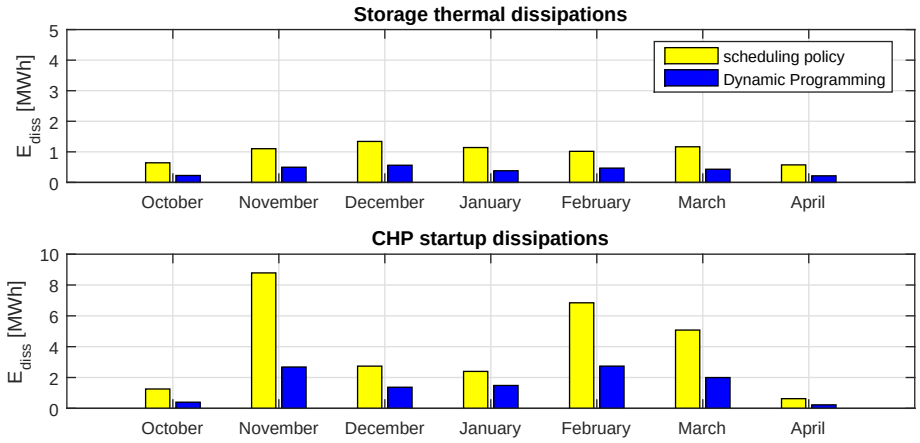


Figure 5.11: Monthly energy dissipations

Accordingly to the analysis conducted above, the beneficial effects of the optimal DP control policy and its peculiarities can be summarized as follows:

- The Combined Heat and Power system is always chosen as the main source of thermal power;
- The thermal storage is used to keep the CHP at nominal operating conditions by absorbing any excess of produced heat. This allows the maximization of the electric efficiency at the cogenerator;
- An optimal use of the thermal storage also leads to a reduction in the number of CHP shut off and startup phases, thus avoiding any unnecessary energy dissipation;
- The Ground Source Heat Pump and the Auxiliary Boiler are eventually used to cover any residual thermal demands that cannot be otherwise met by the use of stored energy or with the CHP working at nominal conditions;
- The Air Source Heat Pump is always kept off.

5.5 Conclusions

The evaluation of the optimal control policy for the load scheduling problem presented in this chapter allowed for the definition of an energetic efficiency benchmark for the investigated multi-source energy plant technology. Even if the Dynamic Programming

algorithm cannot be applied for the online optimization of the plant, the knowledge of the optimal policy allows to highlight the critical issues of the previously developed rule-based scheduling policies and suggested some modifications for improving its efficiency.

Moreover, a methodology for evaluating the optimal control strategy may play a central role even during the design phase of the plant. During the sizing procedute in fact, the knowledge of the maximum theoretical performance of the plant are essential to perform a fair comparison between different design parameters combinations (e.g. the size of the different machines, see Table 5.1), by decoupling the effect that such parameters have on plant performance from those that the same parameter may have in determining the sub-optimality degree of a non-optimal management policy. A size optimization similar to that performed by Barbieri et al. [1], which relies on a heuristic control algorithm, may in fact lead to the pursue of a certain combination of design parameters that is more tailored to maximize the optimality of the heuristic policy itself rather than that of the comprehensive system.

Finally, the optimal control DP formulation and the implementation may be properly exploited to easily reformulate a Model Predictive Control strategy. If in fact a shorter optimization horizon (e.g. daily based) is chosen, and prediction models for the external disturbances are developed, based on data from weather forecast for atmospheric conditions and historical data for energy demand, it is possible to iteratively apply the already implemented Dynamic Programming algorithm in a receding horizon fashion, thus optimizing the daily scheduling. Due to the reduced number of states and inputs to the plant and the relatively long sample time of the system, the algorithm would not require any additional simplification for an efficient implementation.

Nomenclature

Abbreviations and acronyms

AB	Auxiliary Boiler
ABS	Absorption Chiller
AC	Auxiliary Chiller
ASHP	Air Source Heat Pump
CAES	Compressed Air Energy Storage
CHP	Combined Heat and Power
COP	Coefficient Of Performance
DP	Dynamic Programming
GHG	Green-House Gases
GSHP	Ground Source Heat Pump
ICE	Internal Combustion Engine
LAES	Liquid Air Energy Storage
MGT	Micro Gas Turbine
MPC	Model Predictive Control
NOCT	Nominal Operative Cell Temperature
PH(E)S	Pump Hydroelectric (Energy) Storage
PV	Photovoltaic
STH	Solar Thermal Heating
XSHP	Generic Source Heat Pump

Symbols

c_{diss}	dissipation coeff.	$[-]$
E	energy	$[kWh]$
f	energy conversion factor	$[-]$
k	time step	$[-]$
J	objective function	$[-]$
P	power	$[kW]$
u	system input	$[-]$
w	external disturbance	$[-]$
x	system state	$[-]$
y	system output	$[-]$
Δt	sample time	$[hour]$
η	efficiency	$[-]$
π	control policy	$[-]$

Subscripts and Superscripts

el	electric
nom	nominal
th	thermal
0	initial
*	optimal

Bibliography

[1] E. S. Barbieri, Y. J. Dai, M. Morini, M. Pinelli, P. R. Spina, P. Sun, and R. Z. Wang, "Optimal sizing of a multi-source energy plant for power heat and cooling generation," *Applied Thermal Engineering*, vol. 71, no. 2, pp. 736–750, 2013. ix, 77, 78, 81, 83, 88, 89, 90, 96

[2] E. S. Barbieri, M. Morini, E. Munari, M. Pinelli, P. R. Spina, and R. Vecci, "Concurrent optimization of size and switch-on priority of a multi- source energy system for a commercial building application," in *69th Conference of the Italian Thermal Machines Engineering Association, ATI2014*, 2014. 78, 79, 80, 81, 82, 83, 84, 88, 89

[3] E. S. Barbieri, P. R. Spina, and M. Venturini, "Analysis of innovative micro-CHP systems to meet household energy demands," *Applied Energy*, vol. 97, pp. 723–733, Sep. 2012. 75

[4] M. Bianchi, L. Branchini, N. Cavina, A. Cerofolini, E. Corti, A. De Pascaie, V. Orlandini, F. Melino, D. Moro, A. Peretto, and F. Ponti, "Managing Wind Variability with Pumped Hydro Storage and Gas Turbines," *Energy Procedia*, vol. 45, pp. 22–31, 2014. 76

[5] A. D. Carvalho, P. Moura, G. C. Vaz, and A. T. de Almeida, "Ground source heat pumps as high efficient solutions for building space conditioning and for integration in smart grids," *Energy Conversion and Management*, vol. 103, pp. 991–1007, Oct. 2015. 75

[6] D. Chiappini, A. L. Facci, L. Tribioli, and S. Ubertini, "SOFC Management in Distributed Energy Systems," *Journal of Fuel Cell Science and Technology*, vol. 8, no. 3, 2011. 76

[7] A. L. Facci, L. Andreassi, and S. Ubertini, "Optimization of CHCP (combined heat power and cooling) systems operation strategy using dynamic programming," *Energy*, vol. 66, pp. 387–400, 2014. 50, 75, 76

[8] D. Y. Goswami and F. Kreith, *Handbook of Energy Efficiency and Renewable Energy*. Boca Raton, FL, USA: CRC Press, 2007. 82

[9] B. Greening and A. Azapagic, "Environmental impacts of micro-wind turbines and their potential to contribute to UK climate change targets," *Energy*, vol. 59, pp. 454–466, Sep. 2013. 74

[10] H. Ibrahim, A. Ilinca, and J. Perron, "Energy storage systemsCharacteristics and comparisons," *Renewable and Sustainable Energy Reviews*, vol. 12, no. 5, pp. 1221–1250, Jun. 2008. 75

[11] S. A. Kalogirou, "Solar thermal collectors and applications," *Progress in Energy and Combustion Science*, vol. 30, no. 3, pp. 231–295, Jan. 2004. 74

[12] B. Kantharaj, S. Garvey, and A. Pimm, "Compressed air energy storage with liquid air capacity extension," *Applied Energy*, vol. 157, pp. 152–164, Nov. 2015. 75

[13] X. Kong, R. Wang, and X. Huang, "Energy optimization model for a CCHP system with available gas turbines," *Applied Thermal Engineering*, vol. 25, no. 2-3, pp. 377–391, Feb. 2005. 76

[14] X. Kong, R. Wang, Y. Li, and X. Huang, "Optimal operation of a micro-combined cooling, heating and power system driven by a gas engine," *Energy Conversion and Management*, vol. 50, no. 3, pp. 530–538, Mar. 2009. 76

[15] R. M. Lazzarin, "Condensing boilers in buildings and plants refurbishment," *Energy and Buildings*, vol. 47, pp. 61–67, Apr. 2012. 75

[16] E. López González, F. Isorna Llerena, M. Silva Pérez, F. Rosa Iglesias, and J. Guerra Macho, "Energy evaluation of a solar hydrogen storage facility: Comparison with other electrical energy storage technologies," *International Journal of Hydrogen Energy*, vol. 40, no. 15, pp. 5518–5525, Apr. 2015. 75

[17] R. Luthander, J. Widén, D. Nilsson, and J. Palm, "Photovoltaic self-consumption in buildings: A review," *Applied Energy*, vol. 142, pp. 80–94, Mar. 2015. 74

[18] V. Marano, G. Rizzo, and F. A. Tiano, "Application of dynamic programming to the optimal management of a hybrid power plant with wind turbines, photovoltaic panels and compressed air energy storage," *Applied Energy*, vol. 97, pp. 849–859, 2012. 76

[19] K. Marik, Z. Schindler, and P. Stluka, "Decision support tools for advanced energy management," *Energy*, vol. 33, no. 6, pp. 858–873, Jun. 2008. 76

[20] R. McKenna, P. Ostman v.d. Leye, and W. Fichtner, "Key challenges and prospects for large wind turbines," *Renewable and Sustainable Energy Reviews*, vol. 53, pp. 1212–1221, Jan. 2016. 74

[21] S. Mitra, L. Sun, and I. E. Grossmann, "Optimal scheduling of industrial combined heat and power plants under time-sensitive electricity prices," *Energy*, vol. 54, pp. 194–211, Jun. 2013. 76

[22] P. S. Moura, G. L. López, J. I. Moreno, and A. T. De Almeida, "The role of Smart Grids to foster energy efficiency," *Energy Efficiency*, vol. 6, no. 4, pp. 621–639, Nov. 2013. 75

[23] L. Navarro, A. de Gracia, S. Colclough, M. Browne, S. J. McCormack, P. Griffiths, and L. F. Cabeza, "Thermal energy storage in building integrated thermal systems: A review. Part 1. active storage systems," *Renewable Energy*, Nov. 2015. 75

[24] L. Navarro, A. de Gracia, D. Niall, A. Castell, M. Browne, S. J. McCormack, P. Griffiths, and L. F. Cabeza, "Thermal energy storage in building integrated thermal systems: A review. Part 2. Integration as passive system," *Renewable Energy*, vol. 85, pp. 1334–1356, Jan. 2016. 75

[25] A. Nosrat and J. M. Pearce, "Dispatch strategy and model for hybrid photovoltaic and trigeneration power systems," *Applied Energy*, vol. 88, no. 9, pp. 3270–3276, Sep. 2011. 75

[26] H. Onovwiona and V. Ugursal, "Residential cogeneration systems: review of the current technology," *Renewable and Sustainable Energy Reviews*, vol. 10, no. 5, pp. 389–431, Oct. 2006. 75

[27] A. Pandey, V. Tyagi, J. A. Selvaraj, N. Rahim, and S. Tyagi, "Recent advances in solar photovoltaic systems for emerging trends and advanced applications," *Renewable and Sustainable Energy Reviews*, vol. 53, pp. 859–884, Jan. 2016. 74

[28] C. Xi, Y. Hongxing, L. Lin, W. Jinggang, and L. Wei, "Experimental studies on a ground coupled heat pump with solar thermal collectors for space heating," *Energy*, vol. 36, no. 8, pp. 5292–5300, Aug. 2011. 76

CHAPTER 6
DESIGN AND SIZING OF A HYBRID HYDRAULIC SOLUTION FOR A MID-SIZE EXCAVATOR

To limit the environmental impact of the various mobile machineries employed in the construction area, tighter regulations have been introduced, limiting the pollutant emissions and the fuel consumption. Many different technologies have been consequently developed to improve the energetic efficiency of earth moving systems. Regarding such topic, the various aspects related to the introduction of a hybrid hydraulic solution for energy recovery are investigated in this chapter, focusing on the case study of a middle size excavator.

The development of an efficient energy recovery subsystem poses different challenging issues. The effect that both the many conceivable different layout configuration for the hybrid solution, and the sizing of the newly introduced components, have on the overall efficiency of the plant must be properly assessed. Finally, the introduction of a new subsystem introduces some additional degrees of freedom to the plant, due to the possibility of actively control the energy recovery procedure. A proper methodology must be therefore developed to assess the effective potential of each one of the proposed layouts, to correctly guide the choice of the manufacturer; as the performance of the control policy adopted to manage the energy recovery system may greatly affect the achieved fuel reduction, a tool to estimate the optimal management policy must be first developed. If in fact the control strategy is not able to guarantee the optimality of the plant management process, the comparison between different solutions may be distorted by the sub-optimality degree of the adopted control policy, which may favour the energy recovery procedure only for certain plant layouts and size configurations.

To obviate to this issue, a methodology that takes advantage of the Dynamic Programming algorithm illustrated in Chapter 4.2 is developed to assess the maximum performance that can be achieved with each different excavator configuration, removing any bias deriving from the implementation of a suboptimal control strategy. Based on a previously developed detailed model of the plant, a simplified control-oriented

one is developed and used to implement the Dynamic Programming algorithm. The procedure is used to estimate the performance of four different hybrid layouts based on the predicted fuel economy improvements over a regulated working cycle which depicts a typical operating maneuver for the investigated machinery.

6.1 Introduction

In the traditional mobile construction machinery layout, an internal combustion engine provides the primary source of energy to the plant. The mechanical energy at the engine shaft is then directly transformed into hydraulic elastic pressure energy by means of a pump. This solution allows to exploit the higher power densities achievable with the hydraulic layout, while maintaining a usage flexibility rather unobtainable with a mechanical powertrain. In this typical layout, the engine is usually kept at constant nominal speed, to maximize its efficiency, and all the users are managed by means of a complex hydraulic circuit which is designed to ensure the desired drivability performance.

With the introduction of tighter regulations for the reduction of environmental pollution from construction machineries, many solutions have been developed by the manufacturers. Due to the fact that the hydraulic circuit represents the most complex and articulated part of the system and is responsible for the management of the energy flow, all the research focuses on improving this part of the system. The benefits on energy efficiency related to the improvement of the single hydraulic components (pump, valves, actuators), achieved by the reduction of mechanical losses and the enhancement of the flow characteristics, have been almost maximized. For this reason the research on mobile machinery energetic improvement moved towards a systematic point of view, where the entire process is viewed and optimized as one, accounting for the interaction between the different components.

The improvement of valve controlled systems and the decentralization of the hydraulic power production represent the two ways adopted to improve the overall efficiency of the system. An example of the former approach can be found with digital hydraulic systems, using on-off valve technologies [17] or with the development of independent metering configurations [18]. These solutions maintains the single pump/multiple actuators layout but operates a redesign of the flow control procedure which, while maintaining the drivability performance of classic hydraulic solutions (e.g. load sensing), allows for a more efficient management of the energy flows, thus reducing the various losses. The decentralization approach on the other

hand bases the reduction of flow throttling losses on a complete redesign of the entire engine-to-actuators process. Ivantysynova et al. [7, 22], for example, conceived a power split hydraulic powertrain that avoids the flow throttling related to the motion control typical of load sensing architectures by introducing a system of multiple variable displacement pumps, one for each actuator, each controlled to produce only the desired amount of flow at the right pressure. Inderelst et al. [14] proposed instead the introduction of multiple hydraulic transformers to control the linear and rotational movements of the actuators.

The adoption of energy recovery technologies represents another interesting option for the efficiency improvement in mobile machinery equipments. As both the kinetic energy from rotating parts (e.g. turrets) and the gravitational energy from excavator tools is usually dissipated during the braking and lowering phases, the integration of a suitable accumulator system represents a promising solution to recover this otherwise wasted energy and use it to reduce the load to the thermal engine during future system operations. Both hydraulic accumulators [11] or battery supplied electric generators [16] are actually used for the development of energy recovery systems. Electrical hybrid solutions for construction machineries have been widely investigated [19, 21] as they offer great usage flexibility and high energy densities. Hybrid hydraulic recovery on the other hand provide a low cost solution and offer higher power density, allowing for a more effective energy recovery from applications, such excavators, involving fast movements and high inertias, proving to be perform in this case better than electrical based solutions [9].

With the introduction of regenerative systems, an additional problem is introduced, related to the definition of the correct strategy for the energy recovery and reutilization process, able to maximize the fuel consumption reduction that can be achieved. Many different approaches have been followed to define the correct energy management strategy for hybrid hydraulic solutions. Simple rule based solutions represent a simple and effective approach that, if properly calibrated, provides effective control algorithms. As an example, Hui et al. [12] developed a rule based torque control strategy for the management of the energy recovery procedure from hydraulic/electric hybrid heavy vehicles. Kim et al. [15] used the results from a Dynamic Programming based optimization of a compound hybrid excavator to enhance the performance of another simple rule based excavator, by exploiting the results deriving from the knowledge of the optimal system management strategy. More complex approaches, exploiting tools based on optimal control theory where also followed. Deppen et al. [8] for example developed a Model Predictive Controller for the optimal

energy management of a hybrid hydraulic vehicle based on a linear approximation of the system and the simplifying assumption that the future driving cycle is constant during the short optimization horizon. A more advanced implementation of the MPC methodology was performed by Feng et al. [10] who developed a Markov chain model for a stochastic estimation of future driving conditions on an hybrid hydraulic vehicle. Bender et al. [1, 2] still focused on the development of a driving prediction model to aid the energy management optimization process, but based their algorithm on the iterative learning of the vehicle driving profiles.

In the present chapter, the development of an hybrid hydraulic energy recovery system for a middle-size, load-sensing drive, hydraulic excavator is investigated, analyzing the fuel consumption reduction performance provided by four different hybridization layouts. As mentioned before, the introduction of an energy recovery system requires the parallel development of a suitable control strategy able to guide the process. The scope of this chapter on the other hand is not that of defining an on-line implementable control strategy but is rather that of comparing different circuital solutions for the development of a regenerative application. It is clear however that the definition of the optimal layout cannot be carried out regardless of the evaluation of the associated optimal control strategy, as the sub-optimality degree that may be introduced by a non-globally optimal policy can distort the comparison, by penalizing a solution that perhaps is just badly managed. To obtain a fair estimation of the true potentiality of each solution, a deterministic Dynamic Programming algorithm is first exploited to obtain the guaranteed optimal control strategy for the plant. While this approach is clearly unfeasible for an online implementation, as it requires the knowledge of the future driving cycle conditions, it represents a valuable tool to derive a benchmark for the performance achievable from a particular solution.

In the following, a mathematical model of the system, developed to represent both the basic configuration and the different hybrid layouts, is briefly described. As this original detailed model is not suitable for the implementation of the DP algorithm, a simplified one is first developed and compared to the original. Then, a deterministic formulation for the dynamic optimization of the fuel consumption reduction is conceived and a procedure for evaluating the optimal control policy for each layout configuration is derived. Finally, based on a Design of Experiment (DoE), the various solutions are sized and compared.

6.2 Excavator layout and test cycle

The middle size excavator taken into account for this study is based on the typical hydraulic machinery layout which can be roughly divided in three parts [13]:

1. A fluid power generation system transform the mechanical power output from an ICE into hydraulic power;
2. A valve block controls the flow of pressurized fluid towards the different end actuators;
3. Hydraulic linear actuators and motors operates the different excavator tools by converting the hydraulic power back into mechanical.

The digging tool layout (Figure 6.1) is that typical of trench excavation procedures and is composed of a boom, an arm, and a bucket, with the relative hydraulic actuators.

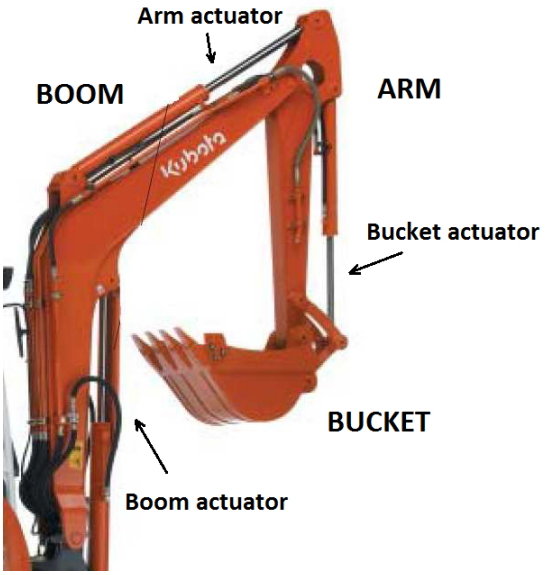


Figure 6.1: Excavator tools configuration for trench digging procedures

The excavator hydraulic circuit is based on the well-established Load-Sensing (LS) architecture [20], where the highest pressure required from the parallel hydraulic drives, named load sensing pressure p_{LS}, is detected and fed back to the pump.

Especially designed hydraulic compensators will in turns adjust the pump delivery pressure p_D to maintain a certain pressure margin Δp_{LS} against the load pressure p_{LS}. This is usually achieved by exploiting a variable displacement pump, which is able to adjust its flow output to obtain the desired deliver pressure.

The main benefit from the adoption of an LS architecture lies in the possibility of achieving excellent drivability performance with the adoption of just a simple calibrated hydraulic valve. As the velocity of an actuator is in fact related to the flow directed towards the component, it is evident that the ability to precisely control such flow ensures the achievement of the desired drivability performances. To understand how the LS architecture grants this, it is necessary to refer to Bernoulli's equation for the flow $\dot{V}$ through an orifice under quasi-steady conditions:

$$\dot{V} = c_d \, \Omega \, \sqrt{\frac{2\Delta p}{\rho}} \tag{6.1}$$

where c_d is a discharge coefficient, Ω is the flow section, Δp is the pressure difference across the orifice, and ρ is the fluid density.

Given that the operator can actively adjust the flow section Ω by opening or closing the orifice, if a constant Δp is maintained (thanks to the LS system), then the operator has the complete control over the amount of fluid flowing through the orifice, as all the other terms in the equation are constant. From an energetic viewpoint, the the LS architecture provides additional benefits if compared to the simpler fixed pump pressure solution, as it allows the pump to operate to a pressure which is just marginally higher to that required from the user. Moreover, if (as usual) a variable displacement pump is adopted, it is possible to exactly match the flow required from the actuator to that generated by the pump itself, thus avoiding the throttling losses otherwise associated to the lamination of the excessive flow produced by a fixed displacement pumps.

While in its basic conception the LS solution is limited to the control of one single actuator, it can be easily expanded to fit the case of multiple parallel users. If the maximum load pressure is detected and set as the load sensing target, it is possible to control the corresponding actuator in the usual load sensing fashion and adjust the pressure difference across the remaining orifices by means of hydraulic compensators, which laminates the fluid to retain the controllability even on the secondary loads. This procedure however, even if fairly simple from a technological point of view, presents the main drawback of increasing the throttling losses.

A detailed ISO schematic representation of the hydraulic load sensing circuit of the excavator is depicted in Figure 6.2.

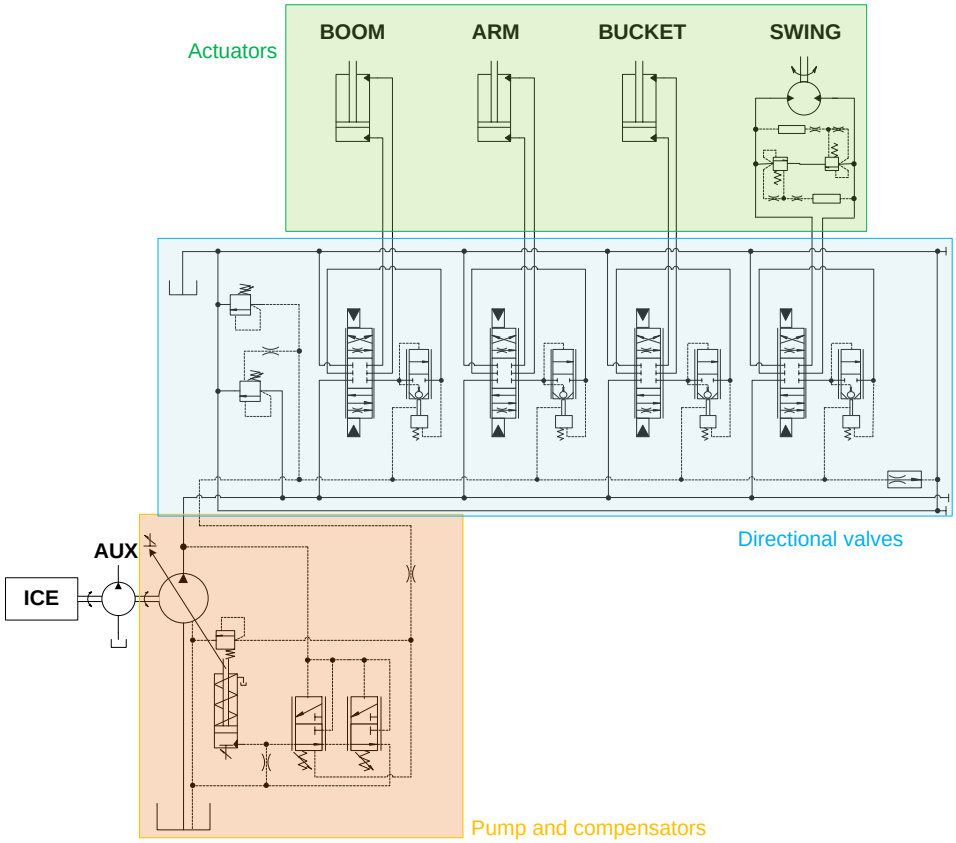


Figure 6.2: ISO scheme for the standard excavator configuration

To evaluate the energetic efficiency of the hydraulic excavator into its standard configuration and assess the possible benefits from the introduction of a hybrid hydraulic recovery system, the definition of a proper test cycle is required. A reference test procedure is derived from the Japanese technical standard JCMAS H020:2007 which is based on a schematic representation of a typical trench excavation procedure. Four different working phases are accounted: trench digging, soil levelling, straight travelling (back and forward), and low idling standby. To limit the effect of operator behaviour and environmental conditions, the test cycle is performed without any soil interaction, lifting an empty bucket, and with all the hydraulic functions at maximum

speed. While this represents an excessive simplification of the digging procedure, it is on the other hand the only effective way to ensure the repeatability of the procedure.

Given the profile of the excavation procedure illustrated in the JCMAS standard, reference position trajectories for the different actuators of the excavator (boom, arm, bucket, and turret) are obtained. Moreover, as the straight travelling and idling standby phases account only for 1/4 of the entire procedure and are not affected by the behaviour of the energy recovery system, they are neglected in the following analysis.

6.3 Excavator model

A detailed physical based lumped parameter model for the simulation of the excavator dynamics has been developed during the years (see [3–6]) and implemented into the AMESim®environment. This detailed approach, while providing excellent correlations to experimental data, cannot be properly used for the development of a control algorithm, due to the excessive number of states required to describe the system and the high complexity of the model. A simplified version of the excavator model has been therefore developed, based on a quasi-static approximation of the hydraulic dynamics. Moreover, the detailed model is based on a "direct causality" representation of the process which, while consistent with the physical reality, has the downside of introducing some additional degrees of freedom that should be accounted in the development of the controller. To further reduce the complexity of the problem, an "inverse causality" approach is followed when deriving the simplified model for the plant. The distinction between direct and inverse causality will become clearer once the models are properly introduced.

6.3.1 Direct causality model

In the following paragraphs, the direct causality models developed for the simulation of the different components of the excavator are briefly described. These models rely on a set of differential equations simulating the pressure rise dynamics in the various hydraulic components and the mechanical dynamics of the moving parts (from valve spools positions to engine shaft speed and excavator tools velocities).

Pump and compensators The excavator equips an axial piston variable displacement pump which is primarily modelled as a simple flow generator. The instantaneous displacement of the pump, V_d, is geometrically related to the swash plate angular position α. Volumetric and mechanical-hydraulic efficiency, η_v and η_{mh}, have been

experimentally determined according to the ISO 4409:1986 standards and correlated through a black box approach to pump speed n, swash plate position α, and pump delivery pressure p_D. Pump outlet flow $\dot{V}_p$ can be calculated as:

$$\dot{V}_p = \eta_v(p_D, \alpha, n) \cdot V_d(\alpha) \cdot \frac{n}{60} \tag{6.2}$$

where Δp is the pressure difference between pump inlet and outlet sections. The torque T_p that the pump applies to the engine shaft is:

$$T_p = \frac{\Delta p \cdot V_d(\alpha)}{2\pi} \cdot \frac{1}{\eta_{mh}(p_D, \alpha, n)} \tag{6.3}$$

The angular position of the swash is controlled by three compensators, which modulates the outlet flow from the pump to obtain the desired value of delivery pressure. The first one is a pressure compensator (PC) and acts as a relief valve by limiting the maximum system pressure. The second one is a flow compensator (FC) which ensures that the desired pressure margin Δp_{LS} is established between pump delivery and load-sensing line. The third one is a torque limiter compensator (TL) which reduces the actual pressure margin Δp_{LS} if the torque required to the engine exceeds a certain value. The three compensators are modeled following a lumped parameter approach relying on geometrical data provided by the manufacturer, which are used to develop detailed component sub-models inside the AMESim®environment. These sub-models accounts for the various chambers and flow areas in the compensators by means of two sets of equations, one for the fluid-dynamic part (FDM) and one for the mechanical geometrical part (MGM) of the component. The first calculates the pressure inside the actuators chambers, given the flow rates in the actuators, while the second evaluates the spool position given the acting forces. As the detailed description of the compensator models and their validation is outside the scope of this thesis, the interested reader is referred to [3]. The overall effect of the compensators is to generate a pressure, p_{act}, which can be correlated to the pump delivery pressure p_D and the load sensing pressure p_{LS}. The actuator pressure thus generates a torque on the swash plate which tends to increase the pump displacement.

The position of the pump swash plate is evaluated from the mechanical equilibrium equation:

$$J_{EQ}\ddot{\alpha} + c\dot{\alpha} = \sum_{n=1}^{N} T_n + T_{act} \tag{6.4}$$

where J_{EQ} is the swash plate inertia, T_n is the torque applied from each of the N

different pump cylinders, c is a viscous friction coefficient, and T_{act} is the torque deriving from the action of the compensators. Thanks to an experimental campaign [6], it was possible to estimate the value of the viscous friction coefficient and a black box correlation between the overall cylinders torque and pump delivery pressure, rotational speed, and swash plate position.

To derive pump delivery pressure, mass conservation equation is applied to a control volume at the outlet of the component, once the flow rate directed toward the valves, $\dot{V}_{valves}$, is known. Figure 6.3 depicts a schematic of the comprehensive pump model.

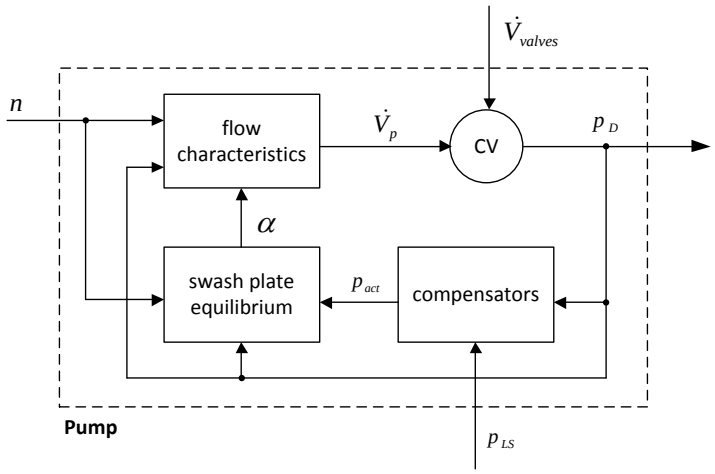


Figure 6.3: Pump direct causality model

Directional valves The equipped load sensing flow sharing valve block (Walvoil®) is composed of a multiple number of identical sections, one for each actuator in the excavator. Figure 6.4, depicts the ISO schematic of a single section of the valve block. The generic valve section i connects the two ports A_i and B_i of the corresponding hydraulic actuator to respectively the pump delivery section, P, and the tank, T. When the operator of the excavator controls a particular movement of the machine, he controls the position x_i of a spool which determines the metering area in the corresponding section of the valve block.

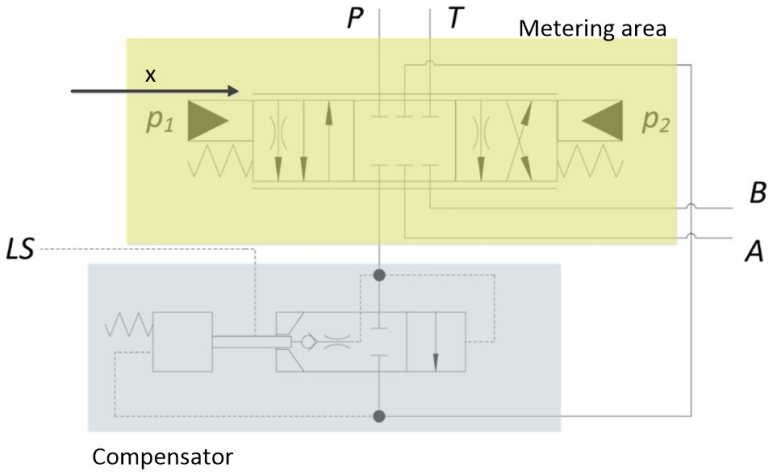


Figure 6.4: Flow control valve ISO scheme

During the functioning of the excavator, the valve block detects the load pressure from the actuator section connected to pump delivery, and sends that signal to the load sensing line, LS. If the pressure detected is the higher among all the actuators, the pump establish the right load sensing pressure margin Δp_{LS} across the valve block metering section. Otherwise, an higher pressure difference is established, based on the load from the most burdened actuator, and a compensator laminates the flow in the valve block to restore the desired pressure difference across the metering area. Thanks to the compensated LS architecture, the fluid flowing through the metering section (and therefore the velocity of the corresponding excavator part) is uniquely determined by the position of the spool (as depicted in Equation 6.1) in both the cases, and the driver has the complete control over the operation.

As the total amount of flow requested to the pump may exceed the maximum deliverable, the valve block incorporates a flow sharing mechanism which, under such circumstances, reduces the pressure margin across all the orifices, by additionally laminating the fluid, and thus ensures that drivability is maintained. In this case the flow rate is evenly reduced in each section and the proportionality between spool position and the flow rate is retained.

Except for the turret, which equips a rotary motor, all the other actuators are equipped with linear hydraulic cylinders and for that reason an asymmetry between the conditions at ports A and B is always present, due to the different areas at the two piston extremities due to the presence of the rod. At each section i, the following nomenclature is adopted to define the four metering flow sections:

- $A_{PA,i}(x_i)$ for the section connecting pump line to port A;
- $A_{PB,i}(x_i)$ for the section connecting pump line to port B;
- $A_{AT,i}(x_i)$ for the section connecting port A to the tank;
- $A_{BT,i}(x_i)$ for the section connecting port B to the tank.

Obviously only the couple $A_{PA,i}(x_i)$, $A_{BT,i}(x_i)$ or the couple $A_{PB,i}(x_i)$, $A_{AT,i}(x_i)$ are open at the same time. Due to the different conditions at the outlet ports, each of the four different sections connecting the P and T ports to the A and B ports presents a different functional relationship with the spool position x_i, which is defined by the valve manufacturer.

During both the active phases and passive phases, the pump has to generate a certain amount of power to move the corresponding tool. While this is obvious for the active phase, when for example a load has to be raised, it would seem reasonable that during the passive lowering phase, the gravitational force would be sufficient to drive the tool down, without any request of energy from the pump. However, to maintain a perfect control on the tools, for both drivability and safety conditions, the free gravitational assisted lowering is not pursued. On the contrary, a counter-pressure is generated on the line returning to the tank by properly defining the corresponding flow section $A_{xT}(x_i)$ so that a large lamination of the flow occurs. A positive pressure is therefore always detected by the load sensing line and the pump has to spend energy in both the lowering and rising phases.

The developed valve model relies on detailed geometrical data from the manufacturer and is based on a white box approach similar to that used for modelling the pump compensators. For a detailed explanation of the model and its validation, the reader can refer to [4]. For each section of the valve block, the flow rate coming from the pump and directed to the corresponding actuator, $\dot{V}_{valve,i}$, and the flow rate coming from the corresponding actuator and directed to the tank, $\dot{V}_{tank,i}$, are calculated, together with the flow to the respective load ports, $\dot{V}_{A,i}$ and $\dot{V}_{B,i}$. The computation is based on the knowledge of the metering spool position x_i (defined by the operator), the pump delivery pressure p_D, the load sensing pressure p_{LS}, and the pressures from the corresponding hydraulic actuator sections $p_{A,i}$ and $p_{B,i}$. This evaluation is carried out for each valve section and the total flow rate required to the pump, $\dot{V}_{valves}$, is equal to the algebraic sum of every section flow rate, $\dot{V}_{valve,i}$. Figure 6.5 depicts the simplified schematic model of each valve section.

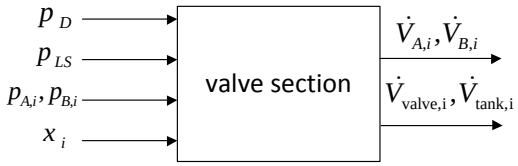


Figure 6.5: Valve section direct causality model

Kinematics and actuators Excavator front tools (boom, arm, and bucket) are modelled with the help of AMESim®Planar Mechanics library, which accounts for the components geometry and inertia and calculates the movement (position y_i and velocity $\dot{y}_i$) deriving from the application of a certain force F_i. This force is applied to each tool by means of a linear hydraulic actuator, also modeled with the aid of AMESim®, following the schematic depicted in Figure 6.6. Given the flow rates input from the corresponding valve section $\dot{V}_{A,i}$ and $\dot{V}_{B,i}$, the actuator model evaluates the pressure in both sides of the actuator $p_{A,i}$ and $p_{B,i}$, and the force F_i.

The swing motion of the turret is simulated in an similar manner, with the only difference that the component is moved by means of a rotary motor.

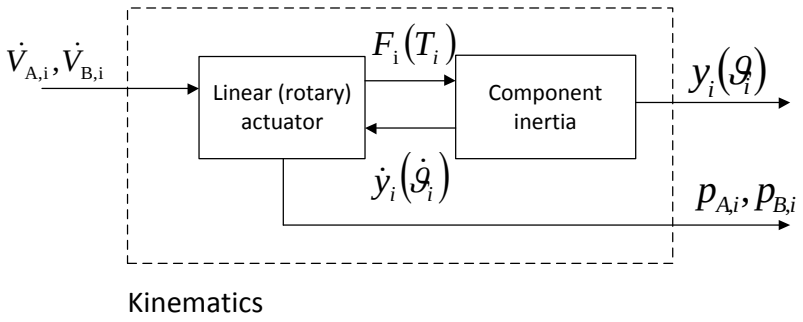


Figure 6.6: Kinematics and actuators direct causality model

The kinematics model is not able to simulate the interaction with the soil and can only be used to simulate excavator free movements. A more detailed description of the kinematic model of the excavator can be found in [3].

Internal combustion engine The internal combustion engine (ICE) is the prime energy source for the entire excavator system. A simple black box approach is adopted to model the component, whit the torque T_{eng} generated from the engine evaluated as a function of the engine speed n and the injected fuel flow rate $\dot{m}_{fuel}$ (see Figure 6.7).

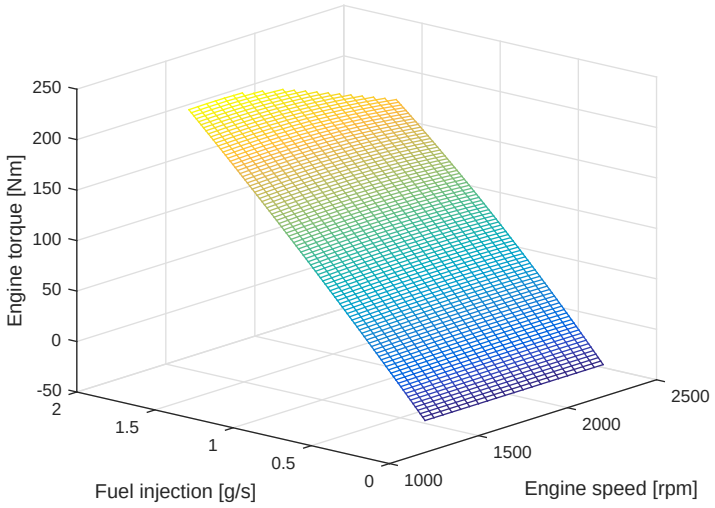


Figure 6.7: Steady state correlation for the ICE torque production

The speed n is then obtained from the dynamic equilibrium at the shaft connecting the engine to the pump:

$$J_{eng}\frac{30}{\pi}\dot{n} = T_{eng} - T_p \tag{6.5}$$

with J_{eng} the equivalent inertia of the whole ICE-pump system.

A PI controller, calibrated to recreate the dynamic performance guaranteed by the manufacturer, adjusts the amount of fuel injected in the engine to ensure that a constant reference speed n_{target}, equal to 2200 rpm, is tracked. The schematic model of the ICE and shaft inertia is depicted in Figure 6.8.

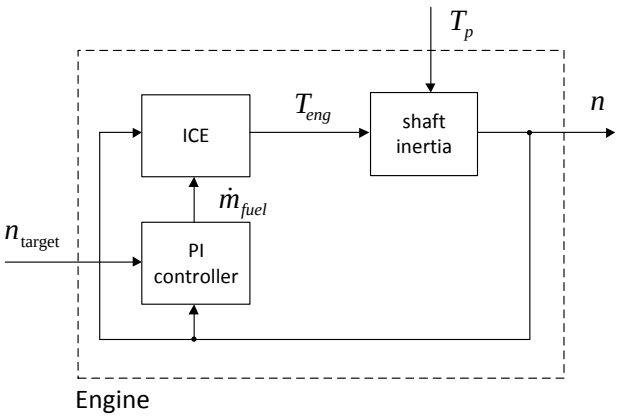


Figure 6.8: Internal combustion engine direct model

Comprehensive excavator model and cycle simulation The models presented in the previous section are assembled to obtain the comprehensive excavator model depicted in Figure 6.9. To track the profiles from the JCMAS duty cycle, a set of PI controllers is introduced to mimic the actions of an operator, who adjusts the position of the various spools depending on the error between actuators real and reference positions.

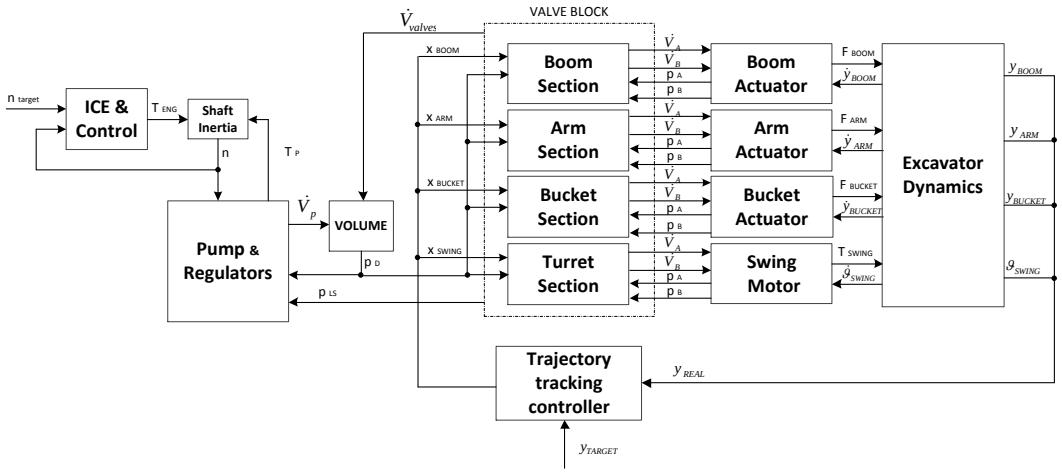
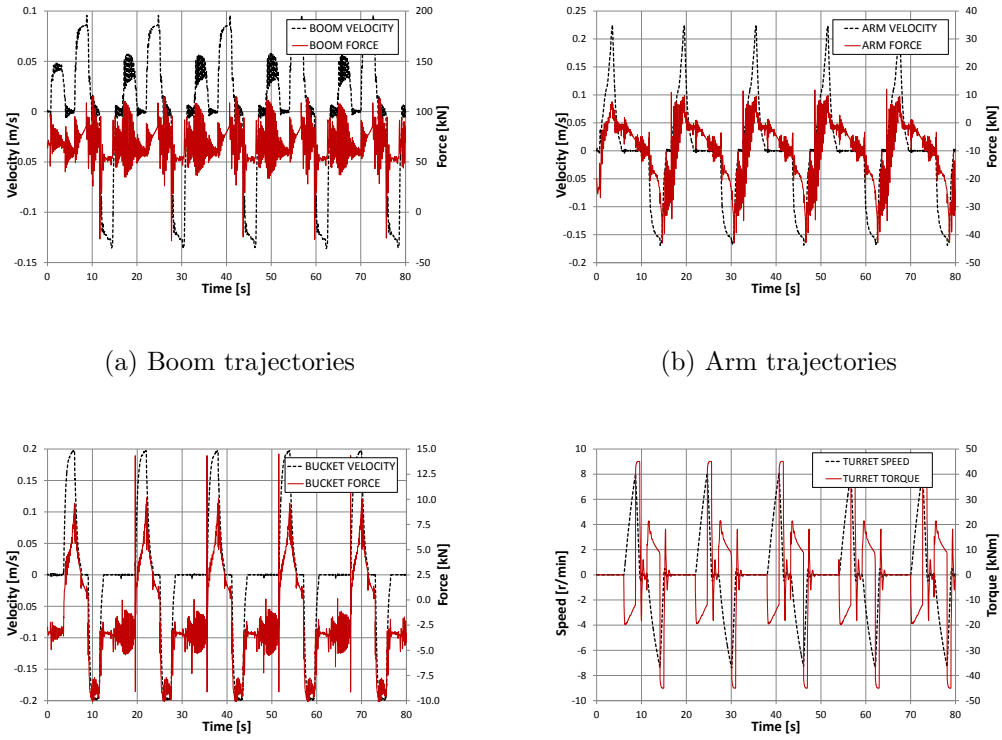


Figure 6.9: Comprehensive excavator direct causality model

The comprehensive excavator model is used to simulate the active phases of the JCMAS trench digging cycle. The derived profiles of actuators position and forces are shown in Figure 6.10 (digging operation) and Figure 6.11 (levelling operation). Note that during the levelling operation, neither the bucket nor the turret are moved.



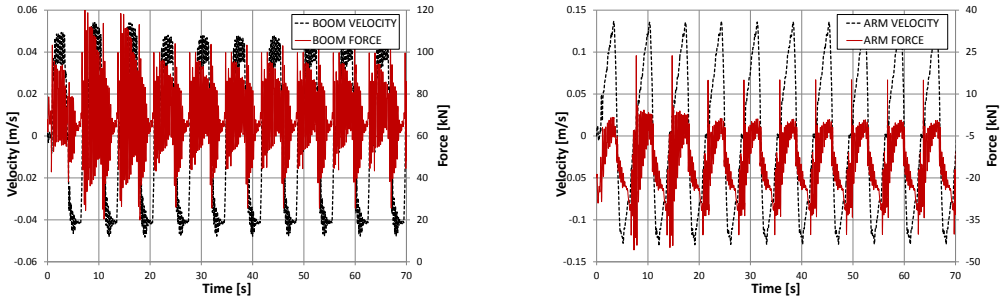
(a) Boom trajectories

(b) Arm trajectories

(c) Bucket trajectories

(d) Turret trajectories

Figure 6.10: Actuators position and resulting forces during the digging transient



(a) Boom trajectories

(b) Arm trajectories

Figure 6.11: Actuators position and resulting forces during the levelling transient

6.3.2 Inverse causality simplified model

To efficiently implement a control algorithm, it is worth to eliminate all the non-essential dynamics from the plant model. This is especially true if the Dynamic Programming algorithm is to be used, as the high number of states in the detailed excavator model (e.g. the pressures in all the valves orifices) would rapidly lead the computational and memory cost of the optimization up to values unsuitable for any calculator (see the so-called "curse of dimensionality", Section 4.2.3). To obviate this, a reduced complexity model has been developed, neglecting all the non-essential dynamics of the excavator, and reducing the entire excavator representation (in its basic configuration) to a set of algebraic sub-models. This procedure is justified by the fact that the dynamics of the hydraulic components are significantly faster than those associated to the motion of the excavator and the regenerative system to be added, which are the only one relevant for a dynamic description of the energy recovery process.

Moreover, the detailed model embeds two sets of PID controllers, one for the torque generation in the ICE (which adjusts the injected fuel to match a target engine speed), and one which simulates the behaviour of the operator, who controls the position of the spools in the valve block to follow the prescribed cycle trajectories. As mentioned before, this approach relies on what is called a "direct causality" representation of the phenomena, where some actions are taken as a delayed consequence

to the system dynamic response. While consistent with the physical reality, this introduces both additional degrees of freedom and additional states to the system (e.g. the integrators and buffers in the PID controllers). To further simplify the problem formulation, these decision processes are now taken based on an "inverse causality" model, which relies on the hypothesis that the real trajectories (i.e. engine speed and excavator tools position) are perfectly tracked so that the value of the decision parameters (i.e. injected fuel and valve positions) can be known instantly, based on a quasi-static dynamic equilibrium at the components. This approach leads to an inversion of the causality of the models, which will be henceforth referenced to as "inverse causality" models.

The different inverse causality sub-models are described in the following paragraphs; the order in which the models are presented follows the causality strategy adopted in describing the system.

Kinematics and actuators Assuming that transient profiles depicted in Figures 6.10 and 6.11 are perfectly matched, the speed $v_i = \dot{y}_i$ and the force F_i at the generic actuator i are known for the entire working cycle. For the generic linear piston types actuators, the flow rate to the two different ports of cylinder can be expressed as:

$$\dot{V}_{A,i} = A_{A,i} \cdot v_i \tag{6.6}$$

$$\dot{V}_{B,i} = -A_{B,i} \cdot v_i \tag{6.7}$$

where $A_{A,i}$ and $A_{B,i}$ are the hydraulic piston areas at the two different piston sections. Chamber pressures are defined according to the sign of piston velocity v_i, based on the following equations:

$$\begin{cases} p_{B,i} = p_{C,i} \\ p_{A,i} = \dfrac{F_i}{A_{A,i}} + p_{B,i} \cdot \dfrac{A_{B,i}}{A_{A,i}} & \text{if } v_i \geq 0 \\ \\ p_{A,i} = p_{C,i} \\ p_{B,i} = -\dfrac{F_i}{A_{B,i}} + p_{A,i} \cdot \dfrac{A_{A,i}}{A_{B,i}} & \text{if } v_i < 0 \end{cases} \tag{6.8}$$

where $p_{C,i}$ is a counter pressure due to resistance introduced by the outlet orifice of the valve section controlling the device.

For the turret hydraulic motor, the correlations are symmetrical and based on the

sign of the angular velocity $\omega_i = \dot{\theta}_i$. If $\omega_i > 0$, i.e. the rotation is clockwise, flow rates and pressures are evaluated following:

$$\dot{V}_{A,i} = \frac{\omega \cdot V_d}{\eta_v} \tag{6.9}$$

$$\dot{V}_{B,i} = -\dot{V}_{A,i} \tag{6.10}$$

$$p_{B,i} = p_{C,i} \tag{6.11}$$

$$p_{A,i} = p_{B,i} + \frac{T_i \cdot 2\pi}{V_d \cdot \eta_{mh}} \tag{6.12}$$

where V_d is the motor displacement, η_v and η_{mh} are respectively the volumetric and mechanical-hydraulic (constant) efficiencies, and T_i is the torque applied to the turret during the working cycle (Figure 6.10). If the rotation is counterclockwise ($\omega_i < 0$), the subscripts $_A$ and $_B$ are switched.

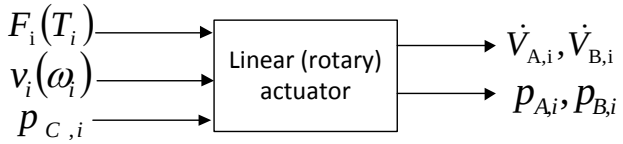


Figure 6.12: Actuators reverse causality model

Directional valves Valve sections reverse causality model depends on the sign of volumetric flow rate $\dot{V}_{A,i}$ towards the actuator. If $\dot{V}_{A,i}$ is greater than zero, then section A_i is connected to the pump line P, and section B_i is connected to the tank line T. The following correlations are adopted in this case:

$$\dot{V}_{valve,i} = \dot{V}_{A,i} \tag{6.13}$$

$$\dot{V}_{tank,i} = -\dot{V}_{B,i} \tag{6.14}$$

$$p_{LS,i} = p_{A,i} \tag{6.15}$$

Moreover, the knowledge of the pump to actuator flow, $\dot{V}_{valve,i}$, allows to determine the pump to port A_i flow area from Bernoulli's equation for steady flows (Equation 6.1) as:

$$A_{PA,i}(x_i) = \frac{\dot{V}_{valve,i}}{c_d}\sqrt{\frac{\rho}{2\Delta p_{LS}}} \tag{6.16}$$

Once the metering section $A_{PA,i}(x_i)$ is known, it is possible to determine the position of the spool, x_i, by simply inverting the correlation provided by the manufacturer. From the knowledge of spool position it is now possible to derive the counter-pressure $p_{C,i}$ at port B_i, still exploiting Bernoulli's Equation 6.1:

$$p_{C,i} = p_T + \frac{\dot{V}_{tank,i} \cdot \rho}{2 \cdot c_d^2 \cdot A_{BT,i}^2(x_i)} \tag{6.17}$$

where p_T is the tank pressure.

If on the other hand $\dot{V}_{A,i}$ is lower than zero, the following correlations are used:

$$\dot{V}_{valve,i} = \dot{V}_{B,i} \tag{6.18}$$

$$\dot{V}_{tank,i} = -\dot{V}_{A,i} \tag{6.19}$$

$$p_{LS,i} = p_{B,i} \tag{6.20}$$

$$A_{PB,i}(x_i) = \frac{\dot{V}_{valve,i}}{c_d}\sqrt{\frac{\rho}{2\Delta p_{LS}}} \tag{6.21}$$

$$p_{C,i} = p_T + \frac{\dot{V}_{tank,i} \cdot \rho}{2 \cdot c_d^2 \cdot A_{AT,i}^2(x_i)} \tag{6.22}$$

Finally, the load sensing signal sent to the pump and the total flow rate from the pump to the valves are evaluated as follows:

$$p_{LS} = \max\{p_{LS,i}\} \tag{6.23}$$

$$\dot{V}_{valves} = \sum_{i=1}^{N} \dot{V}_{valve,i} \tag{6.24}$$

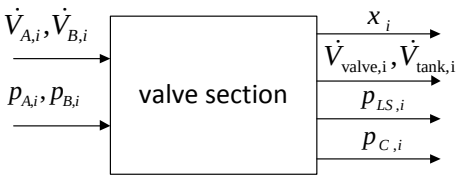


Figure 6.13: Valve section reverse causality model

Pump and compensators Following a quasi-steady approach, the pump outlet flow $\dot{V}_p$ is set equal to the flow toward the valves, $\dot{V}_{valves}$, while pump delivery pressure p_D and the pump differential pressure Δp can be evaluated by means of the following simple algebraic equations:

$$p_D = p_{LS} + \Delta p_{LS} \tag{6.25}$$

$$\Delta p = p_D - p_T \tag{6.26}$$

As pump speed n is assumed constant, it is possible to derive the pump torque T_p by inverting Equations 6.2 and 6.3, as:

$$T_p = T_p(\Delta p, p_D, \dot{V}_p, n) \tag{6.27}$$

Finally, given the desired load sensing margin $\Delta p_{LS,target}$, the maximum permitted system pressure p_{max} and the maximum pump torque T_{max}, it is possible to estimate the real margin Δp_{LS} achieved, due to the effect of the compensators, as follows:

$$\Delta p_{LS} = \begin{cases} \Delta p_{LS,target} & \text{if } p_D < p_{max} \ \& \ T_p < T_{max} \\ p_{max} - p_{LS} & \text{if } p_D \geq p_{max} \\ \Delta p_{LS,target} - \Delta p_{LIM}(\dot{V}_p, p_{LS}, T_p) & \text{if } T_p \geq T_{max} \end{cases} \tag{6.28}$$

The schematic of the reverse causality model of the pump and compensators is depicted in Figure 6.14.

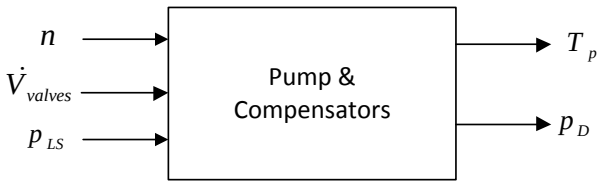


Figure 6.14: Pump and compensators reverse causality model

Internal combustion engine Assuming that the engine controller can always keep the engine speed at the constant reference value, the torque delivered by the engine equals that required from the pump ($T_{eng} = T_p$) and the corresponding fuel mass flow rate $\dot{m}_{fuel}$ is obtained by inverting the correlation used in the direct ICE model (see Figure 6.7). The schematic of the inverse model is depicted in Figure 6.15.

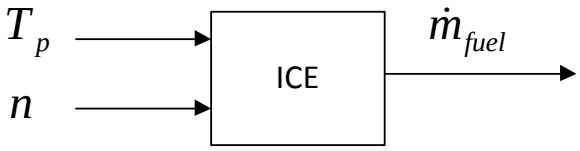


Figure 6.15: ICE reverse causality model

Comprehensive inverse causality model and validation Figure 6.16 depicts the complete model of the basic excavator layout in the reverse causality representation. The model is implemented into the Simulink®environment and is used as the basis for the implementation of the deterministic Dynamic Programming algorithm.

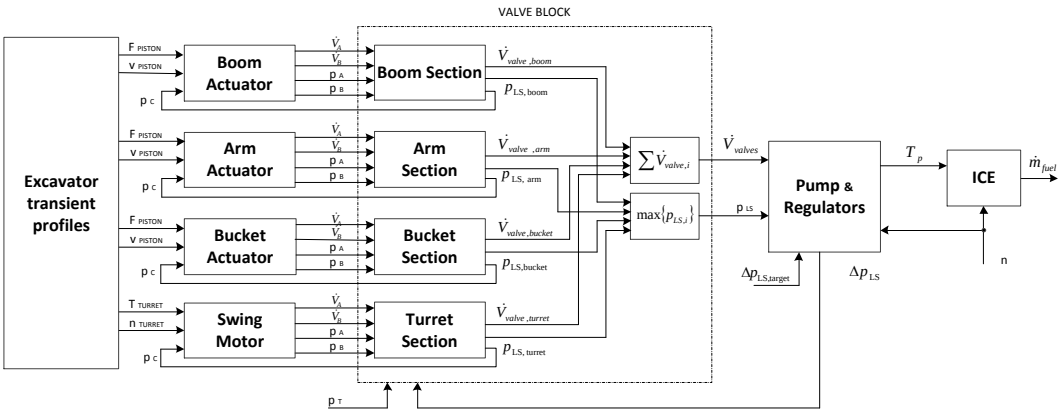
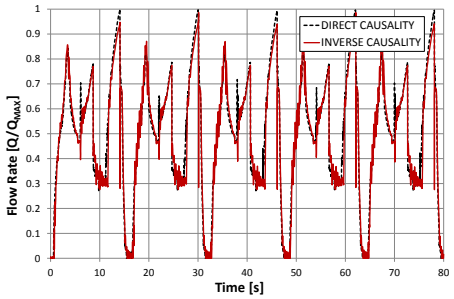
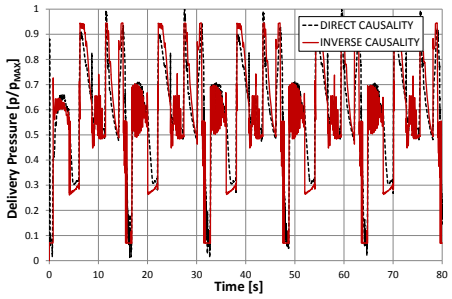


Figure 6.16: Comprehensive excavator reverse causality model

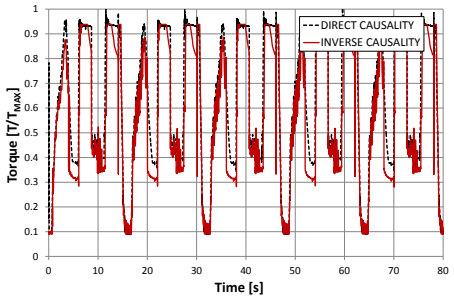
Before proceeding to the modelling of the hybrid configurations, the inverse causality model has been firstly used to simulate the JCMAS cycle with the standard excavator configuration and the results were compared to those from the direct model, to assess the validity of the simplifying assumptions. Figure 6.17 reports the comparison between the predictions from the direct and inverse causality models regarding pump flow rate, pump pressure, engine torque, and fuel consumption. Despite the quasi-steady simplification, the overall estimated fuel consumption differs for less than 1% between the two simulations.



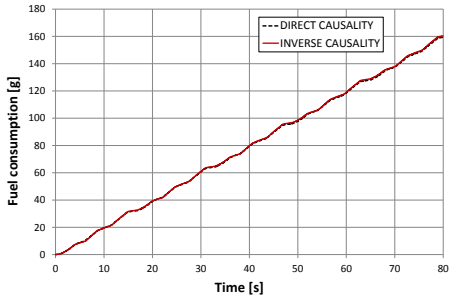
(a) Pump flow rate



(b) Pump delivery pressure



(c) Engine torque



(d) Fuel consumption

Figure 6.17: Direct and reverse causality models comparison

6.4 Hybrid layouts

The four conceived hybrid hydraulic configurations for energy recovery are presented in this section. While many different combination and solutions may be developed for the depicted layout, the present investigation focuses on solutions exploiting the energy recovery from the boom and the turret. As these components have the higher inertias, they are the natural candidates for the implementation of an energy recovery mechanism. All the four proposed layouts use a bladder type hydraulic accumulator which intercepts the flow directed towards the tank from the boom or the turret, and stores the available hydraulic energy. A fixed displacement external gear hydraulic motor is then equipped in a series configuration with the pump. This motor uses the pressurized fluid in the accumulator to generate a torque which relieves the load request to the ICE. Both the accumulator and the motor are simulated inside the direct causality model by means of built-in components sub-models from the AMESim®library. A brief description of the reverse causality recovery system sub-models is instead given below.

Hydraulic accumulator The hydraulic bladder type accumulator is used to recover the energy from the lowering phase of the boom and the braking of the turret. The accumulator is pre-charged with gaseous nitrogen, whose compression is simulated using the adiabatic ideal gas law:

$$p_g \cdot V_g^{\gamma} = \text{const} \tag{6.29}$$

with V_g the volume of the gas in the accumulator, p_g its pressure, and γ the adiabatic index for an ideal diatomic gas ($\gamma = 1.4$).

The following simplifying hypothesis are assumed about the conditions in the accumulator:

- the gas and the hydraulic fluid are always under quasi-steady mechanical equilibrium conditions, i.e. $p_{acc} = p_g$, with p_{acc} the pressure of the hydraulic fluid inside the accumulator;
- the total volume of the accumulator, V_{tot}, is occupied by the nitrogen gas and the hydraulic fluid, i.e. $V_{tot} = V_g + V_{hyd}$;
- from mass conservation equation for the hydraulic fluid, the rate of change of the fluid contained inside the accumulator is equal to the net incoming flow, i.e. $\frac{dV_{hyd}}{dt} = \dot{V}_{acc,in}$;

- the accumulator is adiabatic, i.e. no heat transfer towards the environment is accounted.

Thanks to the hypothesis above, it is possible to differentiate Equation 6.29 and obtain the following dynamic model for the hydraulic accumulator:

$$\frac{dp_{acc}}{dt} = \gamma \cdot \frac{p_{acc}}{V_{tot} - V_{hyd}} \cdot \dot{V}_{acc,in} \tag{6.30}$$

Hydraulic motor The hydraulic motor model relies on the basic quasi-steady approach for simulating fixed displacement machines. Both the volumetric and mechanical hydraulic efficiencies of the machine are in this case assumed constant. The flow rate elaborated by the machine $\dot{V}_m$ and the torque produced T_m are equal to:

$$\dot{V}_m = \eta_v \cdot V_d \cdot \frac{n}{60} \tag{6.31}$$

$$T_m = \frac{\Delta p \cdot V_d}{\eta_{mh} \cdot 2\pi} \tag{6.32}$$

6.4.1 Hybrid layouts configurations

A brief description of the four layouts, named respectively A, B, C, and D, is reported below, along with a schematic representation of the associated inverse causality model. Table 6.1 reports a summary of the main aspects of each layout.

Table 6.1: Energy recovery layouts

Configuration name	**Actuators used for energy recovery**	**Number of accumulators**
A	Boom	1
B	Turret	1
C	Boom, Turret	1
D	Boom, Turret	2

Hybrid layout A The first hybrid layout, whose ISO hydraulic schematic layout is depicted in Figure 6.18, is conceived to recover energy from the boom lowering phase. A digital ON/OFF control valve $\mathbf{u_1}$ is used to eventually bypass the line from boom outlet to tank towards the hydraulic accumulator. Another digital ON/OFF valve $\mathbf{u_2}$ is used to feed the hydraulic motor with the pressurized fluid. Finally, a pressure relief valve is also introduced to limit the maximum pressure in the hydraulic storage.

To decouple the boom discharge port from the pressure inside the accumulator and retain the controllability on the lowering phase, a variable flow control orifice (VCO) is placed between the tool and the accumulator. The VCO spool displacement is proportional to the main spool position $\mathbf{x_{boom}}$ in the boom control valve section.

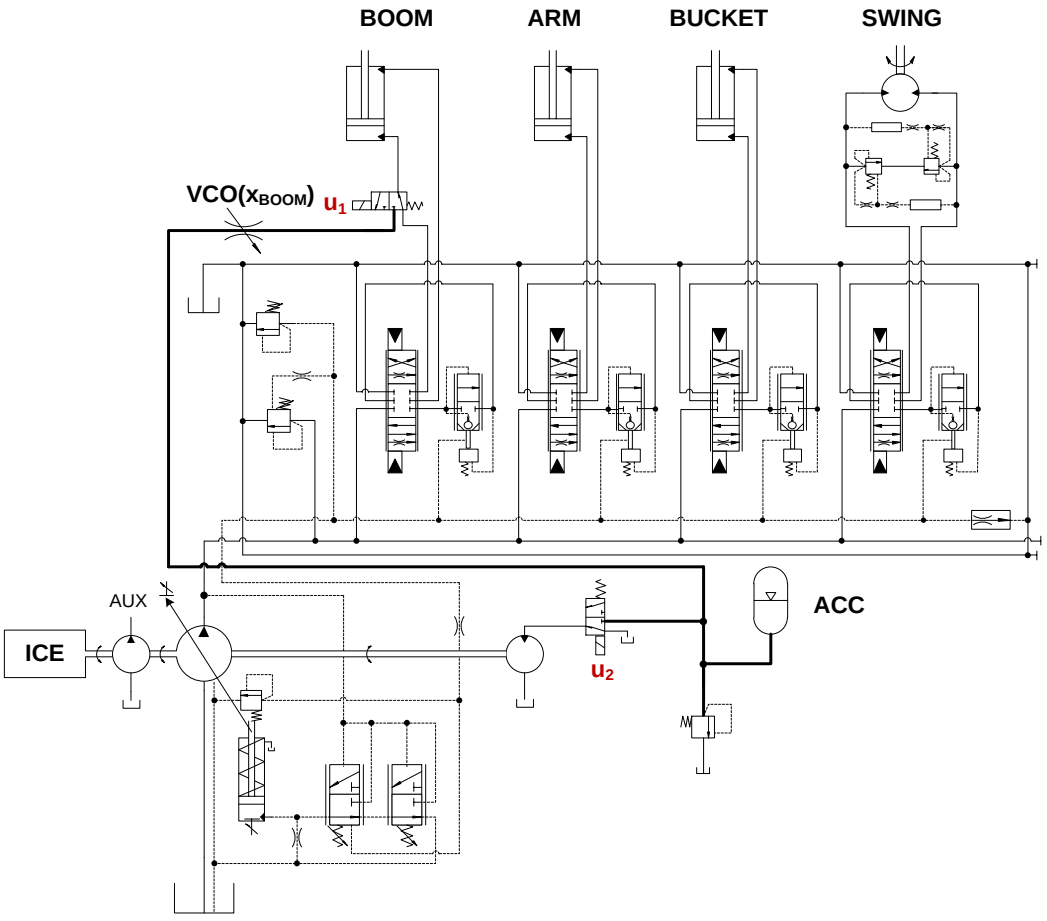


Figure 6.18: ISO schematic for hybrid layout A

The reverse causality model for the hybrid layout A is based on the effect that the two ON/OFF valves have on the fluid flowing towards and from the accumulator. If $\mathbf{u_1} = 0$, the flow from the boom discharge port (from now on associated to port A) is directed to the tank and the usual basic model applies. If on the other hand the valve is open ($\mathbf{u_1} = 1$) the flow $\dot{V}_A$ is directed to the accumulator and a different counter-pressure p'_C is present at boom port A, defined by the VCO characteristic as follows:

$$p'_C = p_{acc} + \frac{\dot{V}_A \cdot \rho}{2 \cdot c_d^2 \cdot A_{VCO}^2(x_{boom}, d_{EQ})} \tag{6.33}$$

where A_{VCO} is the VCO flow area which depends on the spool position x_{boom} and the VCO equivalent diameter d_{EQ}.

The second valve controls the flow from the accumulator to the motor, equal to zero if the valve is OFF ($\mathbf{u_2} = 0$), and equal to the flow $\dot{V}_m$ from Equation 6.32 if the valve is ON ($\mathbf{u_2} = 1$). Moreover, when $\mathbf{u_2} = 0$, the pressure difference across the motor is null and no torque is produced, while if $\mathbf{u_2} = 1$ the pressure difference is $\Delta p = p_{acc} - p_T$ and the torque generated can be evaluated by means of Equation 6.32. The resulting reverse causality model for hybrid configuration A is depicted in Figure 6.19.

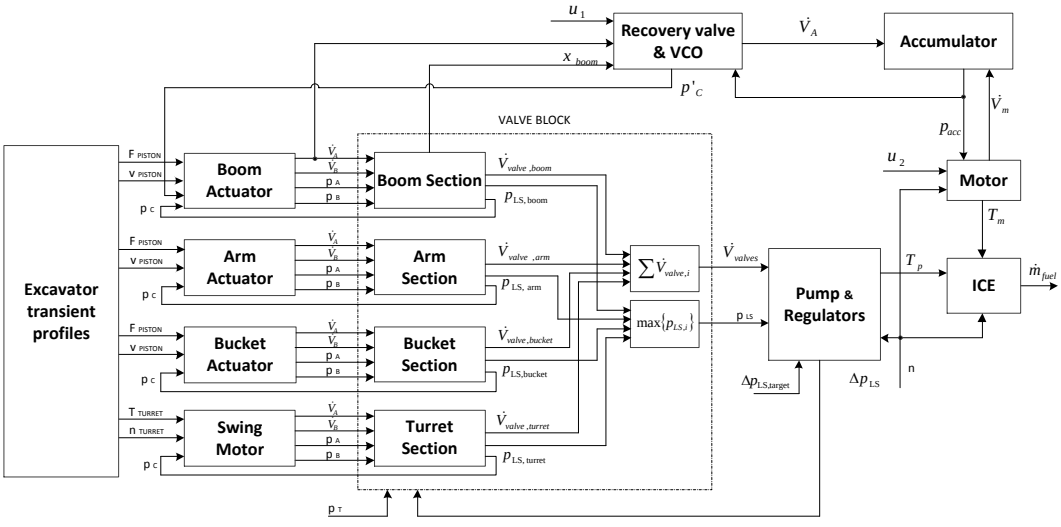


Figure 6.19: Hybrid layout A reverse causality model

Hybrid layout B The second hybrid layout is conceived to recover energy from the swing braking phase. In this case there is no need for an active control of the recovery phase as the procedure can be solely based on the position of the turret spool $\mathbf{x_{turret}}$. A digital ON/OFF control valve $\mathbf{u_1}$ is still present to manage the energy reutilization process by controlling the flow from the accumulator to the hydraulic motor. The ISO schematic of hybrid layout B is depicted in Figure 6.20.

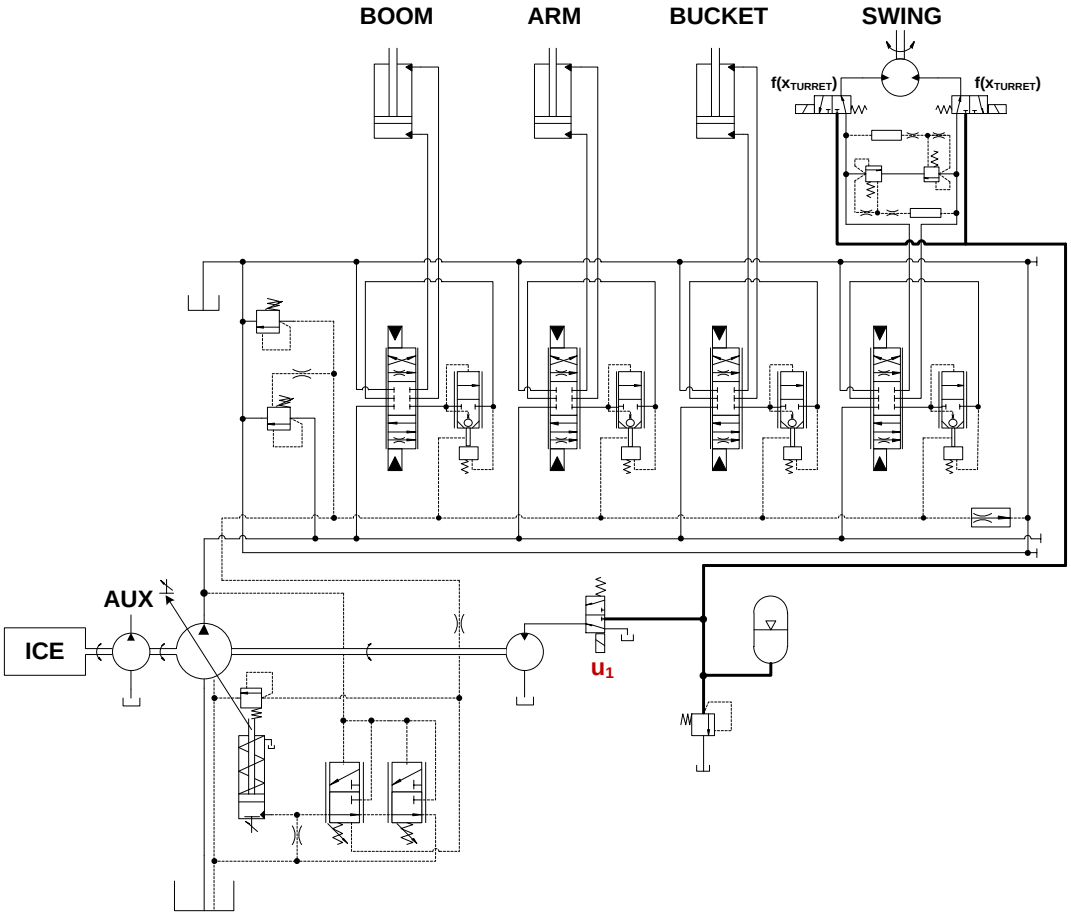


Figure 6.20: ISO schematic for hybrid layout B

The schematic of the reverse causality model for the simulation of hybrid layout B is depicted in Figure 6.21.

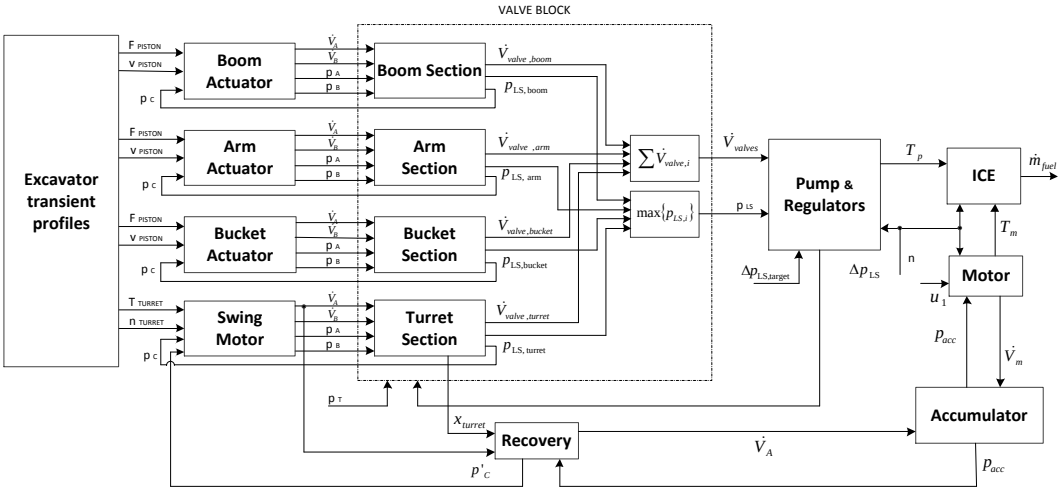


Figure 6.21: Hybrid layout B reverse causality model

Hybrid layout C The hybrid layout proposed in configuration C is a combination of layouts A and B, conceived to accumulate the energy recovered from the boom lowering and the turret braking into a single accumulator. A digital ON/OFF valve $\mathbf{u_1}$ controls the flow from the boom, while the recovery from the turret is automated. A second digital ON/OFF valve $\mathbf{u_2}$ manages the operation of the hydraulic motor. Figure 6.22 depicts the ISO schematic of the hydraulic circuit used in layout C.

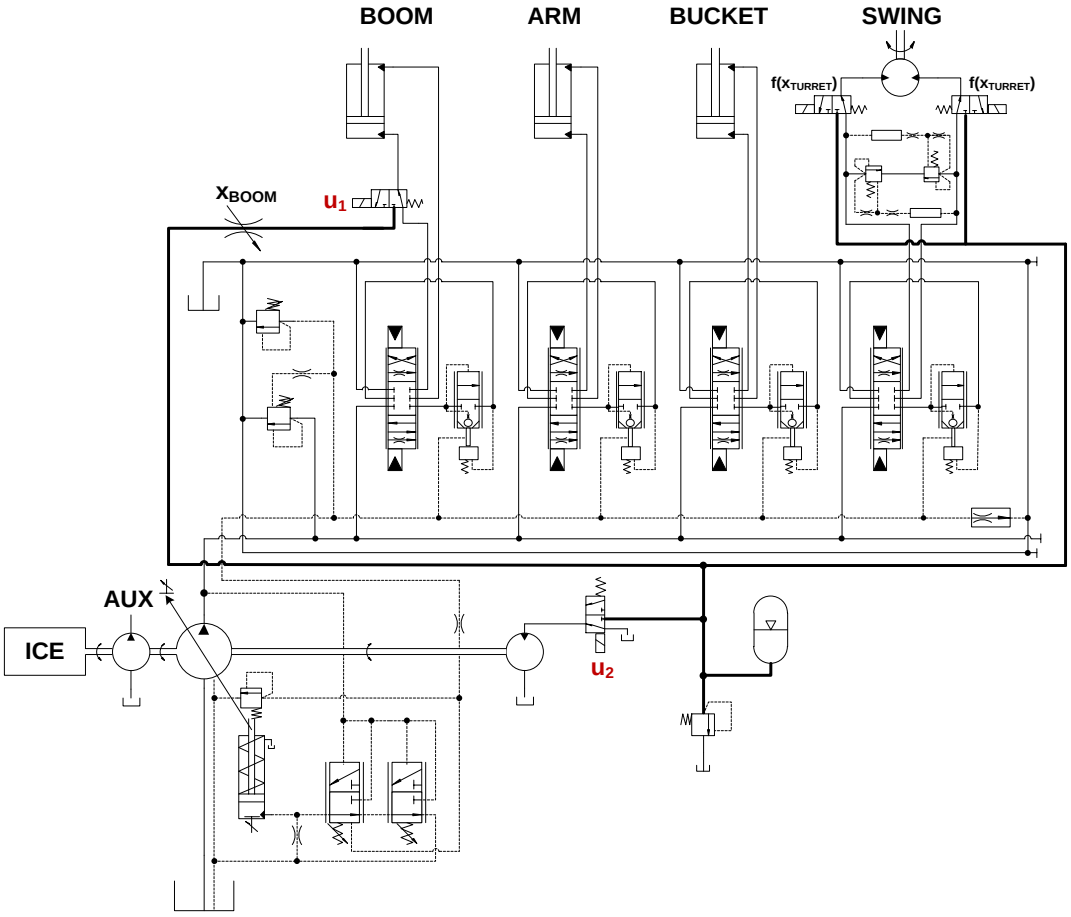


Figure 6.22: ISO schematic for hybrid layout C

Hybrid layout D As for the hybrid layout C, the layout D is conceived as a combination of the first ones. In this case however, two accumulators are adopted, one for each actuator. Four different control actuators are now necessary: the first one, $\mathbf{u_1}$ controls the recovery from the boom actuator; the second one, $\mathbf{u_2}$, controls the access to the input port of the hydraulic motor; the third and fourth ones, $\mathbf{u_3}$ and $\mathbf{u_4}$, controls the connection between valve $\mathbf{u_2}$ and the two accumulators. The ISO schematic of this solution is depicted in Figure 6.23.

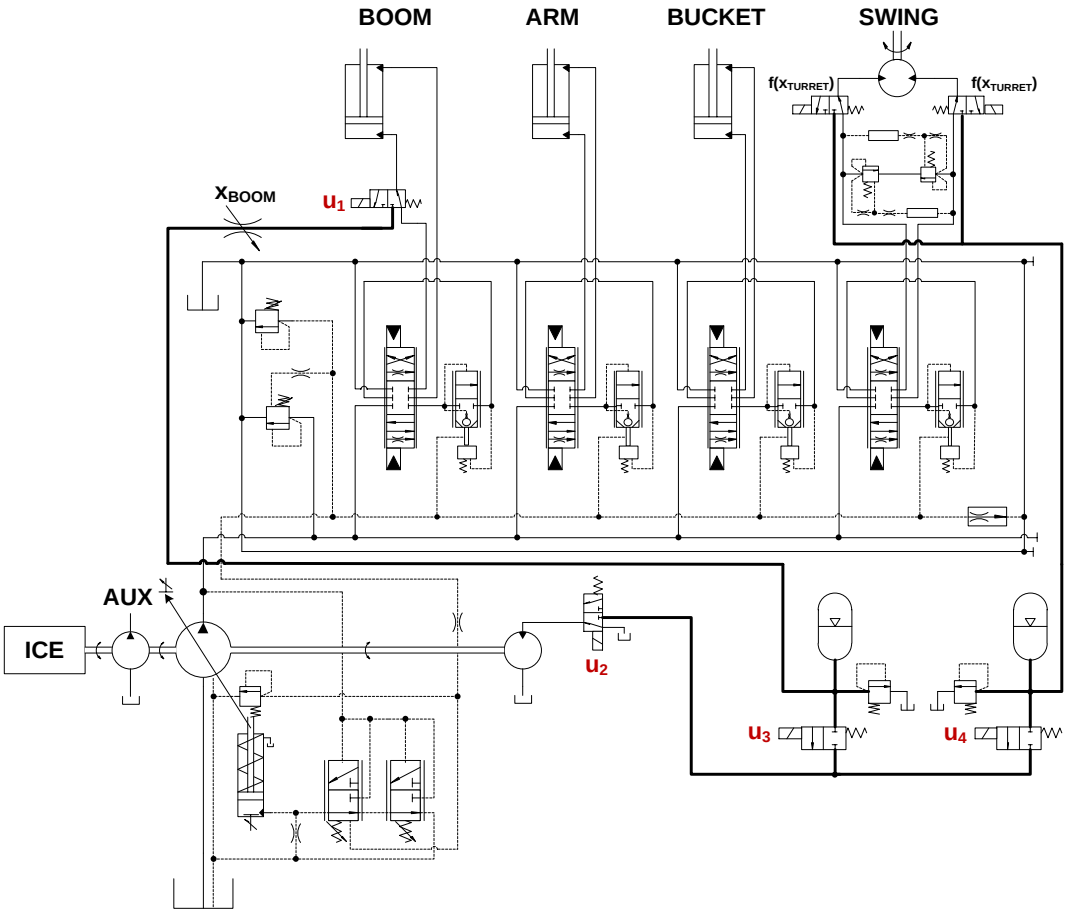


Figure 6.23: ISO schematic for hybrid layout D

6.4.2 Hybrid layout sizing

Parallel to the determination of the optimal layout configuration, a comprehensive comparison would also require the proper sizing of the energy recovery components introduced. Three parameters affecting the behaviour of the hybridized solution have been identified, namely:

- the volume of the bladder accumulator V_{tot};
- the minimum working pressure for the accumulator, p_{min}, equal to the pre-charge pressure;
- the equivalent hydraulic diameter d_{EQ} of the VCO.

The sizing of these components is carried out by means of a DoE (Design of Experiment) procedure, which will allow the identification of the optimal combination of these parameters. The sizing optimization is carried out for layouts A and B, and the results obtained for these configurations will guide the sizing for layouts C and D. The search ranges for each of the three parameters are summarized in Table 6.2.

Table 6.2: Range for the DoE parameters based on the hybrid configuration

		Layout	
		A	B
Parameter	V_{tot} $[L]$	$[2.5, 4, 5, 6, 10]$	$[2.5, 4, 5, 6, 10]$
	p_{acc} $[bar]$	$[10 : 5 : 50]$	$[10 : 10 : 100]$
	d_{EQ} $[mm]$	$[1 : 0.5 : 8]$	[——]

6.5 Dynamic Programming control algorithm

The development of the hybrid layouts for the energy recovery procedure introduced some alternative degrees of freedom to the problem, i.e. the trajectories of the digital valve controls u_i during the working cycle. The management of these inputs greatly affects the performance of the excavator with respect to the achieved fuel consumption reduction, as they actively control the process of energy recovery and reutilization. It is clear now why, to obtain a fair comparison between the four different layouts,

the knowledge of the optimal valve control trajectories is essential. To this extend, the reference maximum fuel consumption reduction connected to each layout and size configuration is estimated based on the optimal digital valve trajectories u_i^* obtained by means of the Dynamic Programming algorithm.

To apply the deterministic DP procedure, the reverse causality equations describing the different excavator layouts are first discretized with a sample time Δt of 0.01 s. Then, these equations are easily rearranged in the form the following state space representation:

$$\begin{cases} x(k+1) = f\big(x(k), u(k), w(k)\big) \\ y(k) = g\big(x(k), u(k), w(k)\big) \end{cases} \tag{6.34}$$

with the vector x representing the state of the system, i.e. the hydraulic pressures p_{acc} in the hydraulic accumulators, whose dynamics are determined by Equation 6.30; the vector u representing the control inputs necessary to manage the recovery process, i.e. the positions of the digital valves described in Section 6.4; w representing the exogenous disturbance to the system, i.e. the profiles of actuators forces, positions, and velocities from the JCMAS working cycle; and finally, y representing a generic combination of system parameters that can be derived from the system equations, e.g. the engine fuel consumption $\dot{m}_{fuel}$, the engine torque T_{eng}, the pump delivery pressure p_D, etc.

Under the hypothesis that the introduction of the generic recovery subsystem does not affect the drivability performance of the excavator, it is possible to assume that, even with the hybrid configurations, the speed and position profiles at the actuators will remain those depicted in Figures 6.10 and 6.11, obtained with the basic excavator configuration. With this assumption, it is possible to efficiently implement the algorithm for the deterministic DP optimization depicted in Chapter 4.2.2.

An appropriate objective function J is defined for the sake of the optimization process:

$$J(\pi, x_0) = \sum_{k=1}^{N-1} \dot{m}_{fuel}\Big(x(k), u(k), w(k)\Big) \cdot \Delta t \tag{6.35}$$

accounting for the overall fuel consumption during the entire JCMAS working cycle. The objective function depends on the initial accumulator pressure x_0 and the generic control policy $\pi = \{u_{t_0}, u_{t_1}, ..., u_{t_k}, ...u_{t_{N-1}}\}$, which determines the digital valves position a the discrete times $t_k = k\Delta t$ with $k \in \mathbb{N}$.

To account for the physical limitations of the system, the optimization is also

subjected to a state constraint on the accumulator pressure, confined in the range $[p_{min}, p_{max}]$ (with the minimum pressure defined by the DoE analysis and the maximum pressure $p_{max} = 4\ p_{min}$) and an output constraint on the ICE torque, whose feasible range is defined accordingly to manufacturers specifications.

$$x(k) \in [x_{min}, x_{max}] \tag{6.36}$$

$$T_{eng}(k) \in [T_{eng,min}, T_{eng,max}] \tag{6.37}$$

The DP algorithm implementation requires a discretization of both the inputs u and the states x. As all the control valves are digital, the inputs are already discretized as they can assume only assume the values 0 and 1. Accumulator pressure on the other hand is discretized over an evenly spaced grid of 50 elements ranging from the minimum accumulator pressure p_{min} to the maximum pressure p_{max}.

For the various proposed layouts and for every combination of the design parameters from the DoE ($p_{acc,min}, V_{tot}, d_{EQ}$), the initial accumulator pressure is chosen equal to the minimum pre-charge pressure; no constraints are set on the value of the final accumulator pressure p_N.

With these assumption, the DP algorithm is able to provide the optimal control policy π^*:

$$\pi^* = \arg\min_{\pi} J(\pi, x0) \tag{6.38}$$

The associated value of overall fuel consumption $m_{fuel,opt} = J(\pi^*, x_0)$ is used to evaluate the performance of the corresponding hybrid layout configuration.

6.6 Layouts comparison

Given the optimal management strategies derived from the application of the DP algorithm, the following results have been obtained for the different hybrid configurations during the digging and loading phase of the JCMAS cycle.

Layout A Regarding the performance of hybrid layout A, the overall improvements on fuel consumption are depicted in Figure 6.24, based on the DoE sizing parameters. The first thing to be noticed is the influence of the VOC hydraulic d_{EQ} on the overall achievable fuel consumption reduction: while with high valve diameters the effect is only marginal, it exists a minimum diameter under which the performance of the energy recovery system are greatly penalized. This is due to the fact that the

VCO valve, in order to maintain the controllability of the boom, has to throttle the fluid, creating a counter-pressure in the corresponding actuator chamber and thus increasing the level of pump delivery pressure that must be maintained to execute the movements. If this counter-pressure is too high, the recovery procedure loses great part of its effectiveness. Focusing on accumulator volume, it is noticeable that the parameter has only a marginal effect on the overall fuel consumption. Finally, when the VOC diameter is higher than the threshold value, an increase in the accumulator pre-charge pressure allows for large increases in fuel consumption reduction, up to +5%.

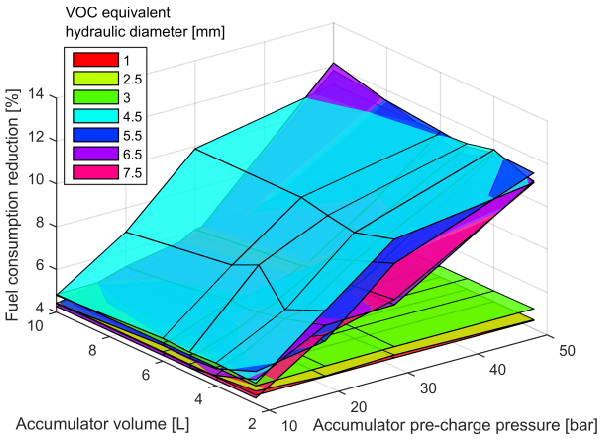


Figure 6.24: Fuel consumption improvements based on DoE parameters for hybrid layout A

The optimal fuel consumption performance for layout A are obtained with an accumulator volume V_{tot} of 10 L, a pre-charge pressure p_{min} of 50 bar, and an equivalent diameter of the VCO valve d_{EQ} equal to 6.5 mm. Figure 6.25 depicts the profile of the accumulator pressure resulting from the optimal management strategy evaluated from the DP algorithm. It is noticeable that, if the VCO valve is correctly sized, it is always convenient to activate the energy recovery during the boom lowering phase (boom velocity lower than zero). Moreover, it is also rewarding, from an energetic point of view, to always reach the maximum accumulator pressure and avoid, during the energy reuse, to completely deplete the storage. Unfortunately, while the recovery procedure shows periodic characteristics, it is not possible to correlate the energy reuse strategy to any particular phase of the JCMAS cycle.

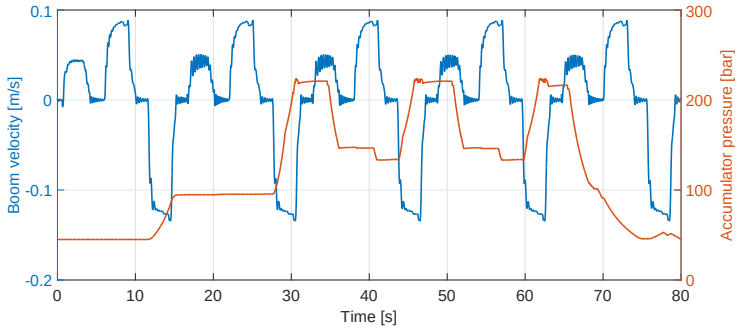


Figure 6.25: Accumulator pressure for optimal hybrid A management solution - digging cycle

Layout B The overall improvements on fuel consumption for layout B are depicted in Figure 6.24, based on the effect of the DoE sizing parameters p_{min} and V_{tot}. Similarly to layout A, the increase of accumulator pre-charge pressure p_{min} allows for a significant increase in fuel consumption reduction, although the effect, +1.5%, is smaller if compared to layout A. The correct choice of the accumulator volume V_{tot}, requires that the component is chosen as little as possible, allowing for an additional increase in fuel economy of about +0.5%. The maximum performances are obtained with a recuperator volume V_{tot} of 2.5 L and an accumulator pre-charge pressure p_{min} of 100 bar.

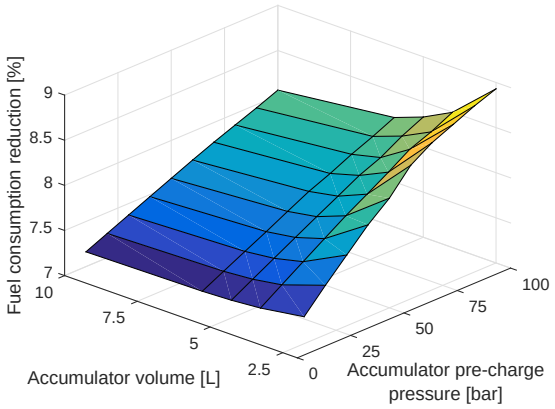


Figure 6.26: Fuel consumption improvements based on DoE parameters for hybrid layout B

Figure 6.27 depicts the profile of accumulator pressure for the layout B, once the DP optimal policy is applied. Due to the configuration layout, the energy that would normally be dissipated by the braking of the turret is always stored in the hydraulic accumulator. The only task of the controller is that of deciding when to deplete the accumulator to reuse that stored energy. While even in this case it is not possible to derive a simple rule based strategy correlating the action of the optimal DP control strategy to some particular phase of the digging cycle, it is still worth noticing that the best strategy involves that the accumulator must be completely filled before attempting to use the stored energy, even if in this particular case the filling process takes more than half of the entire digging process, due to the relatively low amount of energy that can be recovered from turret braking. Moreover, once the accumulator is completely filled and the energy stored can be reused, it is still worth not to excessively decrease its pressure.

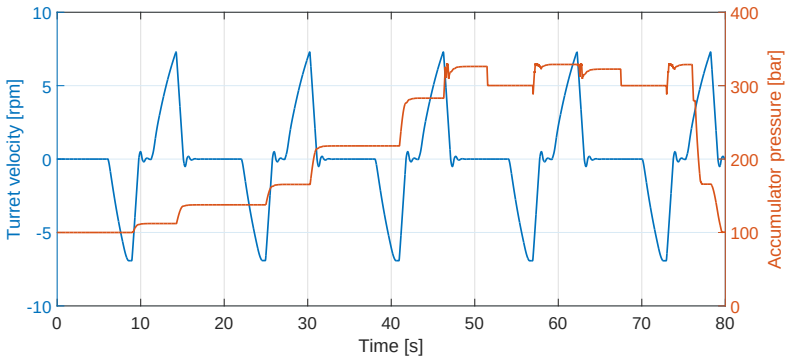


Figure 6.27: Accumulator pressure for optimal hybrid B management solution - digging cycle

Layout C In case of hybrid layout C configuration, the accumulator volume is chosen to be equal to 2 L due to the large improvements that this solution involves for layout B, while having only a marginal effect on the optimality of solution A. The accumulator pre-charge pressure is set to the maximum possible (100 bar) and the VOC optimal valve diameter from layout A is adopted. Figure 6.28 depicts the optimal trajectory of the accumulator pressure for layout C. Even for this excavator

configuration, the strategy of completely fill and the marginally deplete the accumulator proves to be the optimal way to control the energy recovery process. The simultaneous recovery from two sources (boom lowering and turret braking) allows for a more rapid refilling of the accumulator.

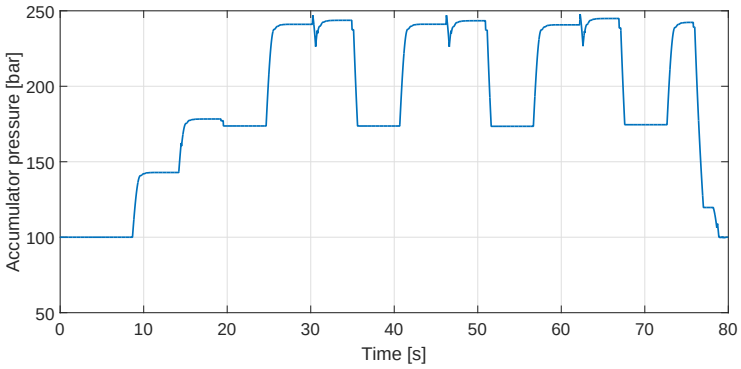


Figure 6.28: Accumulator pressure for optimal hybrid C management solution - digging cycle

Layout D With the adoption of layout D, two different accumulators are employed, for decoupling the energy recovery procedure from the two users. Each accumulator is designed based on the optimal corresponding single user layout configuration. The trajectories of optimal accumulator pressure are close to those obtained from the single actuator/accumulator cases (layout A and B).

6.7 Results

The effective reduction of fuel consumption obtained from the application of the hybrid recovery optimal strategies for the entire JCMAS cycle are reported in Table 6.3. The first thing that can be noticed is that the recovery from the turret alone produces very low decreases in the overall excavator fuel consumption. For layout B in fact, even if the recovery procedure is optimized, the energy that can be recovered from turret braking phase is relatively low, due to the reduced velocities of the component during its rotation. Moreover, as the turret is moved only during the digging and loading phase, the effectiveness of the energy saving system is greatly reduced as

no energy recovery is possible during the levelling phase. The simultaneous recovery from both users adopted in layout C is able to guarantee just a marginal increase in fuel savings of about + 1-2 % compared to the single recovery from the boom actuator. The decoupled strategy adopted in layout D on the other hand involves greater reductions of about + 3-4 %, still if compared to the boom alone recovery strategy from layout A.

Table 6.3: Hybrid layouts fuel saving percentage

Hybrid layout	Actuators used for recovery	# of accumulators	Fuel savings
A	Boom	1	10-11 %
B	Turret	1	2-3 %
C	Boom-Turret	1	11-12 %
D	Boom-Turret	2	13-14 %

From a purely energetic point of view, layout D is obviously the best solution for the recovery of wasted energy and fuel consumption reduction among the various layouts proposed in this study. From a more practical point of view, it is to be noted that solution D requires the adoption a complicated circuital layout and the introduction of many additional components. While layout B can be easily ignored due to the significantly low amount of fuel savings potential correlated to the adoption of that particular configuration, a more complete analysis of the problem is required to effectively decide which of the remaining three solutions is worth to be adopted for the implementation on the machine, taking into account the economic aspects regarding the introduction of the additional components.

6.8 Conclusions

The study presented in this chapter focused on the design and optimization of a hydraulic hybrid solution for the energy recovery and fuel consumption reduction on a middle size excavator application. Four different circuital designs have been analyzed and sized to assess the best achievable performance. With the introduction of an energy recovery system, the number of degrees of freedom of the problem is augmented and the action of the control actuators responsible for the management of the

recovery procedure directly affect the fuel savings performance of the corresponding solution. To operate a fair comparison between the different layouts, a methodology to assess the certified maximum achievable performance of each hybrid solution has been developed, relying on the estimation of the optimal energy management control policy by means of the deterministic Dynamic Programming algorithm. The methodology relies on a simplified lumped parameter model of the excavator which can be used to evaluate the dynamics of the energy recovery process and the fuel consumption reduction over a regulated standard working cycle. Once the optimal strategy is evaluated, it is possible to associate a benchmark value for fuel savings capabilities to each configuration, thus providing an effective methodology for the comparison of the different solutions and the sizing of the newly introduced circuital components.

Nomenclature

Subscripts and superscripts

A, B	hydraulic actuator ports
acc	hydraulic accumulator
act	hydraulic actuator
C	hydraulic port counter-pressure
D	pump delivery section
eng	engine
EQ	equivalent
g	gas
h	hydraulic fluid
LIM	torque limiter
m	hydraulic motor
mh	mechanical-hydraulic
p	pump
T	tank
v	volumetric
0	initial

Abbreviations and acronyms

DoE	Design of Experiments
DP	Dynamic Programming
FC	Flow Compensator
FDM	Fluid-Dynamic Model
ICE	Internal Combustion Engine
JCMAS	Japanese Construction Mechanization Association Standard
LS	Load-Sensing
MGM	Mechanical-Geometrical Model
PI	Proportional Integral
PC	Pressure Compensator
TL	Torque Limiter compensator
VCO	Variable flow Control Orifice

Symbols

A	port section	$[m^2]$	T	torque	$[Nm]$
c	viscous friction coeff.	$[Nm \cdot s]$	u	control input	$[-]$
c_d	discharge coefficient	$[-]$	$\dot{V}$	volumetric flow rate	$[^{m^3}/_s]$
d	diameter	$[m]$	v	linear velocity	$[^m/_s]$
F	force	$[N]$	V_d	machine displacement	$[^{m^3}/_{rev}]$
J	rotational inertia	$[kg \cdot m^2]$	x	spool position	$[m]$
	objective function	$[-]$		system state	$[-]$
n	rotational speed	$[^{rev}/_{min}]$	y	actuator position	$[m]$
p	pressure	$[Pa]$		system output	$[-]$

w	disturbance	$[-]$
y	output	$[-]$
α	swash plate angular position	$[rad]$
γ	adiabatic index	$[-]$
Δp	pressure difference	$[Pa]$
Δt	sample time	$[s]$
η	efficiency	$[-]$
π	control policy	$[-]$
ρ	density	$[{}^{kg}/_{m^3}]$
θ	angular position	$[rad]$
ω	angular velocity	$[{}^{rad}/_{s}]$
Ω	metering area	$[m^2]$

Bibliography

[1] F. Bender, M. Kaszynski, and O. Sawodny, "Drive Cycle Prediction and Energy Management Optimization for Hybrid Hydraulic Vehicles," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 8, pp. 3581–3592, Oct. 2013. 104

[2] F. A. Bender, M. Kaszynski, and O. Sawodny, "Location-based energy management optimization for hybrid hydraulic vehicles," in *2013 American Control Conference*. IEEE, Jun. 2013, pp. 402–407. 104

[3] P. Casoli and A. Anthony, "Gray box modeling of an excavators variable displacement hydraulic pump for fast simulation of excavation cycles," *Control Engineering Practice*, vol. 21, no. 4, pp. 483–494, Apr. 2013. 108, 109, 113

[4] P. Casoli, A. Anthony, and L. Ricco, "Modeling Simulation and Experimental Verification of an Excavator Hydraulic System - Load Sensing Flow Sharing Valve Model," Sep. 2012. 108, 112

[5] P. Casoli, A. Gambarotta, N. Pompini, and L. Ricco, "Coupling excavator hydraulic system and internal combustion engine models for the real-time simulation," *Control Engineering Practice*, vol. 41, pp. 26–37, Aug. 2015. 108

[6] P. Casoli, N. Pompini, and L. Riccò, "Simulation of an Excavator Hydraulic System Using Nonlinear Mathematical Models," *Strojniški vestnik - Journal of Mechanical Engineering*, vol. 61, no. 11, pp. 583–593, Oct. 2015. 108, 110

[7] N. Daher and M. Ivantysynova, "Energy analysis of an original steering technology that saves fuel and boosts efficiency," *Energy Conversion and Management*, vol. 86, pp. 1059–1068, Oct. 2014. 103

[8] T. O. Deppen, A. G. Alleyne, K. Stelson, and J. Meyer, "An energy management strategy for a hydraulic hybrid vehicle," in *2012 American Control Conference (ACC)*. IEEE, Jun. 2012, pp. 1335–1341. 103

[9] M. Erkkilä, F. Bauer, and D. Feld, "Universal Energy Storage and Recovery System - A Novel Approach for Hydraulic Hybrid," in *The 13th Scandinavian International Conference on Fluid Power*, Linköping, Sweden, 2013, pp. 45–52. 103

[10] D. Feng, D. Huang, and D. Li, "Stochastic model predictive energy management for series hydraulic hybrid vehicle," in *2011 IEEE International Conference on Mechatronics and Automation*. IEEE, Aug. 2011, pp. 1980–1986. 104

[11] T. H. Ho and K. K. Ahn, "Design and control of a closed-loop hydraulic energy-regenerative system," *Automation in Construction*, vol. 22, pp. 444–458, Mar. 2012. 103

[12] S. Hui, Y. Lifu, J. Junqing, and L. Yanling, "Control strategy of hydraulic/electric synergy system in heavy hybrid vehicles," *Energy Conversion and Management*, vol. 52, no. 1, pp. 668–674, Jan. 2011. 103

[13] M. Inderelst, "Efficiency improvements in mobile hydraulic systems," Ph.D. dissertation, Aachen, Techn. Hochsch., 2013. 105

[14] M. Inderelst, S. Sgro, and H. Murrenhoff, "Energy recuperation in working hydraulics of excavators," in *22nd FPMC Symposium*, Bath, England, 2010. 103

[15] H. Kim, J. Choi, and K. Yi, "Development of supervisory control strategy for optimized fuel consumption of the compound hybrid excavator," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 226, no. 12, pp. 1652–1666, May 2012. 103

[16] T. Lin, Q. Wang, B. Hu, and W. Gong, "Research on the energy regeneration systems for hybrid hydraulic excavators," *Automation in Construction*, vol. 19, no. 8, pp. 1016–1026, Dec. 2010. 103

[17] M. Linjama and K. Huhtala, "Digital hydraulic power management system towards lossless hydraulics," in *The Third Workshop on Digital Fluid Power, October 13 - 14, 2010, Tampere, Finland*, 2010, pp. 5–22. 102

[18] A. Shenouda and W. Book, "Energy Saving Analysis Using a Four-Valve Independent Metering Configuration Controlling a Hydraulic Cylinder," Nov. 2005. 102

[19] T. Wang, Q. Wang, and T. Lin, "Improvement of boom control performance for hybrid hydraulic excavator with potential energy recovery," *Automation in Construction*, vol. 30, pp. 161–169, Mar. 2013. 103

[20] D. Wu, "Modeling and experimental evaluation of a load-sensing and pressure compensated hydraulic system," Ph.D. dissertation, University of Saskatchewan, 2004. 105

[21] Q. Xiao, Q. Wang, and Y. Zhang, "Control strategies of power system in hybrid hydraulic excavator," *Automation in Construction*, vol. 17, no. 4, pp. 361–367, May 2008. 103

[22] J. Zimmermann and M. Ivantysynova, "Reduction of engine and cooling power by displacement control," in *6th FPNI-PhD Symposium*. West Lafayette, USA: Fluid Power Net Publications (FPNI), 2010. 103

CHAPTER 7
DYNAMIC OPTIMIZATION OF A ORGANIC RANKINE CYCLE PLANT FOR LIGHT AUTOMOTIVE APPLICATION

Over the course of the years, regulations for CO_2 emissions for passengers cars and light duty trucks became tighter and tighter while the price of carbon-based fuels became subject to extreme uncertainities. Consequently, it is of great interest, for both the automotive manufacturers and the private users, the development of new methods for the increase of vehicles fuel economy. Many technologies has been developed during the years [11] for the reduction of fuel consumption; they are based on various approaches, such as the reduction of energy dissipations in the engine, the improvement of engine thermodynamics (HCCI [21], GDI [41], VVT [25], VVA [37]), the development of new strategies for engine thermal management (TMS [13], warm-up time reduction [36]), the improvement of the transmission system (Dual Clutch Transmission [53], CVT [33]), and the reduction of aerodynamic drag [12]. Many techniques have also been proposed for the harvesting of the kinetic energy which is irreversibly wasted during vehicle braking. These techniques are mainly based on the hybridization of the vehicle powertrain (electric [43, 44], mechanic [15], hydraulic [55]).

Even with the aforementioned technologies however, a considerable percentage of the energy produced on a vehicle is still dissipated as thermal energy through coolant and exhausts due to the physical limits of a thermodynamic process. With approximately 65 % [11] of the energy produced in the combustion chamber dissipated to the environment, the recuperation and harvesting of such wasted energy is becoming of increasing interest for the automotive industry.

The present chapter investigates a solution of Waste Heat Recovery (WHR) from light vehicle applications based on the usage of an Organic Rankine Cycle plant, a technology which is receiving increasing attention from the automotive industry. In the first sections, the basic layout of an ORC plant is presented and a suitable mathematical model for evaluating the performance of the plant is described. In the following section, the problem of the optimal energy management of the ORC

system is investigated under realistic driving condition, and a model based predictive controller is developed. The performance of this control strategy are finally compared to those deriving from a static optimization based algorithm and used to assess the possible improvements on fuel economy that this plant is expected to bring.

7.1 Waste Heat Recovery in vehicles

Although many improvements have been made on the increasing of combustion process efficiency, it became more and more difficult to achieve significant results in fuel economy improvements focusing only on in-cylinder processes. On the other hand, with a great amount of exergy still present at the exhausts side (with temperature ranging from 500 to 900 °C), the idea of harvesting this source of otherwise wasted energy has become very attractive. Three mechanism are actually implemented for Waste Heat Recovery, namely:

1. Turbocompounding;
2. Thermoelectrics;
3. Organic Rankine Cycles.

Turbocompounding Turbocompounding is based on the same principles of turbocharging (which to a certain extent may be intended as WHR strategy too), however, in the compound solution, the turbine is directly connected either to the engine itself or to an electric generator. Turbocompounding is mainly used in heavy duty trucks as it provides a cheap and reliable WHR solution [1, 59]. In 1991 Scania firstly commercialized a compounded system for its 6-cylinder 11 L DTC 1101 Diesel engine, providing a 1-3% fuel economy improvement at high loads actual driving conditions [23]. Scania and Cummins released together in 1995 another 12 L six-cylinder engine specially designed for turbocompounding [57]. Since then, many others manufacturers have integrated the turbocompounding technology in their high load applications, such as Caterpillar's 15 L 21st Century Truck engine in 1998 [58], Isuzu's Ceramic IDI in 2002 [58], Volvo's D12-500TC in 2002 [58], Daimler's Detroit Diesel DD15 in 2008 [57]. Nowadays, the option of generating electrical power rather than mechanically linking the turbine to the crankshaft is the main object of the research. The electric solutions offer the advantage that the speed of the turbine can be freely controlled, allowing the optimization of the power generation process, with

simulations results showing a 5-10 % reduction in fuel consumption based on the case of a C-15 Caterpillar medium duty truck engine [24].

Thermoelectrics Thermoelectric technologies use the Peltier effect to convert energy from a thermal gradient (in this case provided by the coolant and exhausts) directly to electrical energy [4]. Even if some applications for Waste Heat Recovery has been developed, they are considered impractical for vehicle applications due to the increased costs, the added system mass, the power consumption of the accessories, and the reduced efficiency (1-3%), [11].

Organic Rankine Cycles Organic Rankine Cycle (ORC) plants are a promising solution for Waste Heat Recovery which is based on the use of a closed thermodynamic cycle where the hot exhausts from the engine are employed as the heat source. High molecular weight organic refrigerants are used as a working fluid due to their low boiling temperature (when compared to water). This property allows for the recovering of heat even from low temperature sources while still maintaining high thermodynamic cycle efficiencies and reduced plant costs. Additional properties of organic fluids may also be inferred from the diagram depicted in Figure 7.1:

- the enthalpy difference between saturated liquid and saturated vapor is much smaller for organic fluids. This is clearly an advantage in heat recovery as it allows the complete vaporization of the fluid even whit low amounts of available energy;

- organic fluids usually present a positive slope vapor saturation line. This type of fluids (named dry fluids, opposed to wet fluids such as water) once vaporized do not produce liquid droplets during the expansion phase, a process which can be very harmful for the mechanical expanders used to extract work during this phase [40]. The adoption of dry fluids avoids therefore the need of reaching relatively high superheated vapor temperatures, necessary to guarantee the safety of the components in the wet fluid case, which sometimes cannot be achieved in WHR systems due to the low temperature of the heat source.

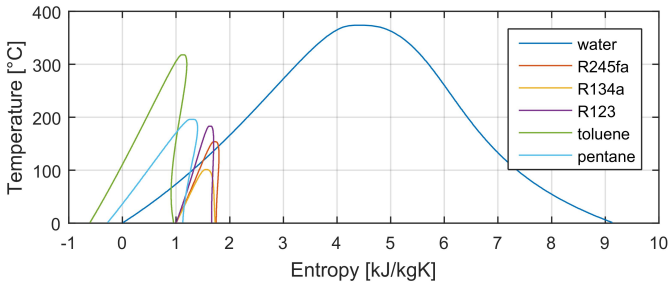


Figure 7.1: T-s diagram for a few organic fluids and water

A generic ORC plant is structured as depicted in Figure 7.2. Four basic components are required, namely a pump, an evaporator, an expander, and a condenser. A recuperator is a standard addition to the basic layout to increase thermodynamic efficiency. The thermodynamic cycle, depicted in Figure 7.3, is summarized as follows:

1-2: The working fluid, in liquid phase, is pressurized by the pump;
2-3: The working fluid is pre-heated in the recuperator;
3-4: Heat from an external source causes the isobaric evaporation of the fluid;
4-5: An expander extracts energy from the working fluid;
5-6: The fluid, still in vapor phase, release part of its energy to the liquid;
6-1: The working fluid is condensed and sub-cooled by means of a coolant.

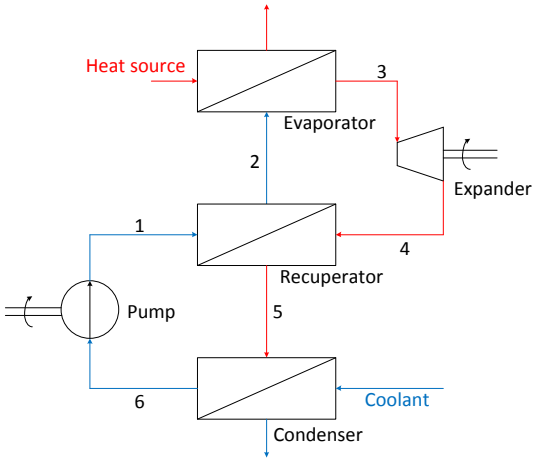


Figure 7.2: ORC plant basic layout

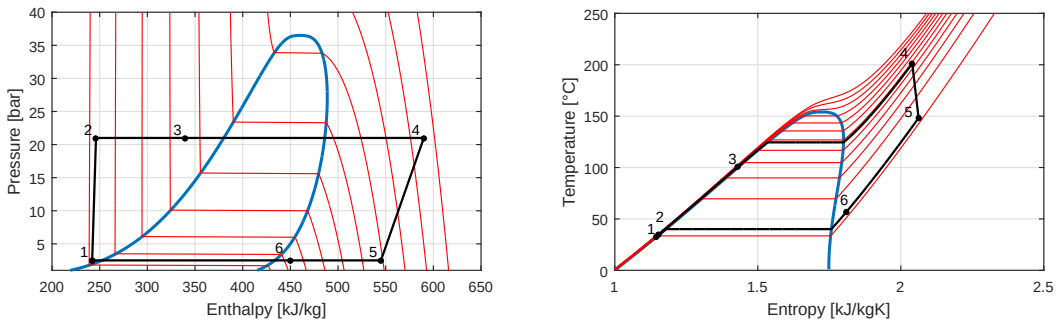


Figure 7.3: p-h and T-s diagrams for a generic ORC cycle

Organic Rankine Cycle plants are currently used in a large number of different industrial applications for the production of energy from low temperature sources, such as solar power plants [19], biomass combined heat and power production [56], geothermal plants [20], and as bottom cycle for various industrial processes involving a great amount of rejected heat (e.g. cement [32], steel [31], glass [9], food processing [3]).

In the automotive field, the first application of ORC technology can be found during the 1970's oil crisis, with the implementation of an ORC plant over a 288 HP Mack 676 truck diesel engine with a predicted 15% fuel economy improvement over a typical duty cycle [47]. After some of the first implementation, a long gap followed before the automotive industry showed again a real interest in Waste Heat Recovery via Rankine systems. Today, due to the renewed interest in high fuel economy and tightening emissions standards, the improving of technologies, and the increase in oil price, many manufacturers have investigated the implementation of Rankine cycle plants for both heavy applications (Cummins [46], Caterpillar [35], Daimler [54], Volvo [27], Renault [17]) as well as for light duty vehicles (BMW [52], Honda [16], Ford [28]), with achieved fuel consumption reductions ranging from 5 to 10 % depending on the system application.

7.2 Organic Rankine Cycle plant layout

The present study focuses on an ORC plant for exhaust waste heat recovery from a turbocharged spark-ignition (SI) engine for passenger car applications. The system layout is illustrated in Figure 7.4: a positive displacement pump is used to control the flow of the refrigerant through a heat exchanger where it receives heat from the hot

exhaust gases coming from the engine. The working fluid adopted in the applications is the refrigerant R245fa, used as it allows to obtain high efficiencies even with low heat transfer form the exhaust side [34]. A bypass valve is added on the exhaust side to eventually decrease the amount of gas entering the evaporator if it is necessary to limit the heat input to the system for safety conditions.

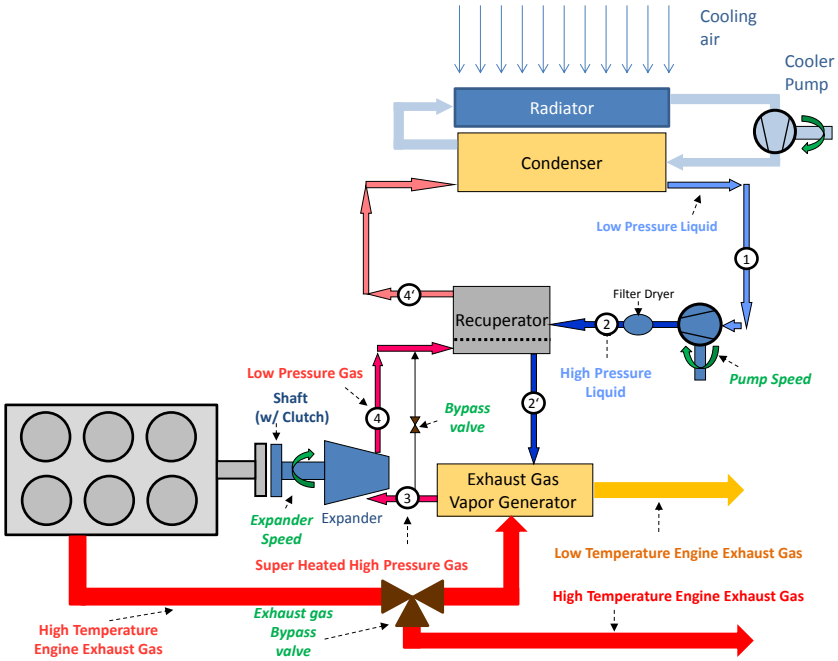


Figure 7.4: Schematic of the ORC plant

Once vaporized, the refrigerant flows through a scroll expander, thus generating mechanical work. It is to be noted that contrary to turbines, which can work at different pressure ratios, positive displacement expanders (such as a scroll one) have to work at a fixed pressure ratio, compensating the pressure variations with isochoric expansion/compression of the fluid at the outlet of the device, with a subsequent decrease in isentropic efficiency [60]. The use of a positive displacement machine for the power generation, though disadvantageous from an energetic point of view, is however justified by its reduced cost and increased robustness, making it attractive for small power applications. At the outlet of the expander, the fluid is directed to

a recuperator which preheats the fluid entering the evaporator, thus increasing cycle efficiency [7, 8].

Finally, the refrigerant runs through a condenser where a low-temperature cooling circuit is used to reject the excess of heat; the coolant flow rate is controlled by an electric pump to guarantee the desired condensing conditions. Ram air is considered sufficient for removing the heat from the coolant and providing the desired temperature at the inlet of the condenser, therefore no additional components are required.

While the ORC system is designed for nominal functioning at highway driving operating conditions, where the evaporator can always generate superheated vapour at its outlet, it may occur that during fast transients (e.g. cut-off conditions, low load, etc.), the heat available at the exhaust side may not be sufficient to obtain a full vaporization of the refrigerant. This condition, which may lead to unsafe behaviors of the expander, is avoided by inserting a liquid phase separator at the evaporator outlet, which eventually bypasses the saturated liquid directly after the expander.

For the application considered in this study, the scroll expander is directly coupled to the engine with a 1:1 gear ratio. This solution reduces the degrees of freedom of the control system, as the velocity of the expander is now coupled with that of the engine, but it eliminates the needs of auxiliary components (such as CVTs, variable slip clutches, or electric generators), which leads to a cost and volume reduction of the entire plant.

7.3 ORC dynamic model

In order to carry on the optimization of the plant and to design the control for the ORC plant, a numerical model for the system is required. A 0-D, lumped parameter physical representation of the plant and has been developed and improved during the years. The model accounts for the key components of the system, base on the representation depicted in Figure 7.5. A black box approach has been used to model the positive displacement machines (i.e. the pump and the expander). The recuperator model is based on a simple $\varepsilon - NTU$ approach while the condenser and the cooling circuit models are based on a quasi-static approach relying on energy conservation equation. The most difficult component to model was the evaporator, which dynamics greatly affect the performances of the plant. In order to describe the behavior of the two phase fluid in this heat exchanger, the Moving Boundary Method (MBM), a technique widely used for the simulation of ORC and Air Conditioning applications, has been used. Furthermore, to improve the model robustness, a Switching Algorithm

has been introduced to comply with non-nominal operating conditions of the evaporator, e.g. the incomplete evaporation of the organic fluid due to low heat inputs from the engine. The thermodynamic properties of the fluid and their derivatives are calculated using the software REFPROP®.

The comprehensive ORC model is able to predict the dynamic response the plant resulting from the effect of the disturbance from the engine (exhaust mass flow rate, exhaust temperature and expander speed) or the action of the control algorithm on the governable inputs (pump speed and exhaus bypass).

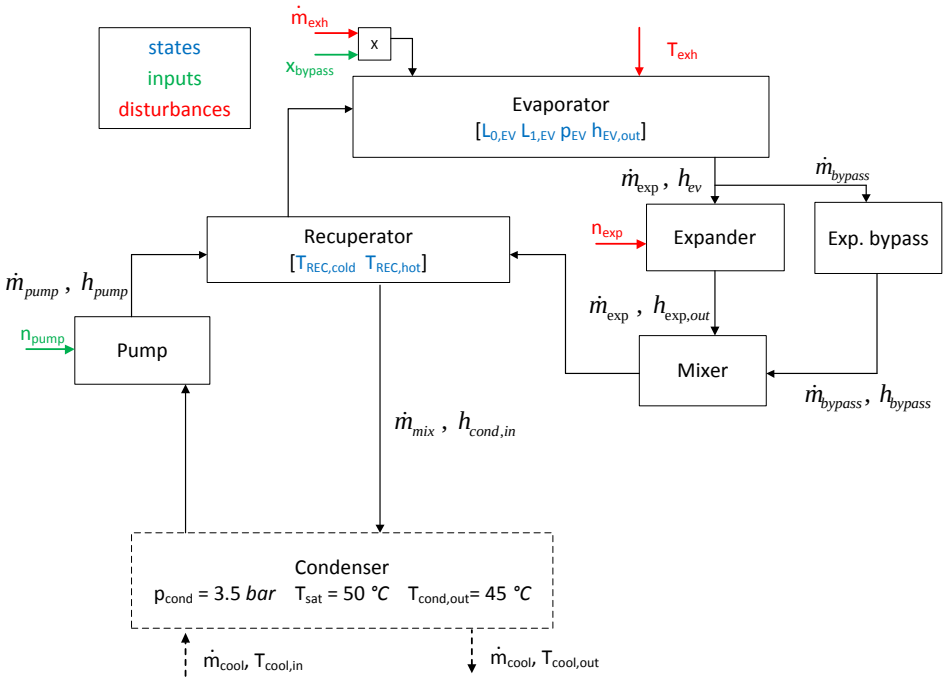


Figure 7.5: Schematic of the ORC plant

7.3.1 Pump and Expander

Pump and expander models, Figure 7.6, are based on a quasi-static, grey box methodology approach [8, 10], where a simple physical based model is integrated with experimental correlations.

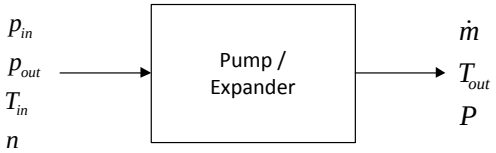


Figure 7.6: Volumetric machines IO scheme

For both of the positive displacement machines, the mass flow rate and the required/generated power can be defined as:

$$\dot{m} = \rho_{in}\ \eta_v(n, \Delta p) V_d\ \frac{n}{60} \tag{7.1}$$

$$P = \dot{m}\ \eta_s(n, \Delta p)(h_{in} - h_{out,s}) \tag{7.2}$$

where

$\dot{m}$ refrigerant mass flow rate;
ρ_{in} inlet fluid density;
η_v volumetric efficiency;
n rotational speed;
Δp inlet/outlet pressure difference;
V_d machine displacement;
P machine power;
η_s isentropic efficiency;
h_{in} inlet enthalpy;
$h_{out,s}$ outlet isentropic efficiency.

For both the pump and the expander, the correlations for the volumetric efficiencies are obtained using data acquired from an experimental campaign. In case of the isentropic efficiency (Equation 7.3b), only the expander is characterized experimentally, while a constant efficiency (0.75) is assumed for the pump.

$$\eta_v = \frac{\dot{m}_{real}}{\rho_{in} V_d\ \frac{n}{60}} \tag{7.3a}$$

$$\eta_s = \frac{h_{in} - h_{out,real}}{h_{in} - h_{out,s}} \tag{7.3b}$$

7.3.2 Expander bypass

The expander bypass acts as a vapor separator, sending the saturated liquid fraction of the flow (if present) from the outlet of the evaporator directly to the recuperator inlet. Bypass mass flow rate is evaluated as:

$$\dot{m}_{bypass} = \dot{m}_{expander} \frac{1 - x_{ev}}{x_{ev}} \tag{7.4}$$

where x_{ev} is the quality of the refrigerant at the outlet of the evaporator.

The saturated liquid flowing through the bypass valve undergoes an isenthalpic expansion and is then mixed to the fluid coming out of the expander. The enthalpy of the refrigerant entering the recuperator is evaluated from energy conservation equation:

$$h_{rec,in} = \frac{\dot{m}_{bypass} h_{sat}(p_{ev}) + \dot{m}_{expander} h_{expander,out}}{\dot{m}_{bypass} + \dot{m}_{expander}} \tag{7.5}$$

where $h_{sat}(p_{ev})$ is the saturated refrigerant enthalpy at the evaporator pressure.

7.3.3 Recuperator

The recuperator model (Figure 7.7) is based upon an effectiveness, ε, correlation, [10, 29]. Due to the high heat transfer coefficients, the heat exchanger is characterized by a rapid response to variations in the flow rates or inlet thermodynamic conditions and a quasi-static approach can be used to model the heat transfer in the component [45]. The heat rejection rate is calculated as:

$$\dot{Q} = \varepsilon \; C_{vapor}(T_{in,hot} - T_{in,cold}) \tag{7.6}$$

where

$\dot{Q}$	heat transfer rate;
ε	effectiveness;
C_{vapor}	minimum vapor heat capacity rate;
$T_{in,cold}$	cold side inlet temperature;
$T_{in,hot}$	hot side inlet temperature.

The effectiveness is related to the refrigerant flow rate and the temperatures of the incoming fluids based on a linear correlation obtained from experimental data. Energy balance equation provides the basis for a dynamic estimation of outlet temperature, which accounts the component thermal inertia, according to:

$$MC\frac{dT_{out}}{dt} = \dot{m}c_p(T_{in} - T_{out}) + \dot{Q} \tag{7.7}$$

where $\dot{Q}$ is the heat absorbed (positive) or rejected (negative) by the fluid, while M is the mass of fluid contained in the heat exchanger, which serves as a calibration coefficient for the dynamic behavior. Equation 7.7 is applied to both the hot and cold sides of the recuperator.

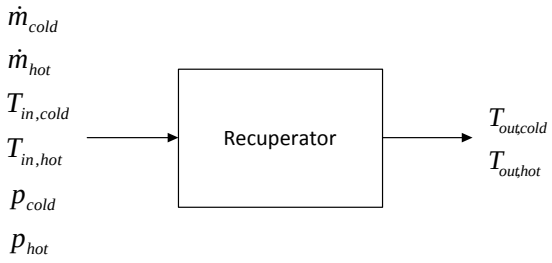


Figure 7.7: Recuperator IO scheme

7.3.4 Condenser and Cooling Circuit

In the condenser, due to the relatively high heat transfer coefficients of the fluids, the dynamics are characterized by a rapid response to variations in the flow rates or inlet thermodynamic conditions. For this reason, as shown in [14], the condensing conditions are scarcely affected by the dynamics of the cycle and therefore a quasi-static approach can be used to model the heat transfer and the outlet thermodynamic conditions of the two fluids crossing each heat exchanger [45]. The heat rejection under steady conditions at the condenser can be therefore expressed as:

$$\dot{Q}_{cond} = \dot{m}_{recuperator}h_{cond,in} - \dot{m}_{pump}h_{cond,out} \tag{7.8}$$

where

$\dot{Q}_{cond}$ condenser heat rejection;

$h_{cond,in}$ refrigerant inlet enthalpy;

$h_{cond,out}$ refrigerant enthalpy outlet enthalpy.

Furthermore, if the variation in the working conditions of the condenser are limited (compared to the evaporator), it is possible to assume a constant condensing pressure [50], chosen in this case equal to 3.5 bar, corresponding to a condensing temperature of 50.6 °C. An additional 5 °C of sub-cooling is assumed at the outlet of the condenser. The following assumptions are also made about the heat transfer in the condenser:

- the super-heated fraction of the vapor at the inlet of the condenser is negligible;
- the cooling circuit is always able to provide a coolant at 30 °C at the inlet of the condenser;
- a constant 5 K of ΔT is present at the pinch point between coolant and refrigerant

Given these assumptions, the heat transfer diagram for the condenser is that depicted in Figure 7.8 and, from energy conservation at the coolant side, it is possible to obtain the coolant required mass flow rate as follows:

$$\dot{m}_{cool} = \frac{\dot{Q}_{cond}}{c_{p,cool}\Delta T_{cool}} \tag{7.9}$$

where

$\dot{m}_{cool}$ coolant mass flow rate;
$c_{p,cool}$ coolant specific heat;
ΔT_{cool} coolant temperature increase (in this case equal to 15.6 K).

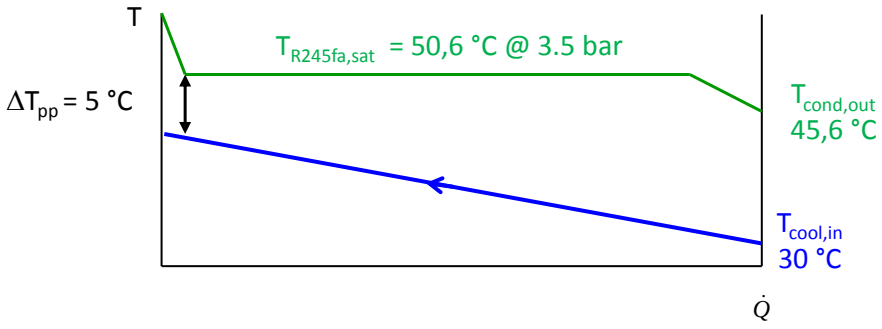


Figure 7.8: Condenser heat transfer

Finally, a linear proportionality is assumed between coolant flow rate and coolant pump power:

$$P_{cool} = k_{cool}\ \dot{m}_{cool} \tag{7.10}$$

where

P_{cool} coolant pump power;

k_{cool} constant chosen to give 1 kW of power consumption at 30 l/min of coolant flow.

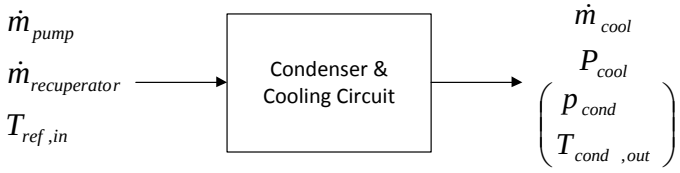


Figure 7.9: Condenser IO scheme

7.3.5 Evaporator

The evaporator, responsible for the recovery of thermal energy form the exhausts, is the component whose dynamics most largely affect the transient response of the entire ORC plant. For this reason, a detailed physical-based approach is followed to model the component. The methodology needs to be robust and accurate, while maintaining in the meantime a fast computational time suitable for control oriented applications. In order to achieve this, the Moving Boundary Method (MBM) is adopted for the description of the heat exchanger. MBM is a low-order 0-D lumped parameter methodology approach which has been widely used for Air Conditioning (AC) applications [38, 42, 64] and has been successfully adapted to the ORC applications [48, 63]; comprehensive literature reviews [5, 6] has shown that this methodology offers the desired accuracy and faster computation times compared to the classic Finite Volume approaches.

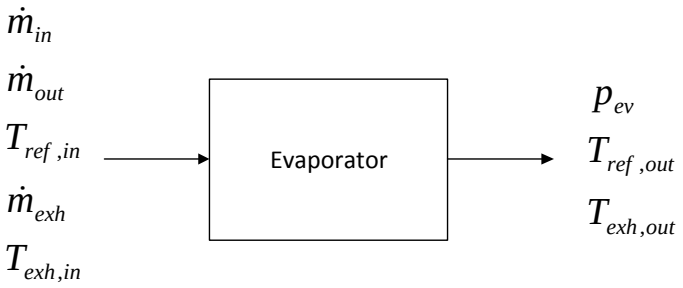


Figure 7.10: Evaporator IO scheme

MBM approach

The MBM approach is based on the following assumptions:

1. The heat exchanger is discretized into interconnected Control Volumes along the axial direction of the flow, with the number of elements corresponding to the number of thermodynamic phases of the refrigerant;
2. In each control volume the thermodynamic properties of the fluid are uniformly distributed;
3. The flow through the heat exchanger is approximated as a one-dimensional flow in a straight tube;
4. The pressure drop along the exchanger is negligible;
5. The dynamics of the exhausts side are negligible;
6. Axial heat transfer in the refrigerant, the wall and the exhausts is negligible, and only radial heat transfer is accounted;
7. The two phase region is considered in thermodynamic equilibrium;

Figure 7.11 depicts the schematic of a generic Control Volume (CV); it is to be noted that with the MBM approach, the boundaries of the CV (z_A, z_B) become additional unknown variables in the model.

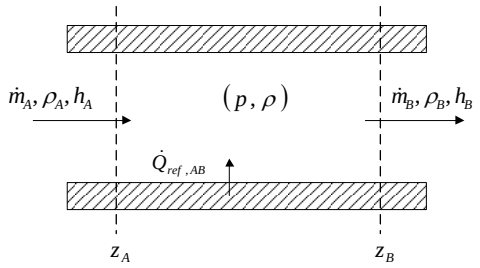


Figure 7.11: Schematic of a Control Volume for MBM

Thanks to the previous assumptions, mass and balance equations can be written for each CV in the following form [30]:

$$\frac{d}{dt}\iiint_V \rho \, dV + \iint_A \rho w_r \, dA = 0 \tag{7.11a}$$

$$\frac{d}{dt}\iiint_V (\rho h - p) \, dV + \iint_A \rho h w_r \, dA = \dot{Q}_{wall} \tag{7.11b}$$

where

V	volume of the CV;
ρ	refrigerant density;
h	refrigerant enthalpy;
p	refrigerant pressure;
A	cross flow area;
w_r	refrigerant relative velocity;
Q_{wall}	wall to refrigerant heat transfer rate.

According to the notation of Figure 7.11, following the procedure outlined in [30], Equations 7.11 can be rewritten as:

$$A\frac{d}{dt}\int_{z_A}^{z_B} \rho \, dz + \rho_A A\frac{dz_A}{dt} - \rho_B A\frac{dz_B}{dt} = \dot{m}_A - \dot{m}_B \tag{7.12a}$$

$$\begin{aligned} A\frac{d}{dt}\int_{z_A}^{z_B} \rho h \, dz - A(z_B - z_A)\frac{dp}{dt} + A\rho_A h_A \frac{dz_A}{dt} - A\rho_B h_B \frac{dz_B}{dt} = \\ = \dot{m}_A h_A - \dot{m}_B h_B + \dot{Q}_{wall} \end{aligned} \tag{7.12b}$$

Under the assumption that all the three fluid thermodynamic phases are present in the expander, the CV scheme of the component is that depicted in Figure 7.12. In this case, the refrigerant enters the evaporator as a sub-cooled fluid (SC), transitions through a two phase (TP) zone, and finally exits as a super-heated vapor (SH). When applying Equations 7.12 to each of the controlling volumes, a system of Differential Algebraic Equations (DAE) representing the entire evaporator is obtained.

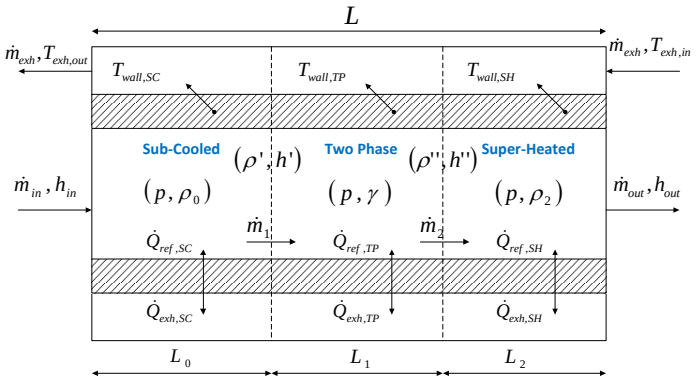


Figure 7.12: Schematic for the three zones evaporator MBM model

Sub-Cooled CV equations For the Sub-Cooled Control Volume, the lumped average thermodynamic properties of the fluid (enthalpy, density, and temperature) are evaluated as follows:

$$h_0 = \frac{1}{z_B - z_A} \int_{z_A}^{z_B} h dz \approx \frac{1}{2}(h_A + h_B) \tag{7.13a}$$

$$\rho_0 = \frac{1}{z_B - z_A} \int_{z_A}^{z_B} \rho dz \approx \rho(p, h_0) \tag{7.13b}$$

$$T_0 = \frac{1}{z_B - z_A} \int_{z_A}^{z_B} T dz \approx T(p, h_0) \tag{7.13c}$$

With these assumptions and using the notation of Figure 7.12, Equations 7.12 can be rewritten as:

$$A(\rho_0 - \rho')\frac{dL_0}{dt} + AL_0\left(\left.\frac{\partial\rho}{\partial p}\right|_{h_0} + \frac{1}{2}\left.\frac{\partial\rho}{\partial h}\right|_p \frac{dh'}{dp}\right)\frac{dp}{dt} + \frac{1}{2}AL_0\left.\frac{\partial\rho}{\partial h}\right|_p \frac{dh_{in}}{dt} = \\ = \dot{m}_{in} - \dot{m}_1 \tag{7.14a}$$

$$A(\rho_0 h_0 - \rho' h')\frac{dL_0}{dt} + AL_0\left[\frac{1}{2}\rho_0\frac{dh'}{dp} + h_0\left(\left.\frac{\partial\rho}{\partial p}\right|_{h_0} + \frac{1}{2}\left.\frac{\partial\rho}{\partial h}\right|_p \frac{dh'}{dp}\right) - 1\right]\frac{dp}{dt} + \\ + \frac{1}{2}AL_0\left(\rho_0 + h_0\left.\frac{\partial\rho}{\partial h}\right|_p\right)\frac{dh_{in}}{dt} = \dot{m}_{in}h_{in} - \dot{m}_1 h' + \dot{Q}_{ref,SC} \tag{7.14b}$$

with the superscript ' indicating that the property evaluated at saturated liquid condition.

Two Phase CV equations In the Two Phase Control Volume, under thermodynamic equilibrium, the temperature of the refrigerant is a function of the pressure, while for the density the following physical correlation subsist:

$$\rho_1 = \frac{1}{z_B - z_A}\int_{z_A}^{z_B} \rho dz = \bar{\gamma}\rho'' + (1 - \bar{\gamma})\rho' \tag{7.15}$$

where

$\bar{\gamma} = \frac{V'}{V''}$ is the average void fraction;

$'$ stands for saturated liquid;

$''$ stands for saturated vapor;

Moreover, if a slip factor μ is defined as:

$$\mu = \frac{\rho''}{\rho'} \tag{7.16}$$

the average void fraction $\bar{\gamma}$ can be expressed, following [30], as:

$$\bar{\gamma} = \frac{(x_{out} - x_{in})(1 - \mu^{2/3}) - \mu\Big(x_{in}(\mu^{-1/3} - 1) + 1\Big)\Big(x_{out}(\mu^{-1/3} - 1) + 1\Big)\beta}{(x_{out} - x_{in})(1 - \mu^{2/3})^2} \tag{7.17}$$

where

$$\beta = \ln \frac{\Big(x_{in}(\mu^{-1/3}-1)+1\Big)\Big(x_{out}(1-\mu)+\mu\Big)}{\Big(x_{out}(\mu^{-1/3}-1)+1\Big)\Big(x_{in}(1-\mu)+\mu\Big)} \tag{7.18}$$

and x_{in}, x_{out} are the inlet and outlet values of refrigerant quality at the boundaries of the TP Control Volume. When both the SC and SH Control Volumes are present, $x_{in} = 0$ and $x_{out} = 1$. In this case, Equation 7.17 is reduced to:

$$\bar{\gamma} = \frac{1-\mu^{2/3}(1+\ln(\mu^{-2/3}))}{(1-\mu^{2/3})^2} \tag{7.19}$$

Mass and energy balance for the TP Control Volume can be therefore derived as:

$$\begin{aligned} &A(\rho'-\rho'')\left(\frac{dL_0}{dt}+(1-\bar{\gamma})\frac{dL_1}{dt}\right)+ \\ &+AL_1\left[\bar{\gamma}\frac{d\rho''}{dp}+(1-\bar{\gamma})\frac{d\rho'}{dp}+(\rho''-\rho')\frac{d\bar{\gamma}}{dp}\right]\frac{dp}{dt} = \dot{m}_1-\dot{m}_2 \end{aligned} \tag{7.20a}$$

$$\begin{aligned} &A(\rho'h'-\rho''h'')\left(\frac{dL_0}{dt}+(1-\bar{\gamma})\frac{dL_1}{dt}\right)+ \\ &+AL_1\left[\bar{\gamma}\frac{d\rho''h''}{dp}+(1-\bar{\gamma})\frac{d\rho'h'}{dp}+(\rho''h''-\rho'h')\frac{d\bar{\gamma}}{dp}-1\right]\frac{dp}{dt} = \\ &= \dot{m}_1h'-\dot{m}_2h''+\dot{Q}_{ref,TP} \end{aligned} \tag{7.20b}$$

Super-Heated CV equations For the Super-Heated CV, the average thermodynamic properties of the fluid are evaluated as in the Sub-Cooled Control Volume:

$$h_2 = \frac{1}{z_B-z_A}\int_{z_A}^{z_B} h\,dz \approx \frac{1}{2}(h_A+h_B) \tag{7.21a}$$

$$\rho_2 = \frac{1}{z_B-z_A}\int_{z_A}^{z_B} \rho\,dz \approx \rho(p,h_2) \tag{7.21b}$$

$$T_2 = \frac{1}{z_B-z_A}\int_{z_A}^{z_B} T\,dz \approx T(p,h_2) \tag{7.21c}$$

Still referencing to Figure 7.12 notation, Equations 7.12 can now be rewritten as:

$$A(\rho'' - \rho_2)\left(\frac{dL_0}{dt} + \frac{dL_1}{dt}\right) + AL_2\left(\left.\frac{\partial\rho}{\partial p}\right|_{h_2} + \frac{1}{2}\left.\frac{\partial\rho}{\partial h}\right|_p \frac{dh''}{dp}\right)\frac{dp}{dt} + \\ + \frac{1}{2}AL_2\left.\frac{\partial\rho}{\partial h}\right|_p \frac{dh_{out}}{dt} = \dot{m}_2 - \dot{m}_{out} \tag{7.22a}$$

$$A(\rho''h'' - \rho_2 h_2)\left(\frac{dL_0}{dt} + \frac{dL_1}{dt}\right) + \\ + AL_2\left[\frac{1}{2}\rho_2\frac{dh''}{dp} + h_2\left(\left.\frac{\partial\rho}{\partial p}\right|_{h_2} + \frac{1}{2}\left.\frac{\partial\rho}{\partial h}\right|_p \frac{dh''}{dp}\right) - 1\right]\frac{dp}{dt} + \\ + \frac{1}{2}AL_2\left(\rho_2 + h_2\left.\frac{\partial\rho}{\partial h}\right|_p\right)\frac{dh_{out}}{dt} = \dot{m}_2 h'' - \dot{m}_{out}h_{out} + \dot{Q}_{ref,SH} \tag{7.22b}$$

Heat transfer and wall temperature models The wall to refrigerant heat transfer terms $\dot{Q}_{ref,SC}$, $\dot{Q}_{ref,TP}$, $\dot{Q}_{ref,SH}$ are evaluated from wall energy balance equation, under the hypothesis that the wall temperature is constant and very close to the refrigerant temperature. This assumption is legit as the heat transfer coefficient at the refrigerant side is much greater than that at the exhaust side. With this assumption, for each control volume and referring to the notation of Figure 7.11, the wall to refrigerant heat transfer can be inferred from wall energy balance equation as follows:

$$\begin{aligned} \dot{Q}_{ref} &= \dot{Q}_{exh} - M_{wall}C_{wall}\frac{dT_{wall}}{dt} \approx \\ &\approx \dot{Q}_{exh} - M_{wall}C_{wall}\frac{dT_{ref}}{dt} = \\ &= \dot{Q}_{exh} - M_{wall}C_{wall}\left(\frac{\partial T_{ref}}{\partial p}\frac{dp}{dt} + \frac{\partial T_{ref}}{\partial h}\frac{dh_A}{dt} + \frac{\partial T_{ref}}{\partial h}\frac{dh_B}{dt}\right) \end{aligned} \tag{7.23}$$

where

$\dot{Q}_{exh}$ exhaust to wall heat transfer;

M_{wall} wall mass;

C_{wall} wall thermal inertia;

T_{wall} wall temperature;

Heat transfer from the exhausts is evaluated with the $\varepsilon - NTU$ approach [29] as a fraction of the maximum theoretical heat transfer:

$$\dot{Q}_{exh} = \varepsilon \, \dot{Q}_{max} = \varepsilon \, C_{exh}(T_{exh,in} - T_{ref}) \tag{7.24a}$$

$$\varepsilon_{counter-flow} = \frac{1 - \exp\Big[- NTU(1 + \frac{C_{exh}}{C_{ref}})\Big]}{1 - \frac{C_{exh}}{C_{ref}} \exp\Big[- NTU(1 + \frac{C_{exh}}{C_{ref}})\Big]} \tag{7.24b}$$

$$\varepsilon_{two\ phase} = 1 - \exp(-NTU) \tag{7.24c}$$

$$NTU = \frac{U A_h}{C_{exh}} \tag{7.24d}$$

$$U = \frac{1}{\frac{1}{\alpha_{ref}} + \frac{A_{h,i}}{A_{h,e}\alpha_{exh}}} \tag{7.24e}$$

where

C_{exh} exhaust heat capacity rate ;
C_{ref} refrigerant heat capacity rate ;
ε effectiveness;
NTU Number of Transfer Units;
U overall heat transfer coefficient;
A_h heat transfer area (i for internal, e for external);
α_{ref} refrigerant heat transfer coefficient;
α_{exh} exhaust heat transfer coefficient.

Heat transfer coefficients are evaluated through Dittus-Boelter correlation for the refrigerant single phase region and the exhausts side:

$$Nu = 0.023 Re^{0.8} Pr^{m} \tag{7.25}$$

with $m = 0.3$ for the exhaust side and $m = 0.4$ for the refrigerant side. For the two phase region, Yan Lin correlation [61] is used:

$$\alpha_{ref} = 1.926 Re_{eq} Pr_L^{1/3} Re_L^{-0.5} \lambda_L / D_h \tag{7.26}$$

where Re_{eq} is Reynolds equivalent number [2] and L indicates that the property is evaluated for saturated liquid conditions.

MBM system implementation Mass and energy conservation Equations for the 3 CVs (7.14, 7.20, 7.22) form a set of DAE which can be expressed in the form:

$$Z(x)\begin{bmatrix}\dot{x}\\ \dot{m}_1\\ \dot{m}_2\end{bmatrix} = f(x,u) \tag{7.27}$$

with

$Z(x)$	a 6x6 matrix obtained Equations 7.14, 7.20, 7.22;
x	evaporator states $x = [L_0 \quad L_1 \quad p \quad h_{out}]'$;
u	evaporator inputs $u = [\dot{m}_{in} \quad \dot{m}_{out} \quad T_{ref,in} \quad \dot{m}_{exh} \quad T_{exh}]'$;
$f(x,u)$	a 6x1 array obtained Equations 7.14, 7.20, 7.22.

Since Z(x) is a function of state variables, it is non-singular and generally invertible [39]; during the simulation of the plant, the matrix Z(x) and the vector f(x,u) are evaluated numerically at each time step and used to obtained the derivative of the states $\dot{x}$, which is then integrated by the solver. The solution process is depicted in Figure 7.13.

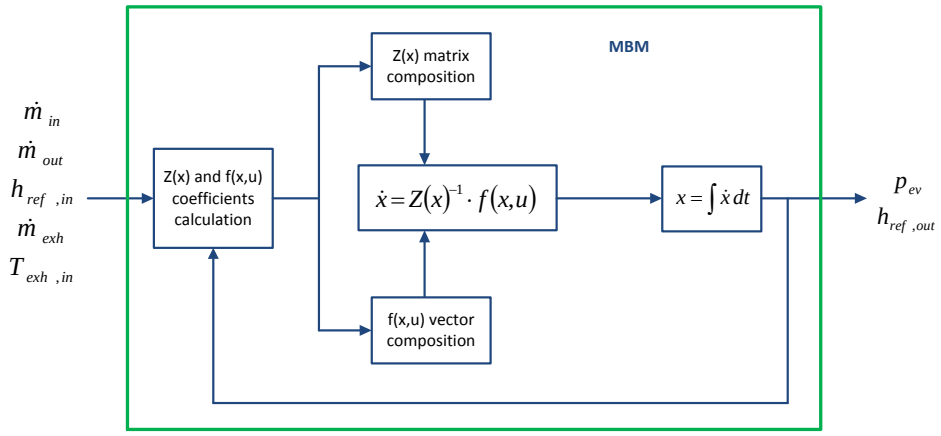


Figure 7.13: Schematic of the MBM implementation

Switching moving boundary The evaporator model developed as far is based on the assumption that three thermodynamic phases of the refrigerant are always present in the heat exchanger. It is possible however that during non-nominal operating conditions (start-ups, fast transients, low engine load, etc.) this assumption may not hold anymore. To deal with these cases and to improve the robustness of the model, the basic MBM model has been extended to include a "switching" logic, a modelling technique already successfully implemented for the simulation of Vapor Compression Cycles (VCC) with MBM models [42]. The resulting evaporator model must be able to adjust its internal structure depending on the inlet/outlet conditions of the working fluid. In order to do so, four additional evaporator sub-models has been added to the original one, according to the distribution of phase regions indicated in Table 7.1. The introduction of a "pseudo-quality" parameter χ evaluated at the inlet and outlet of the evaporator allows to distinguish between the different conditions:

$$\chi = \frac{h - h'}{h'' - h'} \tag{7.28}$$

The pseudo-quality parameter χ is defined in the range $-\infty < \chi < +\infty$, with $0 < \chi < 1$ representing the usual fluid quality x for a two phase fluid under thermodynamic equilibrium.

Table 7.1: Sub-models library for switching MBM

Model name	**# of CV**	**Inlet condition**	**Outlet condition**	**CV lengths**
SC+TP+SH	3	$\chi_{in} < 0$	$\chi_{out} > 1$	$L_0, L_1, L_2 > 0$
SC+TP	2	$\chi_{in} < 0$	$0 < \chi_{out} < 1$	$L_0, L_1 > 0$ $L_2 = 0$
TP+SH	2	$0 < \chi_{in} < 1$	$\chi_{out} > 1$	$L_1, L_2 > 0$ $L_0 = 0$
SC	1	$\chi_{in} < 0$	$\chi_{out} < 0$	$L_0 > 0$ $L_1, L_2 = 0$
TP	1	$0 < \chi_{in} < 1$	$0 < \chi_{out} < 1$	$L_1 > 0$ $L_0, L_2 = 0$

To obtain the four additional evaporator sub-models, mass and energy conservation equations have been recalculated according to the I/O configuration of each control volume in each of the four additional configurations as depicted in the following paragraphs.

SC+TP equations

Sub-cooled

Same as Equations 7.14

Two phase

$$A\bar{\gamma}(\rho' - \rho'')\frac{dL_0}{dt} + AL_1(\rho'' - \rho')\left(\frac{\partial\bar{\gamma}}{\partial x_{out}}\frac{\partial x_{out}}{\partial h_{out}}\right)\frac{dh_{out}}{dt} + \\ + AL_1\left[\bar{\gamma}\frac{d\rho''}{dp} + (1-\bar{\gamma})\frac{d\rho'}{dp} + (\rho'' - \rho')\left(\frac{\partial\bar{\gamma}}{\partial p} + \frac{\partial\bar{\gamma}}{\partial x_{out}}\frac{\partial x_{out}}{\partial p}\right)\right]\frac{dp}{dt} = \\ = \dot{m}_1 - \dot{m}_{out} \tag{7.29a}$$

$$A\bar{\gamma}(\rho'h' - \rho''h'')\frac{dL_0}{dt} + AL_1(\rho''h'' - \rho'h')\left(\frac{\partial\bar{\gamma}}{\partial x_{out}}\frac{\partial x_{out}}{\partial h_{out}}\right)\frac{dh_{out}}{dt} + \\ + AL_1\left[\bar{\gamma}\frac{d\rho''h''}{dp} + (1-\bar{\gamma})\frac{d\rho'h'}{dp} + (\rho''h'' - \rho'h')\left(\frac{\partial\bar{\gamma}}{\partial p} + \frac{\partial\bar{\gamma}}{\partial x_{out}}\frac{\partial x_{out}}{\partial p}\right) - 1\right]\frac{dp}{dt} = \\ = \dot{m}_1 h' - \dot{m}_{out}h_{out} + \dot{Q}_{ref,TP} \tag{7.29b}$$

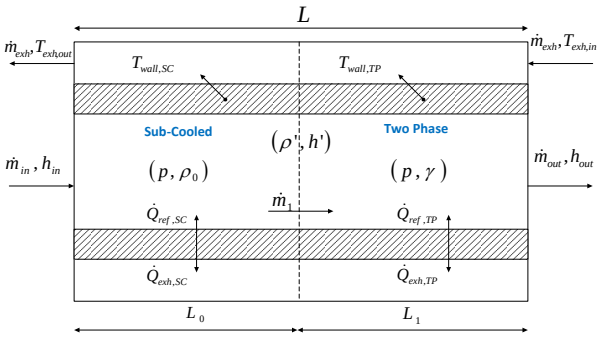


Figure 7.14: SC+TP evaporator sub-model

TP+SH equations

Two phase

$$A(1-\bar{\gamma})(\rho'-\rho'')\frac{dL_1}{dt}+AL_1(\rho''-\rho')\left(\frac{\partial\bar{\gamma}}{\partial x_{in}}\frac{\partial x_{in}}{\partial h_{in}}\right)\frac{dh_{in}}{dt}+$$
$$+AL_1\left[\bar{\gamma}\frac{d\rho''}{dp}+(1-\bar{\gamma})\frac{d\rho'}{dp}+(\rho''-\rho')\left(\frac{\partial\bar{\gamma}}{\partial p}+\frac{\partial\bar{\gamma}}{\partial x_{in}}\frac{\partial x_{in}}{\partial p}\right)\right]\frac{dp}{dt}= \tag{7.30a}$$
$$=\dot{m}_{in}-\dot{m}_1$$

$$A(1-\bar{\gamma})(\rho'h'-\rho''h'')\frac{dL_1}{dt}+AL_1(\rho''h''-\rho'h')\left(\frac{\partial\bar{\gamma}}{\partial x_{in}}\frac{\partial x_{in}}{\partial h_{in}}\right)\frac{dh_{in}}{dt}+$$
$$+AL_1\left[\bar{\gamma}\frac{d\rho''h''}{dp}+(1-\bar{\gamma})\frac{d\rho'h'}{dp}+(\rho''h''-\rho'h')\left(\frac{\partial\bar{\gamma}}{\partial p}+\frac{\partial\bar{\gamma}}{\partial x_{in}}\frac{\partial x_{in}}{\partial p}\right)-1\right]\frac{dp}{dt}= \tag{7.30b}$$
$$=\dot{m}_{in}h_{in}-\dot{m}_1h''+\dot{Q}_{ref,TP}$$

Super-heated

Same as Equations 7.22

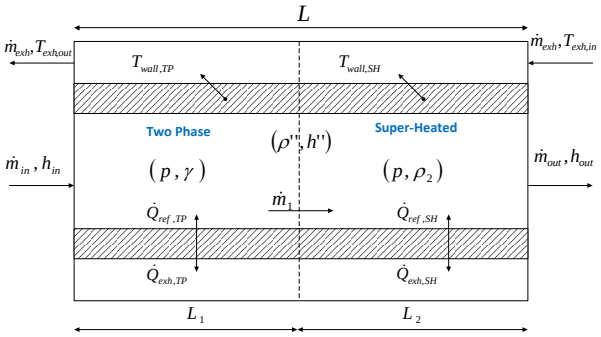


Figure 7.15: TP+SH evaporator sub-model

SC equations

Sub-cooled

$$AL_0 \frac{\partial \rho}{\partial p}\bigg|_{h_0} \frac{dp}{dt} + \frac{1}{2} AL_0 \frac{\partial \rho}{\partial h}\bigg|_{p} \frac{dh_{in}}{dt} + \frac{1}{2} AL_0 \frac{\partial \rho}{\partial h}\bigg|_{p} \frac{dh_{out}}{dt} = \dot{m}_{in} - \dot{m}_{out} \tag{7.31a}$$

$$\begin{aligned} AL_0 \left(h_0 \frac{\partial \rho}{\partial p}\bigg|_{h_0} - 1 \right) \frac{dp}{dt} + \frac{1}{2} AL_0 \left(\rho_0 + h_0 \frac{\partial \rho}{\partial h}\bigg|_{p} \right) \frac{dh_{in}}{dt} + \\ + \frac{1}{2} AL_0 \left(\rho_0 + h_0 \frac{\partial \rho}{\partial h}\bigg|_{p} \right) \frac{dh_{out}}{dt} = \dot{m}_{in} h_{in} - \dot{m}_{out} h_{out} \end{aligned} \tag{7.31b}$$

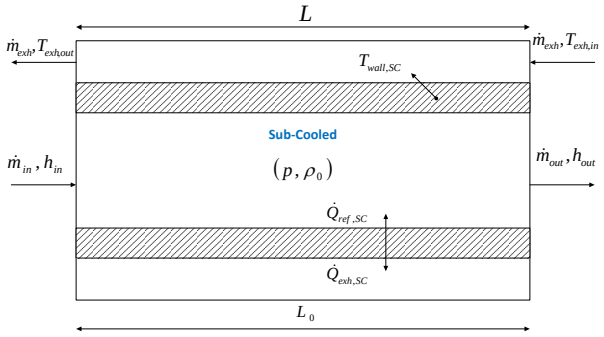


Figure 7.16: SC evaporator sub-model

TP equations

Two phase

$$AL_1\left[\bar{\gamma}\frac{d\rho''}{dp}+(1-\bar{\gamma})\frac{d\rho'}{dp}+(\rho''-\rho')\left(\frac{\partial\bar{\gamma}}{\partial p}+\frac{\partial\bar{\gamma}}{\partial x_{in}}\frac{\partial x_{in}}{\partial p}+\frac{\partial\bar{\gamma}}{\partial x_{out}}\frac{\partial x_{out}}{\partial p}\right)\right]\frac{dp}{dt}+$$
$$+AL_1(\rho''-\rho')\left(\frac{\partial\bar{\gamma}}{\partial x_{in}}\frac{\partial x_{in}}{\partial h_{in}}\right)\frac{dh_{in}}{dt}+AL_1(\rho''-\rho')\left(\frac{\partial\bar{\gamma}}{\partial x_{out}}\frac{\partial x_{out}}{\partial h_{out}}\right)\frac{dh_{out}}{dt}=\dot{m}_{in}-\dot{m}_{out} \tag{7.32a}$$

$$AL_1\left[\bar{\gamma}\frac{d\rho''h''}{dp}+(1-\bar{\gamma})\frac{d\rho'h'}{dp}+(\rho''h''-\rho'h')\left(\frac{\partial\bar{\gamma}}{\partial p}+\frac{\partial\bar{\gamma}}{\partial x_{in}}\frac{\partial x_{in}}{\partial p}+\frac{\partial\bar{\gamma}}{\partial x_{out}}\frac{\partial x_{out}}{\partial p}\right)\right]\frac{dp}{dt}+$$
$$+AL_1(\rho''h''-\rho'h')\left(\frac{\partial\bar{\gamma}}{\partial x_{in}}\frac{\partial x_{in}}{\partial h_{in}}\right)\frac{dh_{in}}{dt}+AL_1(\rho''h''-\rho'h')\left(\frac{\partial\bar{\gamma}}{\partial x_{out}}\frac{\partial x_{out}}{\partial h_{out}}\right)\frac{dh_{out}}{dt}=$$
$$=\dot{m}_{in}h_{in}-\dot{m}_{out}h_{out}+\dot{Q}_{ref,TP} \tag{7.32b}$$

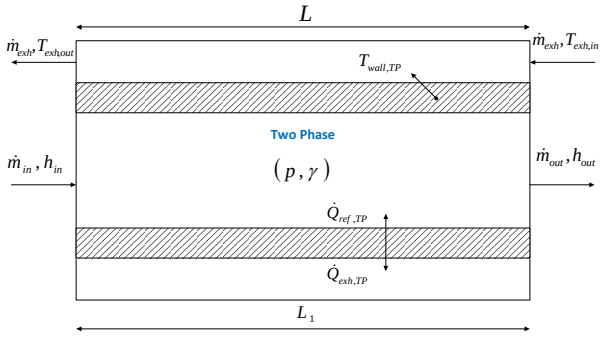


Figure 7.17: TP evaporator sub-model

Switching algorithm

The switching between the different evaporator sub-models is controlled by the finite state machine algorithm depicted in Figure 7.18. The switching happens based on the value of the inlet/outlet pseudo-quality χ or when a Control Volume length is lower than a determined threshold (calibrated to guarantee numerical stability in the model). To prevent the chattering between two different states, conditions on the derivatives of pseudo-quality and lengths are also accounted, following the approach developed by McKinley et al. [42].

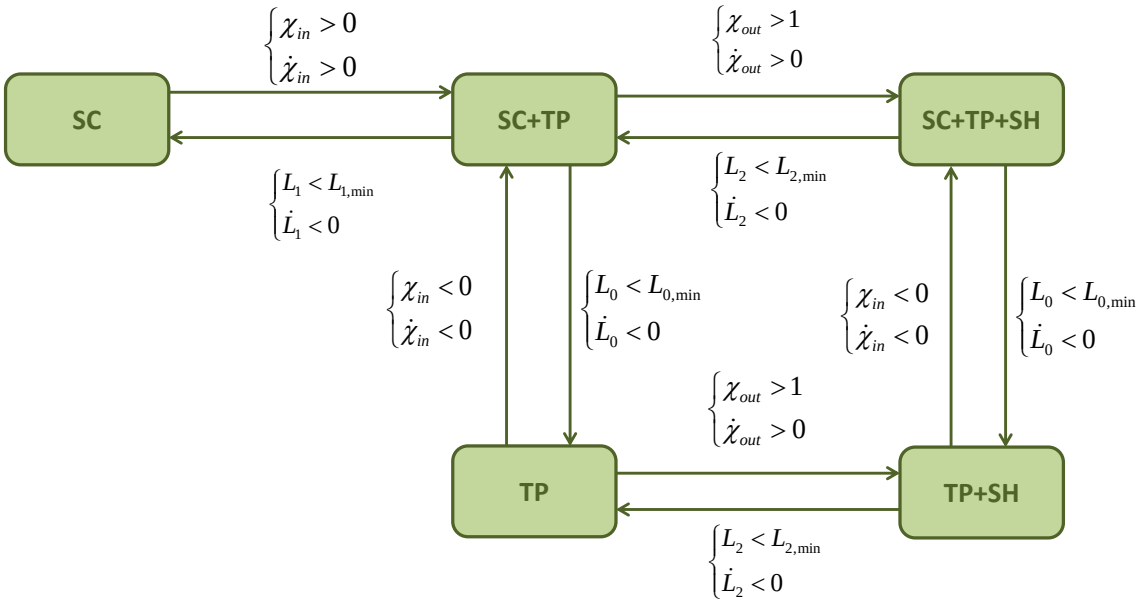


Figure 7.18: Finite State Machine for MBM switching algorithm

7.3.6 Comprehensive ORC model

The various sub-component models described in the previous sections are connected together following the plant schematic depicted in Figure 7.5. Under the assumption that the matrix $Z(x)$ from evaporator Equation 7.27 is non-singular (guaranteed by the switching modeling technique [10]), the set of equations describing the ORC system may ideally be rearranged in the following explicit non-linear state space form:

$$\dot{x}(t) = f\big(x(t), u(t), w(t)\big) \qquad x(0) = x_0$$
$$y(t) = h\big(x(t), u(t)\big) \tag{7.33}$$

where $x(t) = [L_0, \quad L_1, \quad p_{ev}, \quad h_{ev}]' \in \mathbb{R}^n$ denotes the state vector, $y(t) \in \mathbb{R}^p$ denotes the system outputs (which represent any combination of the other parameters estimated from the model, e.g. the temperature of the fluid, the refrigerant mass flow rate, etc.), $u(t) = [n_{pump}, \quad x_{bypass}]' \in \mathbb{R}^m$ denotes the control inputs, and $w(t) = [m_{exh}, \quad T_{exh}, \quad n_{eng}]' \in \mathbb{R}^s$ denotes the exogenous disturbances from the engine.

7.4 Transient optimization

While ORC plants for stationary applications and heavy duty vehicles are usually designed to predominantly operate at steady-state conditions, where best performance can be achieved [8], in Waste Heat Recovery applications for light vehicles, the variability of the heat source and the necessity of harvesting the maximum amount of energy available in every situation pose an interesting problem for the development of an effective management strategy for this kind of plants. It is therefore of considerable interest for the automotive industry to both investigate the behaviour of ORC systems in presence of significant load transients and to analyze the possible strategies to maximize the fuel economy increase under such conditions. Contributions in this field, while relatively limited, are increasing in the recent years. The majority of the works focuses on disturbance rejection control policies which aim to maintain certain desired operation parameters (mainly evaporation pressure and temperature) despite the variability of the heat source. For instance, Peralez et al. [49] combined a feedback PID controller with an inverse reduced MBM model for the feed-forward generation of control references from measured disturbances, with the aim of tracking a desired SH temperature at the evaporator outlet. Feru et al. [18] used a linearized plant model to develop an MPC controller and a Kalman filter for reference tracking of expander inlet quality. Hou et al. [26] developed a supervisory predictive controller for the evaporator, while Zhang et al. [62] applied a generalized predictive control strategy. More recently Hernandez et al. [22] developed a perturbation based Extremum Seeking (ES) algorithm to find the optimal evaporation temperature online while the plant is functioning.

The methods previously described are very effective in guiding the process to a target set-point rapidly and reliably, providing good stability and disturbance rejection

performances, and they are able to guarantee good energetic efficiencies, especially under steady state condition. However, they all rely on the traditional partitioning of the processes of information management, decision making, and system control system into two different layers. The first layer performs a steady-state economic optimization of the plants variables (whose economic objective is in this case the maximization of the harvested energy), and the second layer, uses the knowledge of the optimal steady-state economic reference set-points to rapidly guide the plants transient and reject the effect of dynamic disturbances. As mentioned in Chapter 4.3.1 however, for many applications this hierarchical separation of information and purpose is no longer optimal or desirable [51]. What is lacking, in fact, is the ability to properly deal with the optimization of the economic parameter during the transient phases. In application such as Waste Heat Recovery, this restriction may pose a serious limitation on plant efficiency, as the process is, for its own nature, very unlikely to operate under long steady state conditions. In the present section a different approach, based on an economical formulation of the MPC method, is adopted, where the task of the control algorithm is to directly optimize the entire plant energy production during a short transient phase, rather than tracking some optimal steady set-points. The algorithm relies on the dynamic model of the plant which predicts the evolution of the system and allows to evaluate the optimal input trajectory. The metaheuristic MPC optimization method based on Particle Swarm Optimization (PSO) described in Chapter 4.3.2 is adopted to overcome the difficulties in solving the corresponding non-linear Receding Horizon optimization problem. Finally, experimental data from a real driving transient are used to provide a set of realistic operating conditions for validating the algorithm, and a simpler traditional control strategy, based on static plant optimization, is used as a benchmark to assess the performance of the developed controller.

7.4.1 Transient profile and problem definition

In order to test the ability of the model to deal with realistic transient conditions typical of automotive applications and to design a control algorithm able to properly deal with these situations, experimental data have been collected from an instrumented vehicle equipped with a 2.0 L turbocharged SI engine and a 6-speed manual transmission during normal on-road operating conditions. Figure 7.20 depicts the specific manoeuvre, composed of various accelerations and decelerations at highway conditions, with an average speed of 65 km/h, a maximum speed of 87 km/h, and a drop at the end to 9 km/h. This manoeuvre gives rise to significant variations

in the amount of heat available to the system, and provides an excellent benchmark for evaluating both the ability of the model to robustly operate in a wide range of operating conditions, as well as the efficiency of the control algorithm in maximizing the energy harvesting while maintaining the plant in safe operating conditions.

The set of experimental data depicted in Figure 7.19, i.e. exhaust flow rate, exhaust temperature, and engine speed, are those necessary to simulate the dynamic response of the ORC plant. In this study, due to the relatively low amount of power recovered from the engine, the effects of the ORC plant on the normal engine functioning are neglected.

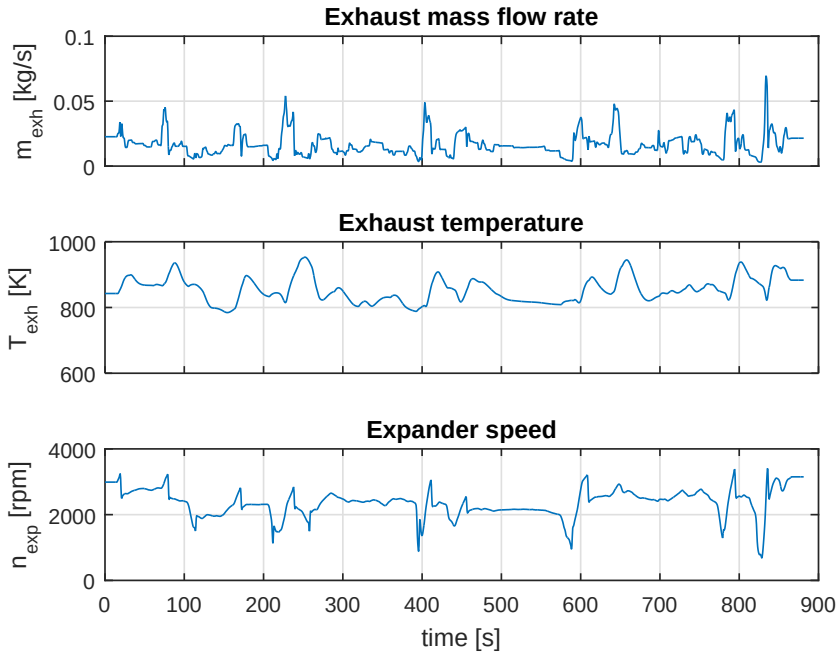


Figure 7.19: Measured engine parameters during the transient test

Figure 7.20: Vehicle speed profile during the transient test

7.4.2 Dynamic PSO based economic n-MPC controller

The goal for the control algorithm is to maximize the energy harvested from the system during the investigated transient operation. The problem can be formulated as an economic MPC problem with the introduction of the following objective function J:

$$\begin{aligned} J(x(t),u(t),w(t)) &= -\int_{t}^{t+T_P} P_{net}(x(t),u(t),w(t))\ dt = \\ &= -\int_{t}^{t+T_P} \Big(P_{exp}(x(t),u(t),w(t)) - P_{pump}(x(t),u(t),w(t)) + \\ &\quad - P_{cool}(x(t),u(t),w(t)) \Big)\ dt \end{aligned} \tag{7.34}$$

where $x(t)$ is the state vector defined in the ORC state space representation Equation 7.33, $w(t)$ is the vector of exogenous disturbances from the transient profile, and $u(t)$ is the vector of control inputs. The control algorithm must be able to estimate the optimal input trajectory $u^*(t) = [n^*_{pump}(t),\ x^*_{bypass}(t)]$ to apply to the system in order to minimize J over the finite prediction horizon T_P. Furthermore, the optimization problem needs to account for the physical constraints of the input variables, as the actuators are bounded by saturation limits, as well as the rate of change of the actuators ($\dot{u}$). Finally, the control algorithm must be able to guarantee safe operating conditions for the plant, which require the limitation of the maximum pressure and

temperature at the outlet of the evaporator and the guarantee of a minimum subcool degree at the evaporator inlet. The physical value of such constraints are defined according to the physical limitations of the different components of the system and are summarized in Table 7.2.

Table 7.2: Summary of ORC System Constraints

Input constraints
$50\ rpm \leq n_{pump} \leq 700\ rpm$
$0 \leq x_{bypass} \leq 1$
Input rate of change constraints
$\left\lvert \frac{dn_{pump}}{dt} \right\rvert \leq 50\ rpm/s$
$\left\lvert \frac{dx_{bypass}}{dt} \right\rvert \leq 0.2\ x_{bypass,max}/s$
Output constraints
$p_{ev} \leq 25\ bar$
$T_{ev,out} \leq 175\ °C$
$T_{ev,in} \leq T_{sat} - 5\ °C$

Following the MPC approach, the evaluation of the optimal control trajectory u^*, which minimizes the cost J, is performed on the basis of a predicted evolution of the system, performed by means of the ORC plant model. Given that a candidate control trajectory $\bar{u}(t)$ is applied, and the system states $x(t_k)$ and disturbances $w(t_k)$ are measured at time t_k, it is possible to estimate the evolution of system states and outputs $\bar{x}(t)$, $\bar{y}(t)$ (with the $\bar{}$ notation used to distinguish the estimated trajectories from the real ones) only if the external inputs are known during the entire prediction horizon. As this knowledge is however unobtainable for real applications, due to the stochastic nature of the disturbance source, the estimation is carried out assuming that external disturbances will remain constant during the prediction horizon $T_P = N_P \cdot T$ ($\bar{w}(t) \equiv w(t_k)$). To obviate to the error introduced with this assumption, only the first part of the optimal estimated control trajectory $\bar{u}^*(t)$ is implemented, then, as a new sample from the system is obtained, the procedure is repeated using the updated values of system real states and disturbances. The procedure is iterated at

each sample time $t_k := kT$, with T the sampling period and $k \in \mathbb{N}$, in the standard Receding Horizon fashion.

As the continuous time representation of $\bar{u}(t)$ would require the solution of a functional optimization problem, the input trajectory is assumed to be piecewise linear, i.e.:

$$u(t) \equiv u(t_k) + \left(\frac{u(t_{k+1}) - u(t_k)}{T}\right)(t - t_k) \qquad \forall t \in [t_k, t_{k+1}),\ k \in \mathbb{N} \tag{7.35}$$

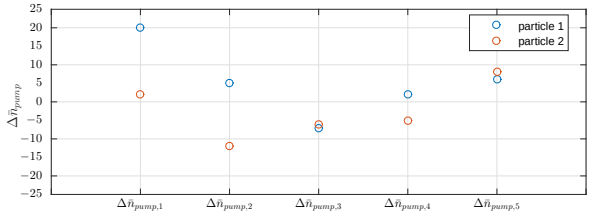
The discretized representation of the control input trajectory allows to univocally define its evolution over the prediction horizon using only $N_P \times m$ parameters, with m the number of control inputs (in this case $m = 2$) and N_P the length of the prediction horizon. The problem can be therefore formulated as a mathematical programming one and solved with the method presented in Chapter 4.3.2. Furthermore, to additionally reduce the complexity of the problem and the computational burden, only the first N_C steps of the input profile are optimized, corresponding to the so-called control horizon $T_C = N_C \cdot T$, and the input trajectory is assumed to stay constant for the remaining part of the prediction horizon ($\bar{u}(t) = \bar{u}(t_k + T_C)\ \forall t \in [t_k + T_C, t_k + T_P]$). Pump speed and exhaust bypass are therefore defined only at integer multiples $k = \{1, 2, ..., N_C\}$ of the basic sample time T, and intermediate values are evaluated based on the definition of piecewise linear function (Equation 7.35).

Also, to better handle the constraints on the inputs rate of change, it is more convenient to parameterize the future control trajectory based on the $N_C \times m$ function variations $\Delta\bar{\mathbf{u}} = [\Delta\bar{u}_1, \Delta\bar{u}_2, ..., \Delta\bar{u}_{N_C}] = [\bar{u}(t_{k+1}) - u(t_k), \bar{u}(t_{k+2}) - \bar{u}(t_{k+1}), ..., \bar{u}(t_{k+N_C}) - \bar{u}(t_{k+N_C-1})]$ rather than on the $N_C \times m$ discrete values of the control inputs.

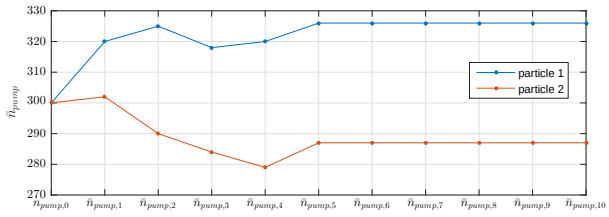
Given the sampled-data structure described above, the evaluation of the optimal combination of the $N_C \times m$ parameters which would minimize the objective function J over the prediction horizon T_P, is carried out by assigning at each particle i of a PSO algorithm a candidate position $\Delta\bar{\mathbf{u}}^i$ and a velocity $\mathbf{v}^i$, with $i \in S = 1, 2, ..., s$. The position of the particles (and the corresponding input trajectories) are then updated based on the PSO algorithm rules.

To better illustrate the correlation between each PSO particle position and the associated input trajectory it is possible to refer to Figure 7.21, where two different particles coordinates and their associated trajectory are depicted. In this simple example, it is assumed that a generic particle represent the update of pump speed trajectory at each sample time T over a prediction horizon N_P of 10 steps, of which the first 5 are to be optimized, and that the initial pump speed value $n_{pump,0}$ is known.

Figure 7.21a depicts the coordinates of two generic PSO particles, and Figure 7.21b depicts the associated pump speed trajectories.



(a) particle coordinates for pump speed



(b) resulting pump speed trajectories

Figure 7.21: Example of pump speed trajectories generation

The correct definition of the search space of each particle and the pre-screening of input trajectories greatly helps to reduce the computational burden by eliminating infeasible trajectories. The constraints on input rate of change are easily accounted by restricting the search range for the PSO particles, thanks to the function variations formulation $\Delta\bar{\mathbf{u}}$, while he trajectories violating the constraint on input saturation are simply marked as unfeasible. Constraints on system outputs on the other hand proved to be more problematic to address. After fast variations of the exogenous disturbances $w(t)$ it may be in fact difficult (or rather impossible) to define any input trajectory able to maintain the system inside the feasible region for the entire duration of the prediction horizon. To maintain the feasibility of the problem, the formulation of the objective function is modified, by turning the hard constraints on evaporator inlet and outlet temperature into soft constraints, which are subsequently embedded into the objective function J as penalty terms of the form:

$$J = -\int_{t_k}^{t_k+T_P} (P_{expander} - P_{pump} - P_{coolant})dt + k_{T_{in}} E_{T_{in}} + k_{T_{out}} E_{T_{out}}$$

$$\text{where:} \qquad E_{T_{in}} = max\left(0, \frac{1}{T_P} \int_{t_k}^{t_k+T_P} \left[(T_{sat} - T_{ev,in}) - \Delta T_{SC,min}\right] d\tau\right);$$

$$E_{T_{out}} = max\left(0, \frac{1}{T_P} \int_{t_k}^{t_k+T_P} \left[T_{ev,out} - T_{ev,out,max}\right] d\tau\right)$$

where $\Delta T_{SC,min}$ and $T_{sat,max}$ are respectively the constraints on evaporator inlet minimum sub-cool and evaporator outlet maximum temperature, defined according to Table 7.2, and $k_{T_{in}}$ and $k_{T_{out}}$ are some weighting coefficients.

To attain satisfactory controller performance, n-MPC parameters must be tuned mainly to guarantee stability, optimality and to comply with the presence of unknown external disturbances. The presence of unpredictable fluctuations in engine conditions represents in fact a primary problem in the application of the algorithm, as the real system evolution (subject to the action of time-varying disturbances) may significantly diverge from the predicted trajectory, thus making ineffective the control action. To limit this issue, a high sample frequency has been adopted, as the punctual knowledge of the real system evolution (obtained from the measured state $x(t_k)$) gives a feedback to the optimization algorithm which can now take the necessary countermeasures to adjust the evolution of the system. A sample time of 1 s has been found to be the optimal compromise between algorithm efficiency and computation time. The control horizon N_C has been set equal to 4 to keep a low dimension search space for the PSO algorithm and make sure that the response of the algorithm is sufficiently fast. Finally, a prediction horizon N_P of 20 steps has been adopted, to guarantee the stability of the algorithm (see Chapter 4.3). An optimal compromise between algorithm precision and computation time can be found by tuning the PSO parameters, with the adoption of a population of 40 particles (a number high enough to allow the proper exploration of the search space), and limiting the number of iteration to 6.

With the proper tuning of the MPC and PSO parameters the algorithm is able to evaluate the optimal input trajectory, which maximize the net energy output harvested from the system, in a short amount of time and compensating the effect of the external disturbances. Figure 7.22 shows an example of the application of the algorithm, depicting the evolution of pump trajectories as the algorithm progresses. In the first iteration of the algorithm, the trajectories are spread out over the whole

search space and then, with every new iteration, they begin to overlap and converge towards the swarm global best. Even if not perfect convergence is attained, experiments show that, as only the first step will be implemented in the receding horizon fashion, the optimality degree of the obtained trajectory is satisfactory enough.

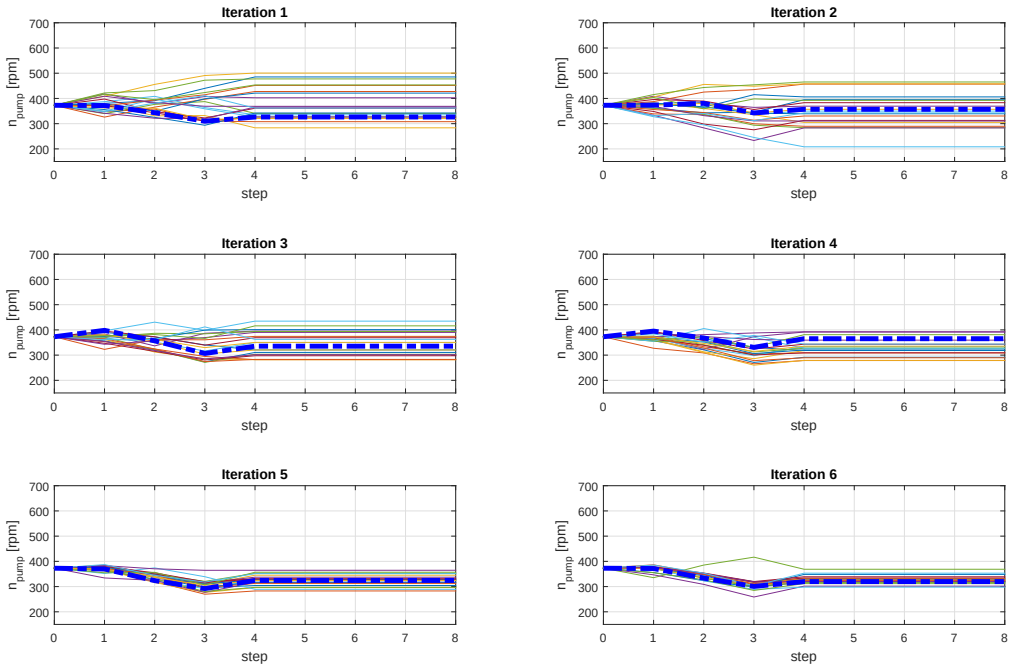


Figure 7.22: Evolution of pump trajectory during PSO optimization

7.4.3 Static PI based controller

To provide a benchmark against which the performance of the developed n-MPC controller can be compared, a simpler control policy is developed, based on the classic feed-forward/feedback structure. The feed-forward part of the controller provides optimal steady states values for control inputs, i.e. pump speed and exhaust bypass, once the engine operating point is known. To derive these correlations, engine disturbances are gridded over a range of operating conditions and the optimal steady state values of control inputs are evaluated, for each point in the grid, by running the PSO algorithm as a static optimizer, with the ORC system subject to constant external disturbances. The objective function to be maximized is still the net power output

and the same constraints depicted in Table 7.2 are accounted. The test grid includes 64 combinations of 4 different exhausts flow rate conditions ([7, 10, 20, 30] g/s), 4 different exhausts temperatures ([780, 830, 880, 930] K), and 4 different engine speeds ([1000, 2000, 2500, 3000] rpm). Optimal steady state values of evaporator pressure and outlet temperature are also evaluated for each point of the grid. As depicted in Figure 7.23, the optimal steady state values for control inputs are mapped into lookup tables as function of the measured exogenous disturbances w. The open-loop optimal input value is then corrected with the help of two PI controllers. The first one adjusts the value of pump speed, n_{pump}, based on the error between measured and steady-state optimal values of evaporator pressure, while the second controller adjusts the exhausts bypass, x_{bypass}, based on the temperature error. Moreover the real inputs to the system accounts for the controller saturations.

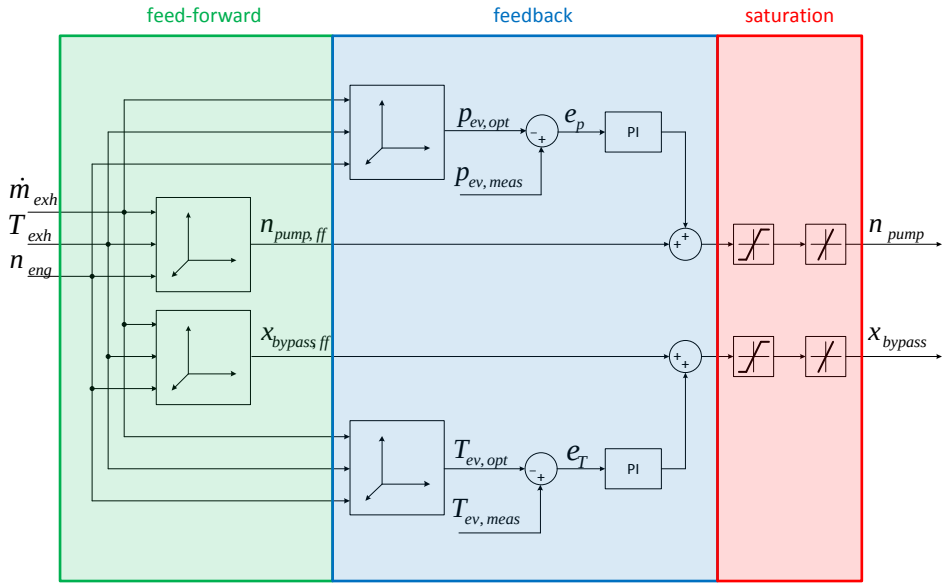
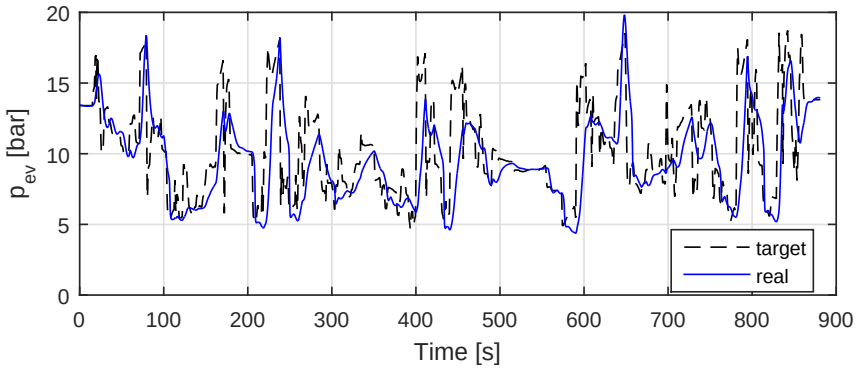
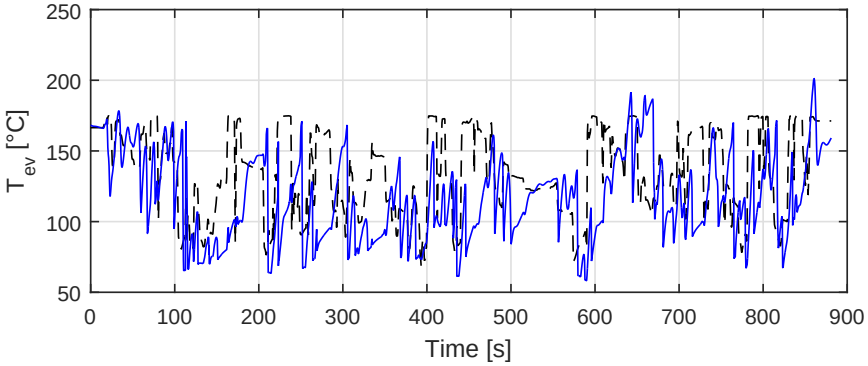


Figure 7.23: Schematic of basic benchmark controller

Figure 7.24 depicts the tracking performances of the PI controller during the analyzed driving transient; it is noticeable that evaporator pressure can be controlled with sufficient accuracy, while evaporator outlet temperature, due to the high thermal inertias of the system, cannot be properly regulated to track the optimal conditions.



(a) Evaporator pressure



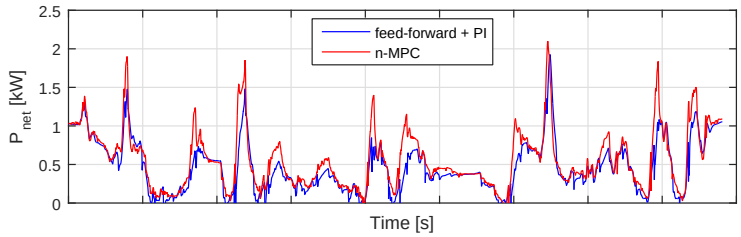
(b) Evaporator outlet temperature

Figure 7.24: Basic control strategy performances

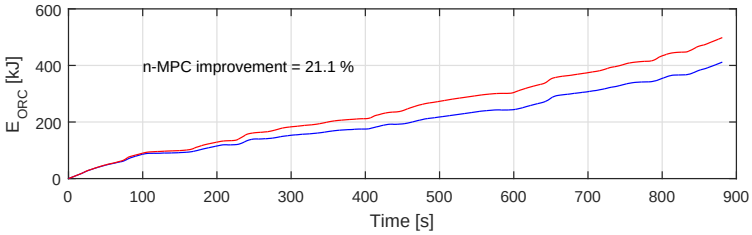
7.5 Simulation results and comparison

The developed n-MPC algorithm and the feed-forward/PI policy are used to control the system during the transient depicted in Figure 7.20. The net power produced and the amount of energy harvested from the system are depicted in Figure 7.25 for both the cases.

When the system approaches steady-state conditions (e.g. at $t = 530\ s$), the performance of the two controllers, based on the amount of net power produced, are almost identical. What on the other hand really makes the difference, is the ability of the n-MPC controller to efficiently and promptly make use a sudden increase of the available heat (e.g. at $t = 220\ s$ or $t = 600\ s$). In these situations, the n-MPC controller exploits its prediction ability to find the input trajectory which leads to the maximum harvesting of energy during the transient. It is worth to notice that this trajectory does not correspond to that which will quickly bring the system to optimal



(a) Net power output



(b) Net energy output

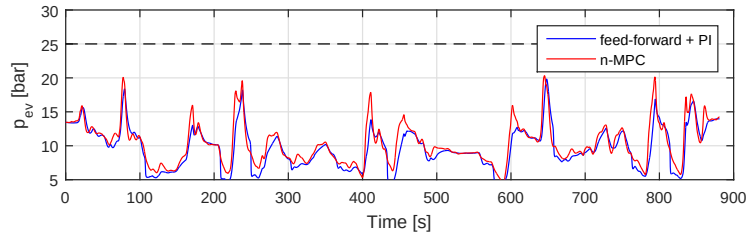
Figure 7.25: Performance comparison over energy harvesting

steady state conditions. The adoption of a dynamic optimal control strategy leads to a consistent increase in the amount of harvested energy from the system, with an overall increase of 21 %.

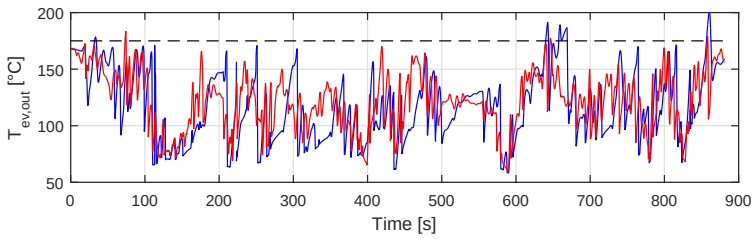
The trajectories of the evaporator pressure are depicted in Figure 7.26a for both the policies; the constraint of maximum pressure never poses a problem due to the relatively low heat input during the entire transient, but it is still interesting to notice that with the n-MPC policy the pressure of the system increases more rapidly during fast transients. This is presumably due to the fact that, during fast transients, it's more convenient to increase the pressure of the system (whose dynamic is faster than the temperature one) to benefit of the more prompt increase in net power output. At steady-state conditions on the other hand, it is rewarding to reduce the pressure and increase the temperature at the outlet of the evaporator, to achieve the best thermodynamic efficiency.

Looking at evaporator outlet temperature (Figure 7.26b) it is noticeable how the n-MPC controller better handles the rare occasions where maximum temperature constraint is violated. The same thing can be noticed, even to a wider extent, if looking at the degree of sub-cooling at evaporator inlet depicted in Figure 7.26c. In this case the basic control policy systematically leads to a systematic violation of this constraint, while the n-MPC algorithm is able to easily prevent this situation. This is due to the fact that, while steady-state target for the feed-forward algorithm are

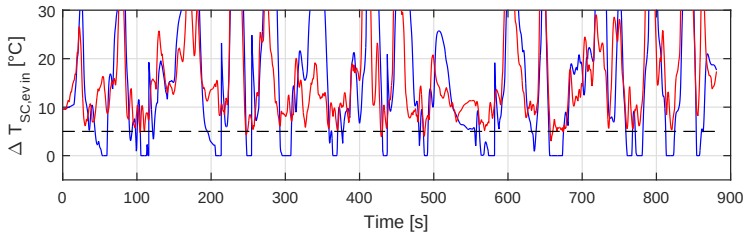
inside the feasibility region, during the transients this condition is not guaranteed due to the limited control authority of the PI feedback loop, which can only try to track the evaporator pressure and outlet temperature. The n-MPC algorithm on the other hand is perfectly capable of selecting the conjoint control input action which is able to guarantee the satisfaction of the evaporator inlet sub-cool constraint.



(a) Evaporator pressure



(b) Evaporator outlet temperature



(c) Evaporator inlet sub-cool degree

Figure 7.26: ORC key parameters

As depicted in Figure 7.27, the n-MPC algorithm is also able to keep an average higher evaporator outlet pseudo-quality χ, with an associated lower usage of the expander bypass valve. Due to the relatively low heat input from the engine, even if using an optimal control strategy, a two phase fluid outlet from the evaporator is still present during some phases of the complete driving cycle.

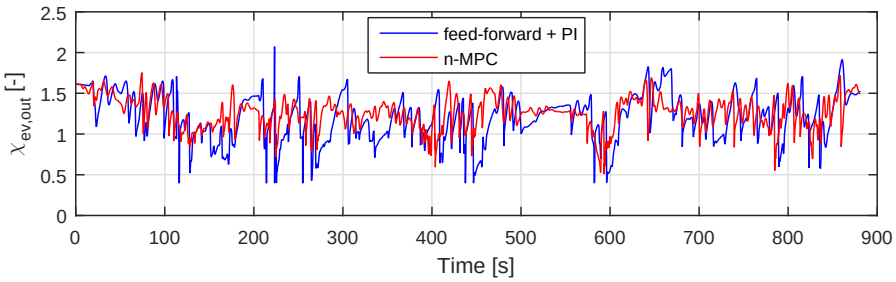


Figure 7.27: Evaporator outlet pseudo-quality

Finally, an estimation on fuel consumption improvements is carried out, based on the assumption that the small amount of power produced by the ORC does not significantly affect the engine operating conditions. Under this assumption, engine brake efficiency η would remain equal to the base solution and may be expressed as:

$$\eta = \frac{T \cdot \omega}{\dot{m}_{fuel} \cdot LHV} \tag{7.36}$$

where T is the engine effective torque in $[Nm]$, ω is the engine speed in $[rad/s]$, $\dot{m}_{fuel}$ is the fuel mass flow rate in $[kg/s]$, and LHV is the fuel Lower Heating Value in $[J/kg]$. If P_{ORC} is the net power produced by the ORC plant, then the mass of fuel saved thanks to the WHR system may be expressed as:

$$m_{fuel,saved} = \int_0^t \frac{T \cdot \omega - P_{ORC}}{\eta \cdot LHV} dt \tag{7.37}$$

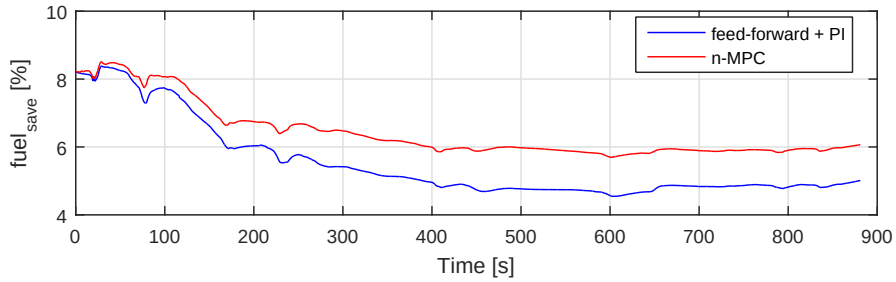


Figure 7.28: Fuel savings

As depicted in Figure 7.28, with the adoption of the n-MPC control algorithm, it is possible to achieve fuel savings percentages of about 6.0 %, over the entire driving manoeuvre, compared to the $\sim$ 4.9 % that can be obtained with the basic feed-forward/PI control policy.

7.6 Conclusions

This chapter presented the development of a nonlinear economic Model Predictive Control algorithm for the dynamic optimization of an Organic Rankine Cycle based Waste Heat Recovery plant for light vehicle applications. The controller relies on a detailed physical model of the plant, based on a switching Moving Boundary Methodology, which is able to predict the dynamic response of the system in presence of large variations in control inputs and external disturbances. Thanks to this model, the knowledge of the system dynamic response may be exploited by the control algorithm to predict the input trajectory which maximizing the net energy harvested from the plant. The arising constrained optimal control problem deriving from a mathematical formulation of the problem is solved by means of a Particle Swarm based nonlinear Model Predictive Controller. The results of the proposed solution are finally compared to those from a benchmark basic control policy, where the set-points of the actuators are obtained from a static optimization of the plant and the optimal steady conditions for plant parameters are tracked by means of two PIs. Both the algorithms are tested over a realistic transient profile, obtained from experimental data collected on a test vehicle. The results of the comparison demonstrate the effectiveness of the n-MPC algorithm to maximize the amount of energy harvested from the engine, providing an increase of 21 % in fuel economy when compared to the basic policy. The n-MPC algorithm is also capable of guaranteeing safe operating conditions for the plant, successfully enforcing the respect of a set of plant parameters constraints, while the basic controller on the other hand frequently fails to do so.

Regarding the possibility of implementing the control policy on-line, the n-MPC strategy strongly suffers from the long computation time required for the solution of the constrained optimal control problem. This is mainly due to the excessive complexity of the plant model. In fact, even if the PSO based algorithm is not affected by the curse of dimensionality (i.e. the computation time is not directly affected by the number of states in the model), the high number of states in the evaporator model still poses a significant problem as it increases the computation time for the simulation of the plant response (due to the great number of equations to be evaluated). The usage of a reduced complexity plant model may be sufficient to decrease the computation

time and make the solution suitable for on-line implementation. Among the possible simplified models, the model order reduction approach followed by Peralez et al. [48] is worth of being cited as it is based on the same MBM approach adopted here, with the introduction of some physical simplifying hypothesis that help to capture only the dominant dynamics of the model. The PSO based metaheuristic approach to the constrained optimal problem solution on the other hand still retains its benefits even regarding on-line applications, as it is able to provide a robust gradient free optimization tool, which does not require an excessive number of calls of the objective function in order to provide the solution.

Even if the complexity of the MBM approach may be excessive for on-line implementation, the developed physical based model is still of primary importance as it provides a reliable model for the testing phase of the control strategies and a starting point for the creation of a reduced complexity control oriented model.

Nomenclature

Subscripts and superscripts

$cond$	condenser
$cool$	coolant
ev	evaporator
exh	exhausts
rec	recuperator
ref	refrigerant
sat	saturation
s	isentropic
v	volumetric
’	saturated liquid
”	saturated vapor
*	optimal

Abbreviations and acronyms

CV	Control Volume
DAE	Differential Algebraic Equation
IO	input/output
MBM	Moving Boundary Method
MPC	Model Predictive Control
ORC	Organic Rankine Cycle
PSO	Particle Swarm Optimization
SC	Sub-Cooled
SI	Spark Ignition
SH	Super-Heated
TP	Two Phase
WHR	Waste Heat Recovery

Symbols

A	area	$[m^2]$	$\dot{Q}$	heat transfer rate	$[W]$
c	specific heat	$[^J/_{kgK}]$	s	specific entropy	$[^J/_{kgK}]$
C	heat capacity rate	$[^W/_K]$	t	time	$[s]$
h	specific enthalpy	$[^J/_{kg}]$	T	temperature	$[K]$
J	objective function	$[-]$		torque	$[Nm]$
L	length	$[m]$		sample time	$[s]$
$\dot{m}$	mass flow rate	$[^{kg}/_s]$	u	control input	$[-]$
M	mass	$[kg]$	U	heat transfer coeff.	$[^W/_{m^2K}]$
n	rotational speed	$[^{rev}/_{min}]$	V	volume	$[m^3]$
NTU	Net Transfer Units	$[-]$	V_d	displacement	$[^{m^3}/_{rev}]$
p	pressure	$[Pa]$	w	disturbance	$[-]$
P	power	$[W]$	w_r	relative velocity	$[^m/_s]$

x	system state	$[-]$
	thermodynamic quality	$[-]$
y	system output	$[-]$
Δ	difference	$[-]$
γ	void fraction	$[-]$
ε	effectiveness	$[-]$
η	efficiency	$[-]$
μ	slip factor	$[-]$
ρ	density	$[kg/m^3]$
χ	fluid pseudo-quality	$[-]$
ω	angular velocity	$[rad/s]$

Bibliography

[1] H. Aghaali and H.-E. Å ngström, "A review of turbocompounding as a waste heat recovery system for internal combustion engines," *Renewable and Sustainable Energy Reviews*, vol. 49, pp. 813–824, Sep. 2015. 147

[2] W. Akers, H. Deans, and O. Crosser, "Condensation Heat Transfer within Horizontal Tubes," *Chem. Eng. Progr.*, 1959. 166

[3] M. Aneke, B. Agnew, C. Underwood, H. Wu, and S. Masheiti, "Power generation from waste heat in a food processing application," *Applied Thermal Engineering*, vol. 36, pp. 171–180, Apr. 2012. 150

[4] S. W. Angrist, *Direct energy conversion.* Boston: Allyn and Bacon, 1976. 148

[5] S. Bendapudi, J. E. Braun, and E. A. Groll, "A comparison of moving-boundary and finite-volume formulations for transients in centrifugal chillers," *International Journal of Refrigeration*, vol. 31, no. 8, pp. 1437–1452, Dec. 2008. 158

[6] J. Bonilla, S. Dormido, and F. E. Cellier, "Switching moving boundary models for two-phase flow evaporators and condensers," *Communications in Nonlinear Science and Numerical Simulation*, vol. 20, no. 3, pp. 743–768, 2015. 158

[7] R. Bracco, S. Clemente, D. Micheli, and M. Reini, "Experimental tests and modelization of a domestic-scale ORC (Organic Rankine Cycle)," *Energy*, pp. 1–12, 2013. 152

[8] T. E. Briggs, R. Wagner, K. D. Edwards, S. Curran, and E. Nafziger, "A Waste Heat Recovery System for Light Duty Diesel Engines," *SAE Technical Paper 2010-01-2205*, 2010. 152, 153, 173

[9] F. Campana, M. Bianchi, L. Branchini, A. De Pascale, A. Peretto, M. Baresi, A. Fermi, N. Rossetti, and R. Vescovo, "ORC waste heat recovery in European energy intensive industries: Energy and GHG savings," *Energy Conversion and Management*, vol. 76, pp. 244–252, Dec. 2013. 150

[10] M. Canova, A. Gambarotta, M. Crialesi Esposito, K. Laboe, and M. Zajac, "Lumped-parameter modeling and analysis of automotive waste heat recovery systems based on an organic rankine cycle," in *14th Internationales Stuttgarter Symposium*, 2014, pp. 1309–1324. 153, 155, 172

[11] F. Chiara and M. Canova, "A review of energy consumption, management, and recovery in automotive systems, with considerations of future trends," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 227, no. 6, pp. 914–936, 2013. 146, 148

[12] Committee on the Assessment of Technologies for Improving Light-Duty Vehicle Fuel Economy, *Assesment of fuel economy technologies for light-duty vehicles*, National Academies Press, Ed., Washington, DC, 2011. 146

[13] E. Cortona, C. H. Onder, and L. Guzzella, "Engine thermomanagement with electrical components for fuel consumption reduction," *International Journal of Engine Research*, vol. 3, no. 3, pp. 157–170, Jan. 2002. 146

[14] M. Crialesi Esposito, N. Pompini, A. Gambarotta, and M. Canova, "A switching Moving Boundary Model for the simulation of ORC plants in automotive applications," in *15th Internationales Stuttgarter Symposium*, 2015, pp. 735–753. 156

[15] D. Cross and C. Brockbank, "Mechanical Hybrid System Comprising a Flywheel and CVT for Motorsport and Mainstream Automotive Applications," in *SAE Technical Paper 2009-01-1312*, Apr. 2009. 146

[16] T. Endo, S. Kawajiri, Y. Kojima, K. Takahashi, T. Baba, S. Ibaraki, T. Takahashi, and M. Shinohara, "Study on Maximizing Exergy in Automotive Engines," in *SAE Technical Paper 2007-01-0257*, Apr. 2007. 150

[17] N. Espinosa, I. Gil-Roman, D. Didiot, V. Lemort, B. Lombard, and S. Quoilin, "Transient Organic Rankine cycle Modelling for Waste Heat Recovery on a Truck," in *24th International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems*, Novi Sad, Serbia, 2011. 150

[18] E. Feru, F. Willems, B. D. Jager, and M. Steinbuch, "Model Predictive Control of a Waste Heat Recovery System for Automotive Diesel Engines," in *18th International Conference on System Theory, Control and Computing*, Sinaia, Romania, 2014, pp. 658–663. 173

[19] J. Freeman, K. Hellgardt, and C. N. Markides, "An assessment of solar-powered organic Rankine cycle systems for combined heating and power in UK domestic applications," *Applied Energy*, vol. 138, pp. 605–620, Jan. 2015. 150

[20] M. Habka and S. Ajib, "Evaluation of mixtures performances in Organic Rankine Cycle when utilizing the geothermal water with and without cogeneration," *Applied Energy*, vol. 154, pp. 567–576, Sep. 2015. 150

[21] A. J. J. Hardy, J. B. Heywood, and T. E. Kenney, "Fuel Economy Benefits and Aftertreatment Requirements of a Naturally Aspirated HCCI-SI Engine System," *SAE International Journal of Engines*, vol. 1, no. 1, pp. 2008–01–2512, Oct. 2008. 146

[22] A. Hernandez, A. Desideri, C. Ionescu, S. Quoilin, V. Lemort, and R. De Keyser, "Towards the optimal operation of an organic Rankine cycle unit by means of model predictive control," in *Proceedings of the 3rd International Seminar on ORC Power Systems*, Brussels, Belgium, 2015, pp. 1–10. 173

[23] H. Hiereth and P. Prenninger, *Charging the Internal Combustion Engine*. Vienna: Springer Vienna, 2007. 147

[24] U. Hopmann and M. C. Algrain, "Diesel Engine Electric Turbo Compound Technology," in *SAE Technical Paper 2003-01-2294*, Jun. 2003. 148

[25] T. Hosaka and M. Hamazaki, "Development of the Variable Valve Timing and Lift (VTEC) Engine for the Honda NSX," in *SAE Technical Paper 910008*, Jan. 1991. 146

[26] G. Hou, R. Sun, G. Hu, and J. Zhang, "Supervisory predictive control of evaporator in Organic Rankine Cycle (ORC) system for waste heat recovery," in *The 2011 International Conference on Advanced Mechatronic Systems*, Zhengzhou, China, 2011, pp. 306–311. 173

[27] T. Howell, "Development of an ORC system to improve HD truck fuel efficiency," in *2011 directions in engine efficiency and emissions research conference*, Detroit, Michigan USA, 2011. 150

[28] Q. Hussain, "Organic rankine cycle for light duty passenger vehicles," in *2011 directions in engine efficiency and emission research conference*, Detroit, Michigan USA, 2011. 150

[29] F. P. Incropera, D. P. DeWitt, T. L. Bergman, and A. S. Lavine, *Introduction to Heat Transfer*. John Wiley & Sons, 2007. 155, 165

[30] J. M. Jensen, "Dynamic Modeling of Thermo-Fluid Systems," Ph.D. dissertation, Technical University of Denmark, 2003. 160, 162

[31] O. Kaka, "Energy and exergy analysis of an organic Rankine for power generation from waste heat recovery in steel industry," *Energy Conversion and Management*, vol. 77, pp. 108–117, Jan. 2014. 150

[32] S. Karellas, A. D. Leontaritis, G. Panousis, E. Bellos, and E. Kakaras, "Energetic and exergetic analysis of waste heat recovery systems in the cement industry," *Energy*, vol. 58, pp. 147–156, Sep. 2013. 150

[33] M. A. Kluger and D. M. Long, "An Overview of Current Automatic, Manual and Continuously Variable Transmission Efficiencies and Their Projected Future Improvements," in *SAE Technical Paper 1999-01-1259*, Mar. 1999. 146

[34] J. E. Körner and T. Kobs, "Waste heat recovery : Low-temperature heat recovery using the Organic-Rankine-Cycle," in *13th Stuttgart International Symposium - Automotive and Engine Technology*, 2013. 151

[35] R. Kruiswyk, "An engine system approach to exhaust waste heat recovery," in *2010 directions in engine efficiency and emissions research conference*, Detroit, Michigan USA, 2011. 150

[36] K. Laboe and M. Canova, "Powertrain Waste Heat Recovery: A Systems Approach to Maximize Drivetrain Efficiency," in *ASME 2012 Internal Combustion Engine Division Spring Technical Conference*. ASME, May 2012, p. 985. 146

[37] T. Lancefield, "The Influence of Variable Valve Actuation on the Part Load Fuel Economy of a Modern Light-Duty Diesel Engine," in *SAE Technical Paper 2003-01-0028*, Mar. 2003. 146

[38] B. Li and A. G. Alleyne, "A dynamic model of a vapor compression cycle with shut-down and start-up operations," *International Journal of Refrigeration*, vol. 33, no. 3, pp. 538–552, May 2010. 158

[39] J. MacArthur and E. Grald, "Unsteady compressible two-phase flow model for predicting cyclic heat pump performance and a comparison with experimental data," *International Journal of Refrigeration*, vol. 12, no. 1, pp. 29–41, Jan. 1989. 166

[40] P. J. Mago, L. M. Chamra, and C. Somayaji, "Performance analysis of different working fluids for use in organic Rankine cycles," *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy*, vol. 221, no. 3, pp. 255–263, Jan. 2007. 148

[41] P. R. Manjunath and R. Ortmann, "MOTRONIC - MED7 Gasoline Direct Injection Technology for Reduction of Fuel Consumption and Emissions," in *SAE Technical Paper 2003-26-0020*, Jan. 2003. 146

[42] T. L. McKinley and A. G. Alleyne, "An advanced nonlinear switched heat exchanger model for vapor compression cycles using the moving-boundary method," *International Journal of Refrigeration*, vol. 31, no. 7, pp. 1253–1264, Nov. 2008. 158, 167, 172

[43] C. Mi, M. A. Masrur, and D. W. Gao, *Hybrid Electric Vehicles*. Chichester, UK: John Wiley & Sons, Ltd, Jun. 2011. 146

[44] J. M. Miller, *Propulsion Systems for Hybrid Vehicles*. Stevenage, Hertfordshire: Institution of Engineering and Technology, 2004. 146

[45] A. Neeraj, F. Chiara, and M. Canova, "Control-Oriented Modeling of an Automotive Thermal Management System," in *2012 IFAC Workshop on Engine and Powertrain Control, Simulation and Modeling*, Oct. 2012, pp. 392–399. 155, 156

[46] C. Nelson, "Exhaust energy recovery," in *2009 directions in engine efficiency and emissions research conference*, Detroit, Michigan USA, 2009. 150

[47] P. S. Patel and E. F. Doyle, "Compounding the Truck Diesel Engine with an Organic Rankine-Cycle System," in *SAE Technical Paper 760343*, Feb. 1976. 150

[48] J. Peralez and P. Tona, "Towards model-based control of a steam Rankine process for engine waste heat recovery," in *2012 IEEE Vehicle Power and Propulsion Conference*, 2012, pp. 289–294. 158, 188

[49] J. Peralez, P. Tona, O. Lepreux, A. Sciarretta, L. Voise, P. Dufour, and M. Nadri, "Improving the Control Performance of an Organic Rankine Cycle System for Waste Heat Recovery from a Heavy-Duty Diesel Engine using a Model-Based Approach," in *52nd IEEE Conference on Decision and Control*, Florence, Italy, 2013, pp. 6830–6836. 173

[50] S. Quoilin, R. Aumann, A. Grill, A. Schuster, V. Lemort, and H. Spliethoff, "Dynamic modeling and optimal control strategy of waste heat recovery Organic Rankine Cycles," *Applied Energy*, vol. 88, no. 6, pp. 2183–2190, Jun. 2011. 157

[51] J. B. Rawlings and D. Q. Mayne, *Model Predictive Control: Theory and Design*. Nob Hill Publishing, LLC, 2012. 36, 62, 63, 174

[52] J. Ringler, M. Seifert, V. Guyotot, and W. Hübner, "Rankine Cycle for Waste Heat Recovery of IC Engines," *SAE International Journal of Engines*, vol. 2, no. 1, pp. 2009–01–0174, Apr. 2009. 150

[53] E. Schneider, J. Müller, M. Leesch, and R. Resch, "Nine-speed hybrid dual-clutch transmission for high-performance front-transverse powertrains," *MTZ worldwide*, vol. 72, no. 6, pp. 26–31, Jun. 2011. 146

[54] S. Singh, "Exhaust heat driven rankine cycle for a heavy duty diesel engine," in *2011 directions in engine efficiency and emissions research conference*, Detroit, Michigan USA, 2011. 150

[55] E. R. Toulson, "Evaluation of a hybrid hydraulic launch assist system for use in small road vehicles," in *2008 IEEE International Symposium on Industrial Electronics*. IEEE, Jun. 2008, pp. 967–972. 146

[56] M. Uris, J. I. Linares, and E. Arenas, "Size optimization of a biomass-fired cogeneration plant CHP/CCHP (Combined heat and power/Combined heat, cooling and power) based on Organic Rankine Cycle for a district network in Spain," *Energy*, vol. 88, pp. 935–945, Aug. 2015. 150

[57] O. Varnier, "Trends and Limits of Two-Stage Boosting Systems for Automotive Diesel Engines," Ph.D. dissertation, Universitad Politecnica de Valencia, 2012. 147

[58] C. Vuk, "Turbo compounding, a technology whose time has come," in *Diesel Engine Emissions Reduction (DEER) Conference*, 2005. 147

[59] W. Weerasinghe, R. Stobart, and S. Hounsham, "Thermal efficiency improvement in high output diesel engines a comparison of a Rankine cycle with turbocompounding," *Applied Thermal Engineering*, vol. 30, no. 14-15, pp. 2253–2256, Oct. 2010. 147

[60] G. Xiaojun and L. Liansheng, "Research on a scroll expander used for recovering work in a fuel cell," *International Journal of Thermodynamics*, vol. 7, no. 1, pp. 1–8, 2004. 151

[61] Y. Y. Yan and T. F. Lin, "Evaporation Heat Transfer and Pressure Drop of Refrigerant R-134a in a Plate Heat Exchanger," *Journal of Heat Transfer*, vol. 121, no. 1, p. 118, 1999. 165

[62] J. Zhang, Y. Zhou, Y. Li, G. Hou, and F. Fang, "Generalized predictive control applied in waste heat recovery power plants," *Applied Energy*, vol. 102, pp. 320–326, 2013. 173

[63] J. Zhang, Y. Zhou, R. Wang, J. Xu, and F. Fang, "Modeling and constrained multivariable predictive control for ORC (Organic Rankine Cycle) based waste heat energy conversion systems," *Energy*, vol. 66, pp. 128–138, Mar. 2014. 158

[64] W. J. Zhang and C. L. Zhang, "A generalized moving-boundary model for transient simulation of dry-expansion evaporators under larger disturbances," *International Journal of Refrigeration*, vol. 29, no. 7, pp. 1119–1127, Nov. 2006. 158