Electric Power Generation, Transmission and Efficiency

Contributors

Gianfranco Chicco	Armando C. Oliveira
B. Chudnovsky	M. Perelman
Alexandre de Assis Mota	Saffa B. Riffat
Zhengwei Fan	R. Saveliev
G. P. Granelli	N. Spitz
E. Korytni	Yuehong Su
Y. H. Luo	Szabolcs Varga
Pierluigi Mancarella	Q. C. Wang
M. Montagna	Tao Lin Xialing Xu
Lia Toledo Moreira Mota	E. Bar-Ziv

Clément M. Lefebvre
Editor

NOVA

ELECTRIC POWER: GENERATION, TRANSMISSION AND EFFICIENCY

ELECTRIC POWER: GENERATION, TRANSMISSION AND EFFICIENCY

CLÉMENT M. LEFEBVRE
EDITOR

Nova Science Publishers, Inc.
New York

For permission to use material from this book please contact us:
Telephone 631-231-7269; Fax 631-231-8175
Web Site: http://www.novapublishers.com

NOTICE TO THE READER

The Publisher has taken reasonable care in the preparation of this book, but makes no expressed or implied warranty of any kind and assumes no responsibility for any errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of information contained in this book. The Publisher shall not be liable for any special, consequential, or exemplary damages resulting, in whole or in part, from the readers' use of, or reliance upon, this material.

Independent verification should be sought for any data, advice or recommendations contained in this book. In addition, no responsibility is assumed by the publisher for any injury and/or damage to persons or property arising from any methods, products, instructions, ideas or otherwise contained in this publication.

This publication is designed to provide accurate and authoritative information with regard to the subject matter covered herein. It is sold with the clear understanding that the Publisher is not engaged in rendering legal or any other professional services. If legal or any other expert assistance is required, the services of a competent person should be sought. FROM A DECLARATION OF PARTICIPANTS JOINTLY ADOPTED BY A COMMITTEE OF THE AMERICAN BAR ASSOCIATION AND A COMMITTEE OF PUBLISHERS.

CONTENTS

PREFACE

This book presents new and important research on electric power and its generation, transmission and efficiency. The world is becoming increasingly electrified. For the foreseeable future, coal will continue to be the dominant fuel used for electric power production. The low cost and abundance of coal is one of the primary reasons for this. Electric power transmission, a process in the delivery of electricity to consumers, is the bulk transfer of electrical power. Typically, power transmission is between the power plant and a substation near a populated area. Electricity distribution is the delivery from the substation to the consumers. Due to the large amount of power involved, transmission normally takes place at high voltage (110 kV or above). Electricity is usually transmitted over long distance through overhead power transmission lines. Underground power transmission is used only in densely populated areas due to its high cost of installation and maintenance, and because the high reactive power gain produces large charging currents and difficulties in voltage management.

A power transmission system is sometimes referred to colloquially as a "grid"; however, for reasons of economy, the network is rarely a true grid. Redundant paths and lines are provided so that power can be routed from any power plant to any load center, through a variety of routes, based on the economics of the transmission path and the cost of power. Much analysis is done by transmission companies to determine the maximum reliable capacity of each line, which, due to system stability considerations, may be less than the physical or thermal limit of the line. Deregulation of electricity companies in many countries has led to renewed interest in reliable economic design of transmission networks.

Short Communication - In recent years, both utilities and the end-users of electric power are becoming increasingly concerned with power quality issues. In order to correctly evaluate the quality of power supply and accurately measure the levels of various power disturbances in current and/or voltage waveforms, additional two aspects of power quality studies have been proposed besides the detection of power disturbance, which are the classification or identification of the types of power disturbances and the measurement of the parameters of power disturbance waveforms. In this chapter, an advanced signal processing technique and artificial intelligence are introduced in this area to realize the functions mentioned above. In details, the Wavelet Transform is used to estimate the parameters of power disturbance and extract the pattern features for classification/identification; on the other hand, artificial neural network is used to classify/identify the types of power disturbances in voltage/current waveforms. Further, the performance of various wavelet functions and artificial neural

networks are investigated. As a result, the combination use of the "Dmey" wavelet and the back propagation (BP) neural network is proposed for archiving desirable performance. A software package was developed by integrating the proposed techniques, and large amount of simulation results obtained using this software package proved the correctness and effectiveness of the applications of the advanced signal processing technique and artificial intelligence to power quality researches.

Chapter 1 - One of the major drivers of the electrical energy system evolution is the widespread adoption of emerging technologies for distributed generation, that are shifting the focus from the centralized production to the local production of electricity. Under the distributed generation paradigm, the present research scenario is more and more emphasising the role of solutions aimed at improving the energy generation efficiency and thus the sustainability of the overall energy sector. From this point of view, the development of multi-generation solutions for small-scale applications (below 1 MW), for instance producing at the same time electricity, heat and cooling power, represents a key asset for improving the performance of the future power system. In fact, the combined production of manifold energy vectors can bring several benefits in terms of energy saving and CO_2 emission reduction, as well as potential enhanced profitability of the plants exploiting the energy mix within the liberalised electricity market framework.

This chapter illustrates a comprehensive approach to multi-generation system characterization and planning. This approach is formulated in terms of the so-called *lambda analysis*, consisting of a unified framework to study multi-generation systems, that extends the classical models based on the analysis of the heat-to-power cogeneration ratio in cogeneration plants. In particular, the representation of the energy interactions within the multi-generation plant is summarized in terms of the transformation of a vector of original energy or cogeneration ratio values into an equivalent set of values, mathematically expressed by means of specifically defined *lambda transforms*. The conceptual scheme presented provides effective characterization and modelling of the production side, the demand side and their interactions in multi-generation systems.

The details of the approach presented are illustrated by reviewing the bulk of alternative schemes and equipment available on the market for setting up multi-generation plants. For each alternative, the suitable equipment models and the expressions of the relevant lambda transforms are presented.

Numerical applications are provided, referred to a multi-generation system for electrical, thermal, and cooling power production. The results highlight the potential of the lambda analysis framework and of the associated lambda transforms as an effective tool to assist the energy system planner.

The availability of such a synthetic and powerful tool is of utmost importance in order to effectively cope with the increasing complexity of the future electro-energetic systems, in which the efficiency enhancement will strongly depend on the integration of the equipment for local combined production of manifold energy vectors.

Chapter 2 - Genetic algorithms, proposed about 40 years ago, have been used as a general purpose optimization technique. In this work, the authors' experience with genetic algorithm based optimization is presented with reference to some electric power system analysis problems.

At first, the fundamentals of genetic algorithms are described: the basic genetic algorithm by John Holland is presented and the function of the three genetic operators of selection,

crossover and mutation is discussed. Among the more recent theoretical developments, the micro-genetic approach by K. Krishnakumar and the algorithm of Chu and Beasley are considered. The former takes advantage from operating with small-sized populations and the latter proposes an effective technique to deal with functional constraints.

The second part of this work is concerned with the description of some applications of the above mentioned genetic algorithm based procedures to power system problems. The topics that are considered in detail are:

- allocation of compensating capacitors in high and medium voltage networks to improve voltage regulation;
- optimization of the topology of EHV networks with the aim of improving security and to overcome the problem of parallel or loop flows; control measures include switching of substation breakers as well as deployment and operation of phase shifting transformers;
- identification of multiple interacting bad data in state estimation, formulated as a combinatorial optimization problem.

The above mentioned items represent a clearly non-exhaustive list of the many application fields where genetic algorithms have been profitably employed, but the authors feel they demonstrate the main advantage of using genetic algorithms with respect to other search methods. Indeed no present day optimizer is so general, robust and flexible to deal with problems so different from each other as the ones considered here.

Chapter 3 - A three-step methodology was developed to provide reliable prediction of a coal's behavior in a utility boiler: (1) Extracting the combustion kinetic model parameters by combining experimental data from a pilot-scale test facility, Computational Fluid Dynamic (CFD) codes and an artificial neural network. While the combustion kinetic parameters used in the model code will not correspond to the combustion rate of a single particle of coal, these parameters do describe the combustion behavior of a "macroscopic" sample of tested coal. (2) Validation of the combustion kinetic model parameters by comparing diverse experimental data with simulation results calculated with the same set of model parameters. (3) The model parameters are then used for simulations of full-scale boilers using the same CFD code. For operational engineering information needed by the utility operator, the authors apply the predicted results to EXPERT SYSTEM, a boiler supervision system developed by Israel Electric Corporation (IEC). Four different bituminous and sub-bituminous coals with known behavior in IEC 550MW opposite-wall and 575MW tangential-fired boilers were used to show the adequacy of the methodology. The predictions are done with the CFD code, GLACIER, propriety of Reaction Engineering International (REI). Preconfigured GLACIER models of the test and full-scale furnaces were purchased from REI and validated by our group. This book chapter will include a detailed description of the methodology, test furnace facility and an example of the experimental and predictive combustion results from the four coals used to test the methodology. In addition, two previously unknown coals will be examined prior to their firing in the utility boilers and prediction of their behavior and operational parameters in the two boilers will be carried out.

Chapter 4 - Electrogasdynamic (EGD) power conversion is a process that converts thermal (internal/kinetic) energy into electric energy, without moving parts. Discrete particles are charged using a corona electrode and transported at high velocity against an electric field,

to produce useful electric power. An important advantage of such a device is that it doesn't include moving parts and so requires very little maintenance.

The basic equations for EGD power conversion are presented, as well as theoretical results for the process. The efficiency of EGD power conversion is calculated for different fluids. The existence of practical limits for conversion is discussed: electric breakdown strength and charge-to-mass ratio. Theoretical results for different working fluids are presented.

A Computational Fluid Dynamics model was developed as a tool to simulate the major characteristics of fluid flow in such a device and to identify the most important factors in the power conversion process. Model results are presented and discussed. The importance of the particle/electric field interaction is evaluated, taking into account turbulent effects. A parametric study to identify the best collector location is carried out.

Experimental results for an EGD apparatus and different operating fluids were also obtained in a test rig. An EGD nozzle was designed, built and experimentally tested. Tests of electric breakdown were carried out for different working fluids (refrigerants). Results are presented and compared to theoretical values. The use of electrospray is also investigated.

Chapter 5 - Load modeling has a significant impact on power systems operation, simulation and analysis. However, little attention has been paid to develop adequate load models and forecasting methods when compared to the effort spent with other power systems related problems. In this context, this chapter presents a review on load models and load forecasting techniques, and also discusses the new trends on these issues. These late tendencies include bottom-up and top-down approaches, gray box identification techniques and the use of fuzzy logic, among others. The discussion emphasizes the main concepts of each method. Especially in distribution networks, load modeling and forecasting may cause some inconvenient because, in order to monitor the residential consumption (in terms of total active power and the type of appliance), many measuring equipment must be installed, leading to undesired costs. This is also troublesome to the residents of the household. In order to minimize these inconvenient, non-intrusive load modeling and forecasting techniques must be applied. This new methodology is also presented and examined in this chapter, in a combination of a bottom-up approach with object-oriented programming techniques.

Chapter 6 - The objectives of this work are to study the primary chemical structure of soot aerosol derived from lump-coal pyrolysis in different experimental conditions in fixed bed. A laboratory-scale movable fixed bed, water-cooled soot aerosol collection system, and electric reactor have been designed and used in the process. Three kinds of coals, sized at 3-5 mm, have been heated in the experiments. Fourier Transform Infrared Spectroscopy has been employed to test functional groups of soot aerosol samples. Infrared spectra from 400 to 4000 cm^{-1} and semiquantitative analysis have been employed. The results of experiments show that contents of hydrogen-bonded are increased, contents of unsaturated hydrocarbons are decreased, and contents of aromatic hydrocarbons are increased with temperature increase; contents of hydrogen-bonded are increased, contents of unsaturated hydrocarbons are increased, and contents of aromatic hydrocarbons are increased early and decreased late with residence time extension; and the contents of unsaturated hydrocarbons derived from soot aerosol samples are higher than those from original coal samples, and contents of hydrogen bonded and aromatic hydrocarbons are different depending on chemical structure of original coals.

Short Communication

POWER QUALITY STUDIES BASED ON ADVANCED SIGNAL PROCESSING TECHNIQUE AND ARTIFICIAL INTELLIGENCE

Tao Lin Xialing Xu and Zhengwei Fan
School of Electrical Engineering, Wuhan University, China

ABSTRACT

In recent years, both utilities and the end-users of electric power are becoming increasingly concerned with power quality issues. In order to correctly evaluate the quality of power supply and accurately measure the levels of various power disturbances in current and/or voltage waveforms, additional two aspects of power quality studies have been proposed besides the detection of power disturbance, which are the classification or identification of the types of power disturbances and the measurement of the parameters of power disturbance waveforms. In this chapter, an advanced signal processing technique and artificial intelligence are introduced in this area to realize the functions mentioned above. In details, the Wavelet Transform is used to estimate the parameters of power disturbance and extract the pattern features for classification/identification; on the other hand, artificial neural network is used to classify/identify the types of power disturbances in voltage/current waveforms. Further, the performance of various wavelet functions and artificial neural networks are investigated. As a result, the combination use of the "Dmey" wavelet and the back propagation (BP) neural network is proposed for archiving desirable performance. A software package was developed by integrating the proposed techniques, and large amount of simulation results obtained using this software package proved the correctness and effectiveness of the applications of the advanced signal processing technique and artificial intelligence to power quality researches.

Keywords: Power disturbance, Wavelet Transform, Artificial neural network.

1. INTRODUCTION

In recent years, microprocessor based control and electronic equipments are widely used in the world, which are sensitive to various power disturbances; on the other hand, the broad application of non-linear loads in power systems has also caused many negative impacts to electric power quality [1]. Therefore, both utilities and power users are becoming increasingly concerned with electric power quality issues. Once a power quality event occurs, besides the detection of power disturbances, the classification or identification of the types of power disturbances in a power quality event and the measurement of the parameters of power disturbance waveforms should also be performed, so as to correctly evaluate the quality of power supply and accurately measure the levels of various power disturbances in current and/or voltage waveforms, which are of significant helps to the mitigation of power quality issues.

Several signal processing and artificial intelligence methods have been used for classification/identification of the types of power disturbances, such as Expert System (ES) [2], Artificial Neural Networks (ANNs) [3], and the combination use of Wavelet Transform (WT) and ANNs. The disadvantage of the first method is that an ES becomes complicated and the searching efficiency decreases significantly while the types of power disturbances increase [4]; on the other hand, transplantation of this method from one case to another is not easy [5].The defect of the 2nd one is that it is relatively difficult to extract pattern features of various power disturbances directly.

The method based on the WT and ANN is studied in this chapter. Wavelet transform was used to locate the beginning and end of a power disturbance in the time domain, and extract pattern features of power disturbances for classification / identification using ANN. The parameters of various power disturbances were also estimated. A software package was developed by integrating the WT and ANN techniques, which was able to realize disturbance identification and parameter measurement.

2. BRIEF INTRODUCTIONS OF WAVELET TRANSFORM AND ARTIFICIAL NEURAL NETWORK

2.1. A Brief Introduction of Wavelet Transform

Wavelet transform is the inner product of the analyzed signal x(t) and a wavelet function $\psi_{a,\tau}(t)$, which is derived from mother wavelet function $\psi(t)$ using dilation and shift operations [6]:

$$\begin{aligned} WT_x(a,\tau) &= \left\langle x(t), \psi_{a,\tau}(t) \right\rangle \\ &= \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-\tau}{a} \right) dt \end{aligned} \tag{1}$$

where

$$\psi_{a,\tau}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-\tau}{a}\right)_.$$

In practical applications, mother wavelet function $\psi(t)$ should meet the admissibility and normality conditions, and should have limited supporting areas in the time-frequency plane. The Wavelet transform is usually conducted as a multi-resolution analysis via the well known Mallat algorithm.

2.2. A Brief Introduction of Artificial Neural Network

Artificial neural network is a kind of new information processing system which can imitate the structure and function of brain cells and brain neurons on the basis of the study of the human brain mechanism. It is made up of a large number of neurons and can be classified into different types by combinations, for example, Back Propagation (BP) neural network, Self Organizing Map (SOM) network, Generalized Regression neural network, Probabilistic neural network , etc [7].

The BP neural network is widely used in many fields because it is simple and it is able to effectively extract useful information. In particular, it is very suitable for the applications with multiple inputs and multiple outputs relations [8].

3. POWER DISTURBANCE CLASSIFICATION/IDENTIFICATION SCHEME BASED ON THE WAVELET TRANSFORM AND ARTIFICIAL NEURAL NETWORK

3.1. The Steps for the Classification/ Identification of Power Disturbance

The classification / identification of power disturbance is a procedure including pre-processing, feature extraction and pattern recognition. Pre-processing stage will de-noise and find/discard bad data, then the feature extraction stage forms the pattern features from power disturbance waveform, which are used in the pattern recognition stage to classify/identify the types of power disturbances. The procedure of power disturbance classification/identification is shown in figure 1.

Among the three stages, it is important to extract pattern features from a disturbance waveform rapidly and effectively. The sampled data from power disturbance waveform can be treated as a time sequence after pre-processing, and the WT, which is an advanced signal processing technique, can be of significant helps to feature extraction from the time sequence. The detailed scheme is introduced in the following.
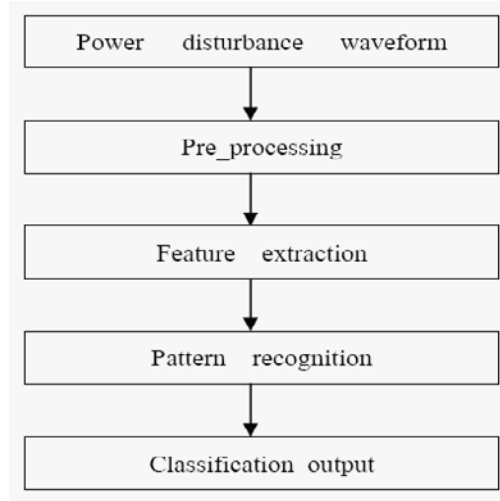
Figure 1. The procedure of power disturbance classification/identification.

3.2. Power Disturbance Feature Extraction Based on the WT

The Energy distributions of power disturbance waveforms in the frequency domain are different, therefore, the energy distributions obtained via the WT can be used as the pattern features for the distinction among various power disturbances [8].

After the M scales decomposition of a pre-processed power disturbance waveform using wavelet filter banks , the wavelet decomposition coefficients $d_k^{(j)}$ can be obtained, where j=1,2,…M, further, the energy distribution sequence can be built up as $E_j = \sum_k \left| d_k^{(j)} \right|^2$ .This procedure can also be applied to a normal voltage or current waveform to obtain sequences $d_k^{'(j)}$ and $E_j' = \sum_k \left| d_k^{'(j)} \right|^2$. Then, the feature for pattern recognition can be described as $P = [p_1, p_2, ..., p_M]$, in which

$$p_j = \frac{E_j - E_j'}{\sum_{j=1}^{M} \left| E_j - E_j' \right|}$$

(2)

After large amount of experiments, the sampling frequency and decomposition scale M are chosen as 1200Hz and 6, respectively to balance recognition capability and computational burden. As an example, a normal voltage signal is a Sine waveform with 50.0Hz frequency and 50.0 Volt amplitude. Thus, the pattern features corresponding to different types of power

disturbances, which are listed in the following, are obtained via (2) and are shown in figure 2 (a)-(f):

1. Voltage swell: amplitude is 70.0V during a 0.1 second disturbance.
2. Voltage sag: amplitude is 28.28V during a 0.1 second disturbance.
3. Voltage interrupts for a short time: amplitude is 0.0V during a 0.1 second disturbance.
4. Transient oscillation: amplitude is 53.55V during the disturbance and the primary frequency is 500.0Hz.
5. Harmonics: the third harmonic with an amplitude of 3.0V, the fifth harmonic with an amplitude of 5.0V, and the seventh harmonic with the amplitude of 7.0V.
6. Voltage fluctuation: Amplitude of the low frequency modulation signal is 0.5V, and the modulation frequency is 8.0Hz.

3.3. Power Disturbance Classification/Identification Based on the ANN

The ANN is used as a classifier in disturbance identification. Firstly, a certain number of pattern features corresponding to various known power disturbances are used to train an ANN, which will have the capability of disturbance identification in the end of this step. Then, whenever a power disturbance occurs, its pattern feature is extracted from the sampled time sequence via the WT. In the end, the obtained pattern feature is input into the trained ANN to get the type information of the power disturbance, which is the output of the ANN.
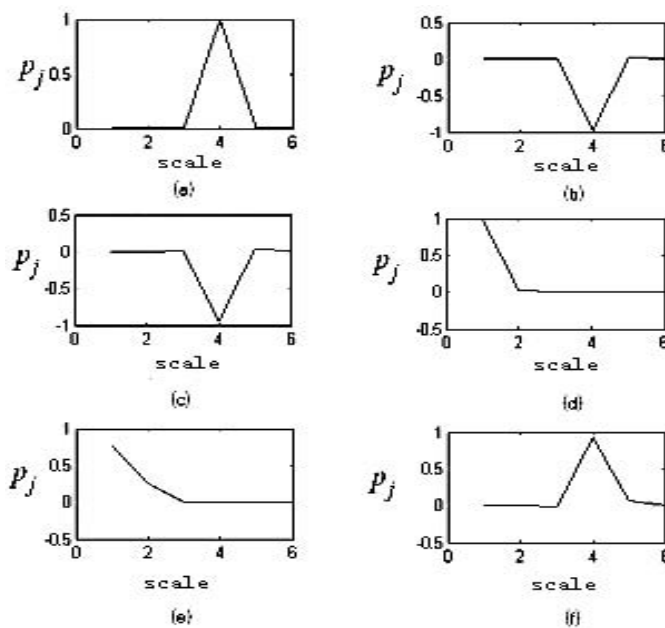

Figure 2. The Energy distributions of different power disturbances.

4. POWER DISTURBANCE PARAMETERS ESTIMATION BASED ON WAVELET TRANSFORM

The types of power disturbances studied in this paper include voltage swell, voltage sag, short time voltage interruption, transient oscillation, harmonics and voltage fluctuations. The parameters of these power disturbances and their meanings are shown in tables 1 and 2, respectively.

As the maximum module values in $d_k^{(j)}$ for all decomposition scales indicate the location of singular point in a disturbed waveform [8], the occurrence and end of a power disturbance can be easily detected through the maximum wavelet decomposition coefficients and, in consequence, the duration of the disturbance can be obtained. Further, the RMS values and amplitudes of the normal period and the duration of a disturbance can also be obtained. Because the WT can de-composite the signal components in a disturbance waveform into different frequency bands, thus the fundamental and harmonic components can be separated in the frequency domain through proper selection of the wavelet decomposition scale. In consequence, the RMS values / amplitudes of the fundamental and harmonic components can be measured in respective frequency bands. The principles of RMS value / amplitude and frequency measurement are introduced in the following sections.

Table 1. Power disturbance parameters

The types of power disturbances	Parameters
Voltage swell	T_s、 T_e、 V_{rms}^0、 V_{rms}^1
Voltage sage	T_s、 T_e、 V_{rms}^0、 V_{rms}^1
Short time voltage interruption	T_s、 T_e、 V_m^0、 V_m^1
Transient oscillation	T_s、 T_e、 V_m^0、 V_m^1、 f_1
Harmonics	T_s、 T_e、 V_{rms}^n
Voltage fluctuation	T_s、 T_e、 V_m^0、 V_m^2、 f_2

Table 2. The meanings of the parameters in table 1

Parameters	Meanings
T_s	The beginning of a power disturbance
T_e	The end of a power disturbance
V_{rms}^0	The RMS value in normal condition

Table 2. (Continued).

Parameters	Meanings
V_{rms}^1	The RMS value during a power disturbance
V_m^0	The amplitude in normal condition
V_m^1	The amplitude during a power disturbance
V_m^2	The amplitude of the low frequency AM signal during voltage fluctuation
V_{rms}^n	The RMS value of the fundamental and harmonic components during harmonics
f_1	The primary frequency during transient oscillation
f_2	The modulation frequency of the low-frequency AM signal during voltage fluctuation

4.1. RMS Value and Amplitude Measurement

The RMS value is defined in the following:

$$U = \sqrt{\frac{1}{N}\sum_{i=1}^{N} v^2(i)}$$

(3)

where N is the number of samples in a cycle of the fundamental component. If the sampling frequency is 1200.0 Hz and the fundamental frequency is 50.0Hz, then N in (3) is 24.

Provided the observation window is 0.5 second, there are 601 samples in total, which forms a sample sequence $V = \{v(i)|1 \le i \le 601\}$. The first 24 samples in the sequence, i.e. $V_{p1} = \{v(i)|1 \le i \le 24\}$, are used to estimate the RMS value at the beginning via (3); similarly, $V_{p2} = \{v(i)|2 \le i \le 25\}$, which is obtained from the sample sequence using a moving window, is used to estimate the instantaneous RMS value at the next sampling time. This procedure repeats until V_{p578} is used to estimate the last RMS value in the original sample sequence. This method can also be applied to the fundamental component and harmonics instead of the sampled disturbance waveform, which are obtained after the WT based de-composition and re-construction procedure.

For a single frequency signal component, its amplitude U_{max} and RMS value U_{rms} satisfies the following relation: $U_{max} = \sqrt{2}U_{rms}$. Thus, the amplitudes of the fundamental component and harmonics can be obtained easily from their respective RMS values.

4.2. Frequency Measurement

The zero crossing detection method is used to estimate the instantaneous frequencies of the fundamental component and harmonics. In the Sine waveform shown in figure 3, the zero crossing points are 1 and 2, their X-coordinates are "a" and "b", respectively. In consequence, the cycle T and frequency f of the single frequency signal component are as follows:

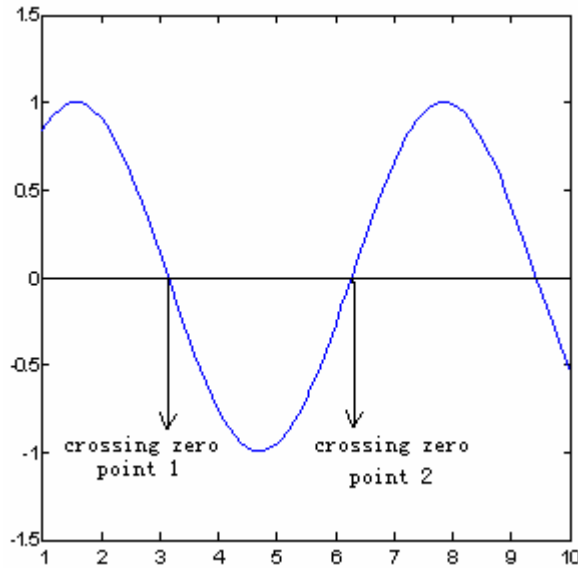$$T = 2 \times (b - a), \quad f = \frac{1}{T} \tag{4}$$

Figure 3. The zero crossing method.

In practical applications, the samples may not contain a zero point due to the sampling interval. In the following situations, namely $v(i) > 0$ and $v(i+1) < 0$, or $v(i) < 0$ and $v(i+1) > 0$, the location of the zero crossing point P can be determined:

$$p = (i + \frac{|v_i|}{|v_i| + |v_{i+1}|}) \times T_s \tag{5}$$

where T_s is the sampling interval.

Apparently, it is easy to form the sequence of zero crossing point $Z = \left\{ z(i) \mid 1 \le i \le M \right\}$ from the original sample sequence, where M is the number of zero crossing points in the sequence. Further, the frequency during any half cycle can be determined based on the locations of two consecutive zero crossing points in sequence Z using (4) and (5). In the end, a frequency sequence $F = \left\{ f(i) \mid 1 \le i \le M - 1 \right\}$ can be obtained.

5. THE SELECTION OF WAVELET FUNCTION FOR POWER DISTURBANCE IDENTIFICATION AND PARAMETER ESTIMATION

In the cases of both disturbance pattern feature extraction and parameter estimation, the WT is used to de-compose various signal components in the disturbance waveform into different frequency bands. Thus, the magnitude frequency properties of the wavelet filter bank should be ideal low pass and band pass. However, this requirement cannot be well satisfied in practical applications and, if the wavelet function used in the WT is not properly selected or constructed, spectrum leakage appears in the scale domain during the WT. Consequently, the energy distribution of a normal or disturbance waveform along the scale axis is different to its actual situation, which leads to the mal-extraction of disturbance pattern features and further affects the correctness of disturbance identification as well as the accuracy of disturbance parameter estimation. Therefore, the first criterion for the selection of wavelet function in this chapter can be concluded that the magnitude frequency property of the filter bank corresponding to the selected wavelet function should be as close as possible to ideal low pass and band pass to avoid or minimize spectrum leakage.

In order to accurately detect the occurrence and end of a power disturbance, the WT based on the selected wavelet function should have good singularity detection capability, which is able to capture even slight distorts or abnormal changes in the signal waveform corresponding to the occurrence and end of a power disturbance. This is the second criterion for the selection of wavelet function in this paper.

Based on the two selection criteria, a "Dmey" wavelet [9] is selected for the WT based disturbance feature extraction and parameter estimation.

6. A SOFTWARE PACKAGE DEVELOPED FOR POWER DISTURBANCE IDENTIFICATION AND PARAMETER ESTIMATION

A software package integrating the above mentioned signal processing and artificial intelligence techniques was developed for power disturbance classification/identification. The user interface of this software package is shown in figure 4.
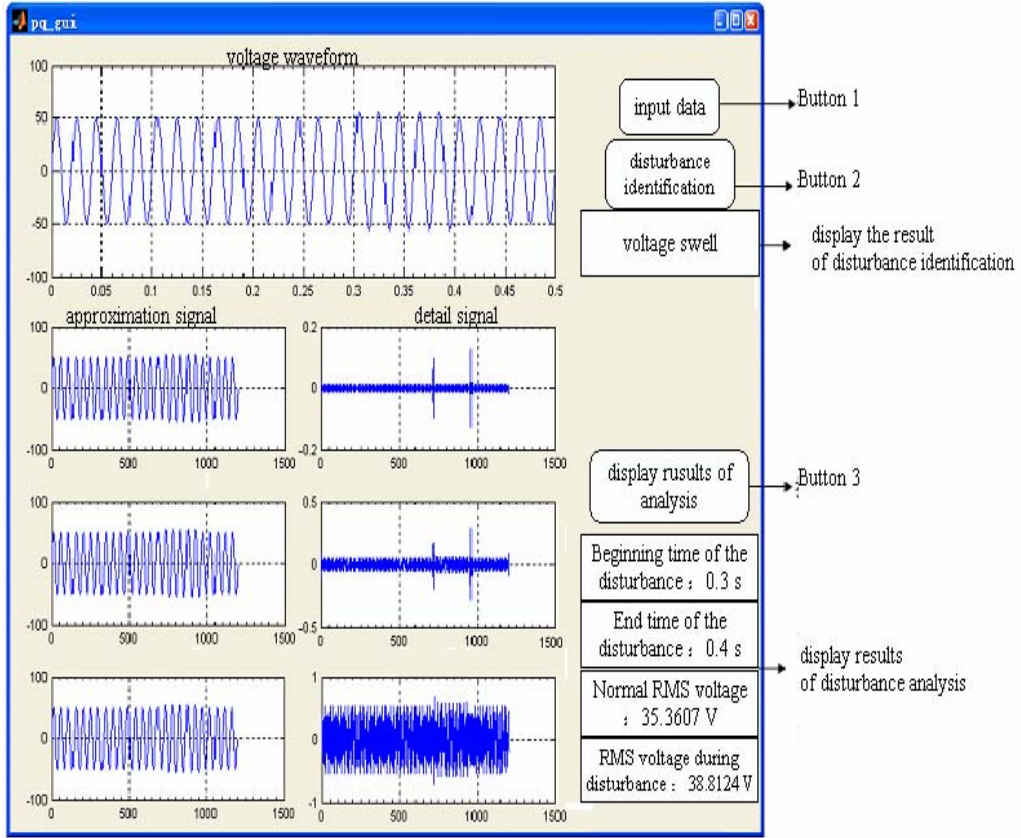
Figure 4. The user interface of the software package for power disturbance identification and parameter measurement.

In this software package, the "Dmey" wavelet based WT is used to extract disturbance pattern features. In order to decrease the computational burden, the de-composition scale used in the WT is 6 while the sampling frequency is 1200.0 Hz. A three layer BP neural network is used as classifier, which has six inputs (neurons) in the input layer and six outputs (neurons) in the output layer corresponding to the six elements in a disturbance pattern feature and six different types of power disturbances, respectively. To balance the accuracy of disturbance identification and computational burden, the number of neurons used in the hidden (middle) layer is three. The network structure is shown in figure 5, the training algorithm employed is the well known Back Propagation algorithm [10].

In detail, P is an input vector, which is the transpose of the disturbance pattern feature; $IW_{1,1}$ is the weight matrix between the hidden and input layers, and a1 is the output of the hidden layer; $LW_{2,1}$ is the weight matrix between the output and hidden layers, and a2 is the final output vector.
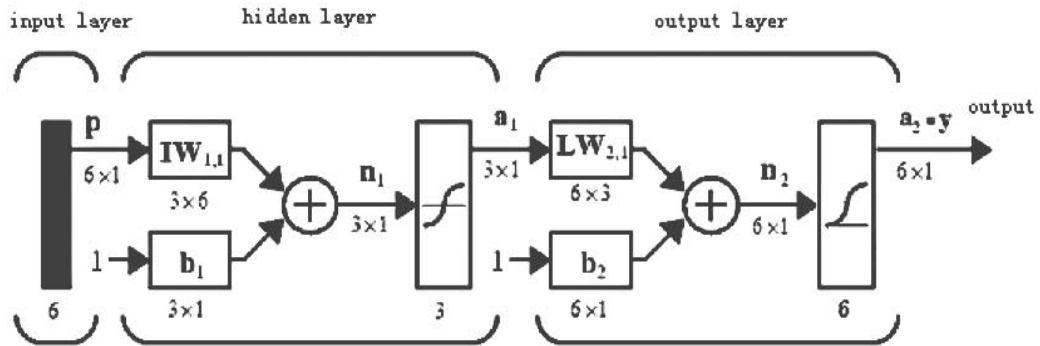
Figure 5. The structure of a 3 layer BP neural network.

Six types of power disturbances are investigated in this paper, which are voltage swell, voltage sag, short time voltage interruption, transient oscillation, harmonics and voltage fluctuation. Whenever a power disturbance pattern feature is obtained using the WT and is sent to the inputs of the BP network based classifier, a vector containing 6 elements appears at the outputs of the trained ANN classifier, in which the largest element indicates the type of the power disturbance according to its location in the vector.

7. SIMULATION VERIFICATIONS AND DISCUSSIONS

90 different power disturbance waveforms were produced using Matlab to exam the performance of the proposed method as well as the developed software package. Among these known disturbance waveforms, each type of the power disturbances discussed in this chapter, which are voltage swell, voltage sag, short time voltage interruption, transient oscillation, harmonics and voltage fluctuation, are related to 15 disturbance waveforms. These benchmarks, in fact, their corresponding pattern features, are used to train the BP network based classifier. The same training benchmarks are again used to test the performance of the whole system and a100% correctness is obtained for the six types of power disturbances. On the other hand, another group of 120 benchmarks are also produced, in which each type of the power disturbances are related to 20 cases. Similarly, these benchmarks, which are different to the training data set, are used to exam the performance of the whole system and the results, i.e. the identification rate for each type of power disturbance, are listed in table 3.

According to table 3, 105 of the 120 testing benchmarks are identified correctly, namely the identification rate reaches 87.5%. As a comparison, the probabilistic neural network has also been used in the system in place of the BP network, the performance of the system was investigated and the identification rate reaches just 85.0%. The main defect of the BP network based classifier is the computational burden due to a slower convergence of the Back Propagation training algorithm. As a help, a modified training algorithm is proposed in [11] to improve the convergence capability of the BP network.

Table 3. The results of power disturbance identification

Disturbance type	The number of Benchmarks	Correctness	Identification rate
Voltage swell	20	20	100.0%
Voltage sag	20	20	100.0%
voltage interruption	20	16	80.0%
Transient oscillation	20	15	75.0%
Harmonics	20	20	100.0%
Voltage fluctuation	20	14	70.0%

Furthermore, as a demonstration of the developed software package, a voltage swell waveform is produced via Matlab. After inputing the sampled waveform using the button "input data", the identification result and disturbance parameters can be obtained and displayed in the dialog windows by using the buttons "disturbance identification" and "display results of analysis", respectively. In the mean time, the profile of the RMS values during the voltage swell is displayed in another pop-up window, which is shown in figure 7. By comparing the estimated disturbance parameters with the objective values, which are listed in table 4, it can be concluded that the proposed methods and developed software package achieves desirable performance in the case of power disturbance identification and parameter measurement.
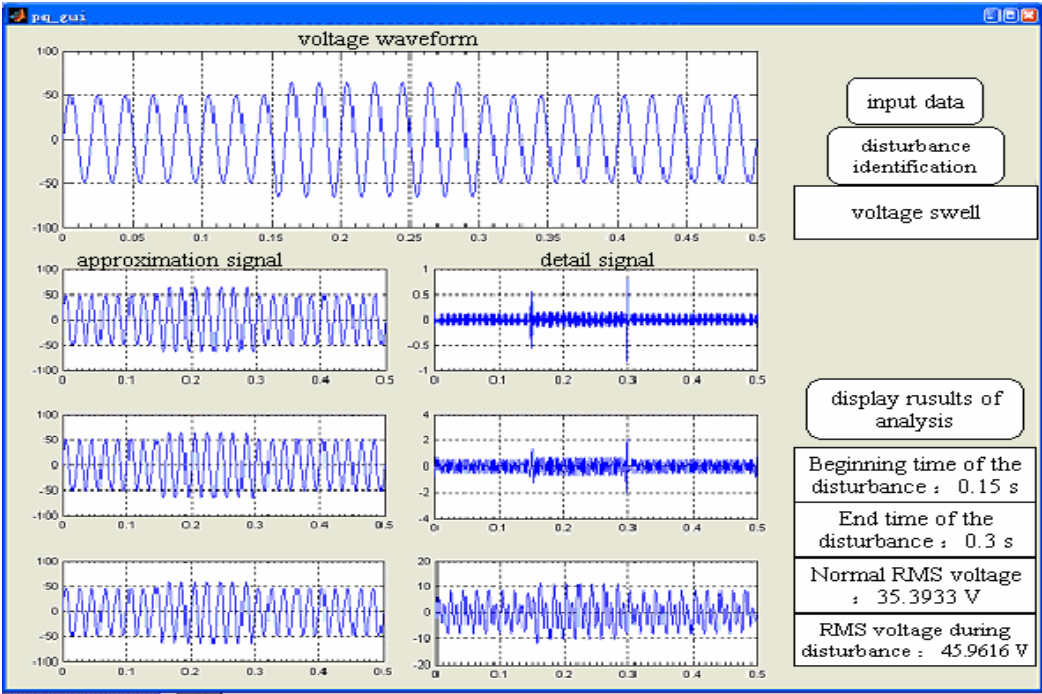

Figure 6. The results of power disturbance identification and parameter measurement in a swell case.
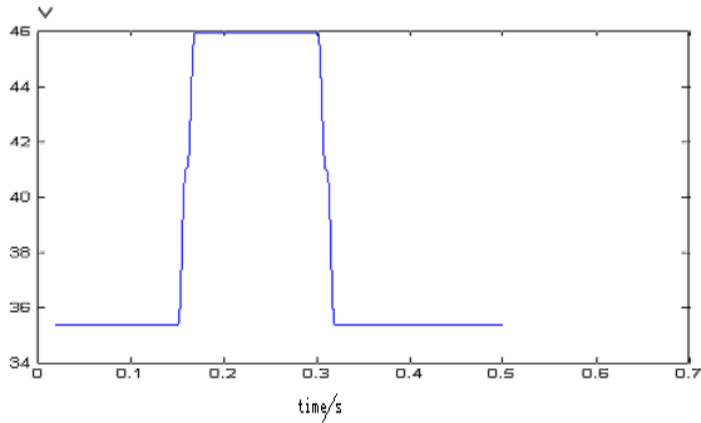
Figure 7. The profile of RMS values during a voltage swell.

Table 4. Power disturbance parameter measurement in a voltage swell case

Items	True value x	Measurement y	$error = \left\|\frac{y-x}{x}\right\| \times 100 \Big/ 100$
The RMS value of normal voltage (V)	35.36	35.39	0.08%
The RMS value of voltage swell (V)	45.97	45.96	0.02%
Time of the occurrence of swell (s)	0.15	0.15	0
Time of the end of swell (s)	0.3	0.3	0

8. CONCLUSIONS

By using the "Dmey" wavelet based Wavelet Transform and Back Propagation (BP) neural network, the types of several important power disturbances were correctly identified, while the identification rate reached 87.5% after 120 case investigations. On the other hand, the parameters of power disturbances were also measured accurately. Furthermore, a software package was developed by integrating the WT and ANN techniques to realize automatic disturbance identification and parameter measurement. As the directions for further studies, the better pattern features for single disturbance identification and better classifier for multiple disturbance identification are worth deep investigating.

REFERENCES

[1]   Surya Santoso, Jeff Lamoree, W. Mack Grady, Edward J.Powers and Siddharth C.Bhatt. *A Scalable PQ Events Identification System*. IEEE Transactions on Power Delivery, April 2000 Vol.15, No.2.

[2]   Surya Santoso, Edward J.Powers, W. Mack Grady and Antony C.Parsons. *Power Quality Disturbance Waveform Recognition Using Wavelet-Based Neural Classifier-Part 1:Theoretical Foundation*. IEEE Transactions on Power Delivery, January 2000 Vol.15, No.1.

[3]   Ouyang Sen , Jianhua Wang , Yingsan Geng , Zhengxiang Song , et al. *Power Quality Classification based on Wavelet Transform and Neural Network.* Advanced Technology of Electrical Engineering and Energy, 2003 Vol.3, 32-36.

[4]   Rafael A. Flores. *State of the Art in the Classification of Power Quality Events, An Overview*. Proceeding of the 10th International Conference on Harmonics and Quality of Power, 2002 Volume 1, 17-20.

[5]   Fusheng Yang. Engineering Analysis and Application of wavelet transform. Beijing, Science Press, 1999.

[6]   Hongyuan Wang , Guodong Shi. ANN Technology and Its Application. Beijing, China Petrol-chemical Press, 2002.

[7]   Xue Mei, Weilin Wu. *Power Quality Classification based on Wavelet and Artificial Neural Network*. Journal of Zhejiang University (Engineering Science), 2004 Vol.10, 1383-1386.

[8]   Yunlin Geng, Yigang He, Qun Wang, Donggen Yang. Detection and location of power quality disturbances by wavelet transform. Relay, 2005 Vol.8.

[9]   Matlab 7.0.1 Wavelet Toolbox. MathWorks Corporation, Sept.2004.

[10]   Matlab 7.0.1 Neural Network Toolbox. MathWorks Corporation, Sept.2004.

[11]   Jianye Sun, Hui Wang. *The Improvement of Learning Algorithm of BP Neural Networks.* Systems Engineering and Electronics, 1994 Vol.6, 41-45.

BIOGRAPHIES

Tao Lin received B.S.E.E., M.E. and PhD degrees from Huazhong University of Science and Technology, China in 1991, 1994 and 1997, respectively. He was with Central China Power Group Co., China from 1997 to 2000. In 2000, he was awarded a postdoctoral research fellowship by the Japan Society for the Promotion of Science (JSPS) to conduct researches at Nagasaki University, Japan. From May 2002 to April 2003, he was with the University of Bath, United Kingdom, as a visiting research fellow. From July 2003 to June 2005, he was with the University of Florida, USA, as a senior research scholar. From Sept. 2005, he is with Wuhan University, China as a full professor. His research interests include power quality, power system relaying, metering and the application of advanced signal processing techniques and artificial intelligence to electric power systems.

Xialing Xu received B.S. degree in Electrical Engineering from Wuhan University, China in 2002. Then she has been with the Wuhan Fiberhome Company, China for three years. From 2005, she was with the school of Electrical Engineering, Wuhan University, China, as a graduate student. Now, she is pursuing Ph.D degree at the same school. Her research interests include distributed power generation, micro-grid and power quality.

Zhengwei Fan received B.S. degree in Electrical Engineering from Wuhan University, China in 2006. Now, he is with Tsuinghua University, China as a graduate student. His research interests include power quality and power system automation.

Chapter 1

CHARACTERIZATION AND PLANNING OF DISTRIBUTED MULTI-GENERATION PLANTS

Gianfranco Chicco[*] *and Pierluigi Mancarella*[*]

Dipartimento di Ingegneria Elettrica, Politecnico di Torino
Corso Duca degli Abruzzi 24, 10129, Torino, Italy

ABSTRACT

One of the major drivers of the electrical energy system evolution is the widespread adoption of emerging technologies for distributed generation, that are shifting the focus from the centralized production to the local production of electricity. Under the distributed generation paradigm, the present research scenario is more and more emphasising the role of solutions aimed at improving the energy generation efficiency and thus the sustainability of the overall energy sector. From this point of view, the development of multi-generation solutions for small-scale applications (below 1 MW), for instance producing at the same time electricity, heat and cooling power, represents a key asset for improving the performance of the future power system. In fact, the combined production of manifold energy vectors can bring several benefits in terms of energy saving and CO_2 emission reduction, as well as potential enhanced profitability of the plants exploiting the energy mix within the liberalised electricity market framework.

This chapter illustrates a comprehensive approach to multi-generation system characterization and planning. This approach is formulated in terms of the so-called *lambda analysis*, consisting of a unified framework to study multi-generation systems, that extends the classical models based on the analysis of the heat-to-power cogeneration ratio in cogeneration plants. In particular, the representation of the energy interactions within the multi-generation plant is summarized in terms of the transformation of a vector of original energy or cogeneration ratio values into an equivalent set of values, mathematically expressed by means of specifically defined *lambda transforms*. The conceptual scheme presented provides effective characterization and modelling of the production side, the demand side and their interactions in multi-generation systems.

---

[*] Tel. +39 011 090 7141, Fax +39 011 090 7199; e-mail gianfranco.chicco@polito.it
[*] pierluigi.mancarella@polito.it

The details of the approach presented are illustrated by reviewing the bulk of alternative schemes and equipment available on the market for setting up multi-generation plants. For each alternative, the suitable equipment models and the expressions of the relevant lambda transforms are presented.

Numerical applications are provided, referred to a multi-generation system for electrical, thermal, and cooling power production. The results highlight the potential of the lambda analysis framework and of the associated lambda transforms as an effective tool to assist the energy system planner.

The availability of such a synthetic and powerful tool is of utmost importance in order to effectively cope with the increasing complexity of the future electro-energetic systems, in which the efficiency enhancement will strongly depend on the integration of the equipment for local combined production of manifold energy vectors.

NOMENCLATURE

Acronym List

AGP	Additional Generation Plant
CCHP	Combined Cooling Heat and Power
CERG	Compression Electric Refrigerator Group
CGP	Cooling Generation Plant
CHCP	Combined Heat Cooling and Power
CHG	Combustion Heat Generator
CHP	Combined Heat and Power
COP	Coefficient Of Performance
DCN	District Cooling Network
DG	Distributed Generation
DH	District Heating
DMG	Distributed Multi-Generation
EDC	Engine-Driven Chiller
EDHP	Engine-Driven Heat Pump
EDS	Electricity Distribution System
EHP	Electric Heat Pump
GAHP	Gas Absorption Heat Pump
GARG	Gas Absorption Refrigerator Group
GDS	Gas Distribution System
HHV	Higher Heating Value
HRC	Heat Recovery Condenser
HRSG	Heat Recovery Steam Generator
ICE	Internal Combustion Engine
LHV	Lower Heating Value
MT	Microturbine

SP   Separate Production
WAHP   Water Absorption Heat Pump
WARG   Water Absorption Refrigerator Group

Symbols

Positional coding is used for symbols. Subscripts with Latin letters represent energy types or end use (c=cooling, d=demand e=electricity, m=multi-generation, t=thermal, y=cogeneration, z=trigeneration) and specify the measuring units; superscripts represent equipment. For additional information, i stands for input and o for output. For Greek letters, η and ε denote efficiency or effectiveness, and λ cogeneration ratio. For energy vectors, with reference to one-hour time steps: for electricity, W is energy [kWh$_e$] or average power [kW$_e$]; for heat, Q is energy [kWh$_t$] or average power [kW$_t$]; for cooling, R is energy [kWh$_c$] or average power [kW$_c$]; for fuel, F is thermal energy [kWh$_t$] or average thermal power [kW$_t$].

1. INTRODUCTION

The diffusion of Distributed Generation (DG) resources [1-4] in recent years has opened new perspectives to the power system development. In particular, the wide range of solutions developed for the local energy production is now evolving far beyond the electricity-only generation. In fact, many DG thermal technologies, *in primis* the Internal Combustion Engine (ICE) [3-6] and the Microturbine (MT) [3-6], besides the classical larger industrial and District Heating (DH) applications [5,7,8], are now available for cogeneration applications also on a small-scale basis (below 1 MW$_e$). The benefits brought by cogeneration, owing to its good overall performance enabling energy and potential economic savings with respect to the Separate Production (SP) of heat and electricity, are well known [8-10]. Hence, nowadays CHP (Combined Heat and Power) [3-10] plants are playing a more and more important role in the worldwide energy sector.

The diffusion of cogeneration paves the way to a wider adoption of trigeneration [11-22], i.e., the combined production of electricity, heat and cooling power, in response to increasingly higher requests of air conditioning from different small-scale users (hospitals, schools, supermarkets, shopping malls, hotels, office buildings, residential blocks, universities, and so forth) [11-14], as well as large-scale users such as airports [15,16]. The rationale behind the set-up of classical CHCP (Combined Heat, Cooling and Power) or CCHP (Combined Cooling, Heat and Power) plants can be somehow seen as an extension of the concept of cogeneration; indeed, in such plants all or a part of the produced thermal energy is used to feed an absorption chiller for cooling power production, thus extending the hours of plant operation and improving the plant performance and economy [12,14,17,22].

Following a novel approach to trigeneration, the authors in previous works [23-26] have analysed various aspects of the trigeneration planning and evaluation when adopting several alternatives for cooling power production, including electric and engine-driven chillers and heat pumps, as well as direct-fired absorption chillers [18,26-31]. This approach can still be generalized to potentially entail an array of heat generation machines on which manufacturers and researchers are working worldwide in order to improve the performance and reduce the

market cost. Putting together all the various alternatives for cogeneration, cooling power generation and heat generation, it is possible to set up a *multi-generation* (or *polygeneration*) *plant*. In addition, as far as thermal equipment (for heat and/or cooling generation) is concerned, several typologies of machines that can be adopted are reversible, so that they can actually operate under both heating and cooling mode, simplifying the plant schemes and improving the plant economy, owing to the possibility of saving the purchase of some equipment. As a further point, the thermal levels at which the heat is produced in heat generation equipment such as heat pumps may be different from the ones available from the cogenerator. In this way, the concept of multi-generation can be seen as referred to the production of electricity, cooling power (chilled water, for instance) and heat (e.g., hot air, hot water, steam, and so on), with cooling and heat in case available at different enthalpy (temperature/pressure) levels.

This chapter contains a comprehensive work concerning an original approach developed for multi-generation system characterization and planning, that we have called *lambda analysis*. This approach extends and generalizes some classical tools, adopted for studying cogeneration systems, to address trigeneration and multi-generation cases. In particular, the lambda analysis enables to assess the effects of heat/cooling generation alternatives on the production side sizing and management. In fact, all the energy vectors involved in the analysis are referred to equivalent energy vectors, in particular electricity and heat (the *two* "co-generated" vectors), so as to more effectively model the characteristics and the *manifold* interactions between generation and demand. More specifically, the lambda analysis is formally described through new mathematical transformations, that we have dubbed *lambda transforms*.

The models developed within the lambda analysis enable to describe the interactions among different energy flows, in particular fuel, cooling, thermal and electrical power, within a multi-generation system and with external energy networks (such as gas and electricity distribution systems, district heating and district cooling networks), as it is often possible to encounter in distributed applications. In this respect, the concept of "distributed" in this chapter refers to the presence of a generation system of small-scale capacity, that, as such, can potentially interact with different energy networks (for instance in urban areas). The focus is set here on the single energy generation point, although in general several energy systems could be scattered to interact with various energy networks, in analogy to what occurs with electricity-only DG. The fact that manifold energy vectors can be produced by the plant and can be potentially exchanged with different energy networks leads then to what can be properly dubbed Distributed Multi-Generation (DMG).

Within this generalized framework, "classical" cogeneration and trigeneration plants are specifically illustrated as special sub-cases, owing to their diffusion and conceptual importance to the development of the multi-generation concepts.

This chapter is structured as follows. Section 2 introduces the general issues, schemes and components concerning classical cogeneration systems, as well as the extension to trigeneration and more generally to multi-generation systems. Section 3 illustrates the main issues concerning the lambda analysis approach for cogeneration system planning. More specifically, the focus is set on properly characterizing and modelling the production side, the demand side and their interactions (including the possibility of exchanging energy flows with different networks) also on a time-domain and off-design basis, two aspects often neglected at a first planning stage. Section 4 generalizes the concepts of the lambda analysis to the CCHP

planning, pointing out the effects of different equipment of the Cooling Generation Plant (CGP) upon the thermal and electrical loads as seen from the cogeneration side, which is the core of the novel approach. The trigeneration case is then generalized, in turn, in order to consider the possibility of adopting thermal generation equipment, as well as of recovering heat from the cooling generation machines. This leads to set up the so-called AGP (Additional Generation Plant) for both cooling and thermal production, which constitutes, along with the CHP side, the multi-generation energy system. This general framework is formalised from the mathematical standpoint by introducing the possible sets of lambda transforms. Section 5 deals with the relevant performance indicators and models for the most used equipment in multi-generation systems. Starting from these models, Section 6 presents explicitly the expressions for the lambda transforms for different relevant plant configurations and equipment. Section 7 discusses how the analysis based on the lambda transforms can be used as a handful tool at the system planning stage. Section 8 illustrates a case study application to a trigeneration system with different equipment alternatives, aimed at exemplifying and highlighting some numerical aspects of the theoretical concepts introduced. Section 9 contains the final remarks.

2. GENERALITIES ON DISTRIBUTED MULTI-GENERATION SYSTEMS

2.1. The Multi-Generation Framework

Planning a multi-generation plant is a challenging task, due to the variety of alternatives available. The adoption of different equipment can be considered at a planning stage within a comparative analysis that takes into account the user's needs and the economic conditions. Starting from the requested load patterns and from the availability of various equipment with different energy performance, the economic profitability of a distributed multi-generation solution also depends on the site-specific characteristics in terms of energy prices and accessibility to the energy networks and markets. In particular, the possibility of feeding different heat/cooling generation equipment with energy vectors produced *in situ* or coming from external networks, as well as selling the produced energy vectors to external networks, represent options that can be effectively investigated at the planning stage and that can change the results of an energy-based assessment [24-26].

Let us first address some concepts referred to cogeneration – the simplest case of multi-generation. In general, the first step towards the planning of an energy system is the analysis of the load patterns. One of the traditional approaches is based on the analysis of the electrical *load duration curve* [4] for both electrical and thermal power [5,12]. In addition, since the duration curve analysis does not allow for considering simultaneity of the loads, the *cogeneration ratio* (heat to power ratio [8]) is used as an auxiliary indicator. Load duration curve analysis, along with cogeneration ratio analysis, are usually sufficient to deem, at a first evaluation, which characteristics a prime mover is supposed to boast in order to satisfy the plant needs. However, often these analyses are run under rated or average conditions [8,32]. In this light, this chapter introduces a more general approach that takes into account off-design conditions and partial load operation. We have dubbed this approach *cogeneration*

lambda analysis, being the Greek letter λ used in many texts (for instance, [8]) to indicate the cogeneration ratio.

Electricity and heat are the types of energy produced in a cogeneration plant, with the corresponding duration curves and indicators. However, from a general viewpoint, a CHP plant could sometimes be seen as a multi-generation plant in which the thermal energy may be produced at different enthalpy levels, in response to specific user's requests. For instance, a small-scale ICE can produce typically hot water at about 80 °C and steam at about 10 bar [3-7]. In this case, more duration curves might be adopted for characterizing both the thermal needs and the thermal generation. Indeed, sometimes the term *trigeneration* is used to point out two different physical energy carriers used for thermal power (for instance, hot water and steam) besides electricity. Thus, when also cooling power is generated (e.g., chilled water at 7 °C for air conditioning purposes), one could even use the term "*quad-generation*".

However, for the sake of simplicity the analysis carried out in this chapter refers to a unique thermal demand curve, ruling out the details of a second-stage further level of planning, involving for instance the production of steam besides "classical" hot water cogenerated by small-scale systems. This approach has the upside of dealing with thermal demand and production likewise electrical demand and production, allowing for both simplified and unified characterization and evaluation techniques. In any case, the tools presented can be easily extended to entail more general options.

When planning a multi-generation system that includes cooling power besides heat and electricity (trigeneration), the duration curve for the cooling load has to be further considered, as well as the interaction among the different types of energy. In fact, it is possible to interpret the effect of different heat/cooling generation equipment as change in the "equivalent" electrical and/or thermal load seen from the CHP unit, thus affecting the selection and regulation of the cogeneration prime mover itself. For instance, in classical trigeneration absorption chillers are fed by heat from the cogenerator in order to produce cooling power. The final outcome is an "aggregate thermal consumption", sum of the "pure" heat load and the heat needed to generate cooling [12].

This order of concepts leads to formulate the *multi-generation lambda analysis* presented here, as a generalization of the *cogeneration lambda analysis* [26] to multiple energy vectors. In particular, as a special and important sub-case, the description of this methodology can be effectively exemplified with reference to a generalized trigeneration system in which electricity, heat and cooling power can be produced according to different combined equipment and plant schemes. Thus, in the sequel the term *trigeneration lambda analysis* is adopted, although the approach can be readily extended to any other kind of energy vector produced. In mathematical terms, the proposed approach leads to the definition of the *lambda-transforms* introduced in Section 4.5.

The lambda analysis and the lambda transforms are able to synthetically describe all the different energy flow interactions within a combined multi-generation energy systems and with the external networks (such as electricity, gas, district heating and district cooling). On the other side, this approach is orientated towards the implementation of numerical codes for time-domain simulations that enable to take into account further issues such as regulation strategies and economic interactions with the energy markets and the energy networks [24-26] within a DMG framework.

2.2. Structure of a Multi-Generation Plant

Figure 1 shows the general layout of a multi-generation plant with the related energy flows (not necessarily present simultaneously). The core of the system is represented through two main blocks (CHP and AGP).

The CHP block contains a cogeneration prime mover and, usually, a Combustion Heat Generator (CHG), for thermal back-up and peak shaving. This block produces electricity W and heat Q to various possible final uses, including the production of cooling power or additional heat after feeding the AGP.

Different possible regulation strategies may influence the sizing and selection of the equipment for both cogeneration and cooling/heat generation sides. As far as the prime mover is concerned, generally speaking the main possible classes of regulation strategies are [26]:

- *base load*: the energy system covers only the constant part of the electricity (heat) load;
- *load following*: the energy system follows the evolution of the electricity (heat) load;
- *peak shaving*: the energy system covers only part of the load during electrical peak conditions;
- *on/off operation*: the prime mover is forcibly kept on or off, usually on the basis of economic considerations.

In addition, considering small-scale CHP units, the performance of ICEs and MTs often becomes poor and more uncertain below 50% of the rated electrical load so as to make it necessary (and advised by the constructors themselves) to turn the group off.

As far as the AGP block is concerned, it can be composed of different equipment for cooling and/or heat production, and can be schematically represented according to two linking modes with respect to the CHP side:

- *separate* (or *parallel*) cooling/heat generation mode: the AGP is "decoupled" from the cogeneration side, i.e., it is fed by energy vectors (typically gas) not produced by the CHP plant (figure 2);
- *bottoming* (or *series*) cooling/heat generation mode: the AGP is cascaded to the *topping* cogeneration plant, in general distinguishing the "*electrical bottoming cycle*" (where the AGP feeding energy vector is electricity) from the "*thermal bottoming cycle*" (where the AGP feeding energy vector is heat) (figure 3).

Some equipment in the AGP block can be reversible, that is, able to operate under *cooling mode* or *heating mode*, typically producing heat for air conditioning at enthalpy levels different from the cogenerator ones.

With reference to figure 1:

- The entry F represents the thermal power contained in the fuel (based on the fuel *LHV* or *HHV* [5,6]); direct fuel-firing is the typical input to the cogeneration side; for small-scale applications, typical fuels are diesel, natural gas, dual-fuel [3,5,6],

and bio-masses [33]; in addition, the AGP can be fed by fuel, typically gas (*separate generation*, figure 2).

- The electrical power *W* can be produced by the cogeneration plant as well as bought in from the Electricity Distribution System (EDS); it can be used to supply the user's needs, to feed the AGP in order to produce heat/cooling power (*electrical bottoming generation*), and finally it can be sold out to the EDS.

- The thermal power *Q* is typically produced by the CHP side to supply the user; in addition, it can also be used to feed the AGP in order to produce additional heat for the user or cooling power (*thermal bottoming cycle*, figure 3); in case, the thermal power can be exchanged with DH networks, or be recovered from the chillers to supply part of the thermal load.

- The cooling power *R* is produced within the AGP to supply the user's need; in addition, exchanges with a DCN are in theory possible; finally, part or all of the cooling power can be fed back into the CHP plant, for instance to pre-cool the intake air of a turbine or MT in order to improve the electricity generation capacity and efficiency of the overall system [34,35].
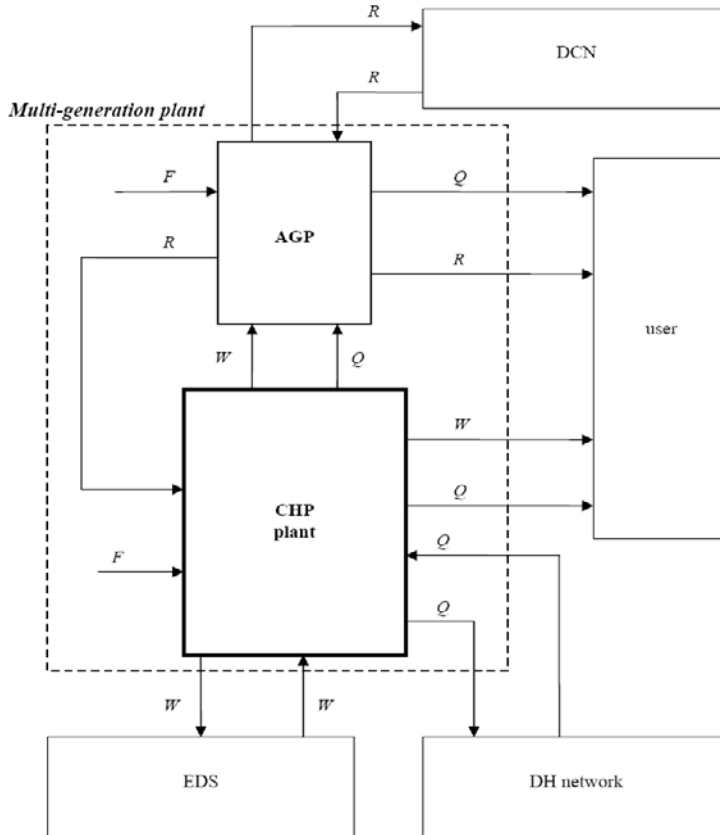
Figure 1. General layout and energy flows in a multi-generation plant.
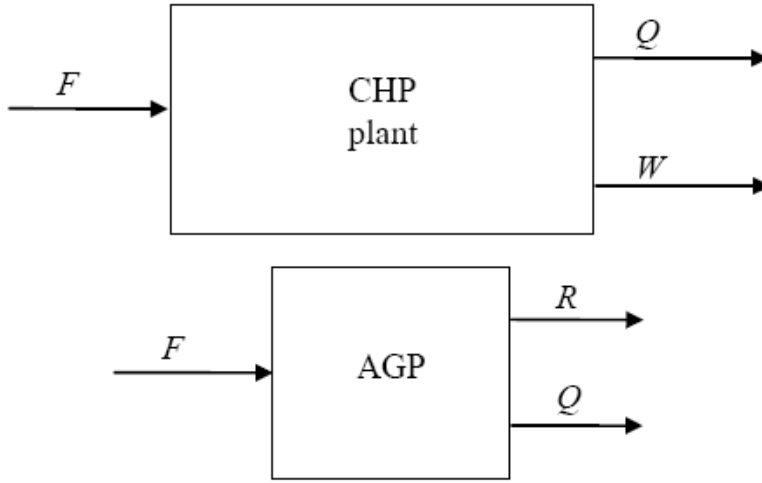
Figure 2. Multi-generation configuration with separate AGP.
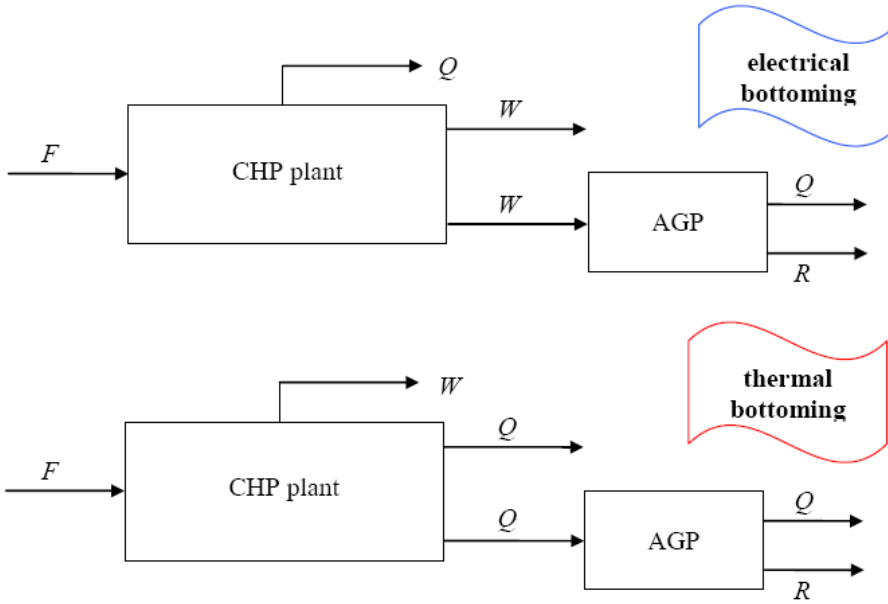
Figure 3. Multi-generation configurations with bottoming AGP.

The possibility of interacting with different energy networks is a peculiar feature of a DMG system. This aspect gives the energy system high potential in terms of energy and economic performance, and at the same time calls for adequate modelling and planning tools, as discussed in the next sections.

2.3. The CHP Block

A more detailed view of the components of the CHP block is shown in figure 4. The final use of the produced energy vectors is to supply the user's needs, or to sell the exceeding electricity to the EDS and, in case, the exceeding heat to a DH network. In particular, the CHP plant indicated as a black-box in figure 1 corresponds to the breakdown shown in figure 4.

The prime mover is the core of the plant. Its only input is the cogeneration fuel thermal power F_y, which is transformed into cogenerated electricity W_y and cogenerated heat Q_y. In turn, the cogenerated heat Q_y can be supplied at different thermal levels, depending on the prime mover and on the recovery heat generation equipment used.
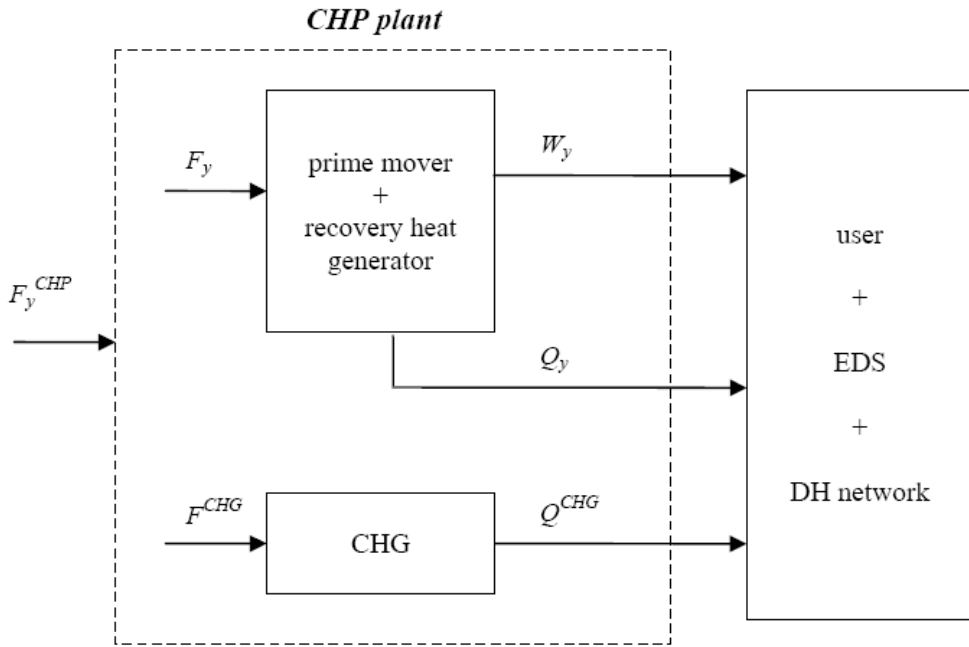
Figure 4. General CHP plant scheme.

Referring to small-scale DG, if the prime mover is a microturbine [3,5,6,36-40], usually the heat is recovered in the form of either hot water or steam. The typical fuel adopted is natural gas, while bio-masses or a mix of bio-mass and natural gas are spreading as valuable alternatives [41,42], above all in the perspective of CO_2 emission reduction [33,42]. If the prime mover is an ICE [3,5,6,36], hot or superheated water can be produced, as well as steam. Typical fuel adopted are natural gas, diesel, dual fuel, and bio-masses; however, more and more stringent emission constraints might limit the use of diesel-fueled units to back-up applications, although with a large potential of being turned into dual-fueled systems [3].

The auxiliary heat generation is usually provided by a CHG group [7,30,31], very often gas-fired, or alternatively oil-fired or diesel-fired. It is in practice a group of boilers, that can produce hot water or steam depending on the user's applications. The thermal capacity is often dwindled onto more units, for higher overall efficiency, reliability and flexibility.

As discussed in detail in the sequel, normally the user's needs do not match the production, due to the plant sizing or to specific operating conditions (for instance, the CHP group is switched off when it is not profitable to produce electricity). In this way, an auxiliary heat generation group is needed for thermal back-up, besides possible peak-shaving applications. Indeed, even if the plant is in theory sized to satisfy the electrical and/or thermal peaks, electrical and thermal production in the prime mover are correlated, whereas thermal and electrical load are relatively more independent of each other. Because of this, positive and negative thermal and electrical unbalances between production and loads are commonplace in cogeneration plants.

As far as electricity is concerned, when the prime mover production fails the load or exceeds the load, it is the EDS (under the hypothesis of grid-connected plant) to make up for the unbalanced load, by injecting/drawing energy to/from the plant. When the thermal production exceeds the load, the heat surplus is wasted to the ambient. When the heat production fails the load, instead, the unbalanced share is produced on site by means of an auxiliary heat generation group. Of course, the possibility of being connected to a DH network may change the operational options of the multi-generation plant.

2.4. The AGP Block

The components of the AGP block depend on the separate or bottoming layout.

If the AGP is *separate* from the cogeneration side, typically the following technologies of chiller/heaters directly fed by fuel can be adopted:

- GARG (Gas Absorption Refrigerator Group) [26-31,43]: the so-called direct-fired absorption chillers are fed by gas, whose thermal content is transformed into cooling effect directly in the machine. The chiller is most of times double-effect [7,27,28,30,31]. The heat, usually discarded by means of a cooling tower, could be in case recovered (for instance, by means of a heat pump to increase the heat temperature to levels suitable for user's applications [20,26]). Often, the machine is reversible and could be used also under heating mode, so as to allow for saving the purchase of additional boilers.

- GAHP (Gas Absorption Heat Pump) [7,28,29]: these machines are born for drawing thermal power from a "free" heat source (typically the atmosphere or ground water) and for supplying it to a hotter ambient after increasing its temperature. Basically, from the outside they can be seen as boilers (they are also directly-fed by fuel) but with efficiency higher than unity. However, unlike a boiler, the output heat is often limited to temperature lower than about 70-80 °C, also to avoid consistent efficiency drops [7,29-31].

- EDC (Engine-Driven Chiller) [18,26,31,44-46]: in this case, a conventional vapour-compression chiller, instead of being driven by an electrical compressor, is driven by a mechanical compressor whose shaft is directly connected to a conventional internal combustion engine. Seen from the outside as a black box, the system, also directly fed by fuel, is completely equivalent to a GARG, although energetically an EDC has the advantage that a part of the fuel input can

be more easily recovered, as in normal cogeneration ICEs. Thus, the machine can provide at the same time high-quality heat and cooling power [31].

- EDHP (Engine-Driven Heat Pump): the EDC is often a reversible machine able to work as a heat pump; in this case, the possibility of recovering thermal power also from the driving ICE allows for enhanced performance as a total energy thermal generator [26,44-46].

Considering a *bottoming* AGP, the equipment mostly used is:

- WARG (Water Absorption Refrigerator Group) [7,18,26-31]: in the classical trigeneration case, the absorption machine is fired by heat (in the form of hot water, superheated water or steam) produced in cogeneration ("thermal bottoming"). The equipment can be both single-effect or double-effect [7,28-31]. Again, the thermal power discharged to a cooling tower could be in theory recovered. As for the GARG and in analogy with the GAHP, also in this case the machine is often reversible and could be used for heating purposes as a WAHP, in case co-fired [28,29].
- CERG (Compression Electric Refrigerator Group) [7,31,32] and EHP (Electrical Heat Pump) [7,8,31,32]: the CERG is the classical solution to produce cooling power, and in a multi-generation system the feeding electricity can be produced in the CHP unit ("electrical bottoming"). In most cases, however, an EHP is adopted, since it is a reversible machine capable to work in both heating mode and cooling mode (as a CERG, then). In this way, the CHP installed electrical capacity and the working hours of the plant throughout the year are optimized, exactly as for the thermal power with a WARG. When adopting a CERG, the discharged thermal power could be sometimes recovered by means of a heat recovery condenser [24,26,30,31,47,48].

In general, besides different plant layouts, adopting different cooling (and/or heating) machines could lead to significantly different interactions among the energy flows. Modelling of these interactions, with special reference to the interface between CHP side and AGP, is one of the key drivers to the planning of the whole multi-generation system.

3. CHARACTERIZATION AND PLANNING OF A COGENERATION PLANT

Various techniques for CHP characterization and planning are well consolidated and widely used in most studies. In particular, the load duration curve analysis is one of the favourite tools for first approximation considerations. The analysis based on the CHP load duration curves and cogeneration ratio represents one of the standpoints to start with in order to carry out the lambda analysis in multi-generation systems. Hence, the concepts illustrated in the following subsections form the basis for the lambda analysis and the generalization to the lambda transforms defined in a multi-vector space.

3.1. Load Duration Curve Analysis

The *load duration curve analysis* [4,5,12] provides a first snapshot of the load levels and, accompanied by equipment efficiency curves and regulation strategies, can also give information on the overall fuel consumption. In particular, on the basis of the (thermal or electrical) duration curves, it is possible to select the prime movers and estimate the duration of their operation under a certain regulation strategy. Some examples are provided in Section 8. It has to be pointed out that this kind of analysis is approximated and represents only a first step in the planning procedure. However, it can start giving hints to the planner about sizing and regulating the prime mover, as well as on the selection of the auxiliary heat generator group, before performing further analyses, for instance based on time-domain simulations [4,26].

Conceptually, there are some limits to the load duration curve approach the planner should be aware of [4,26]:

- In a comparative analysis there is no explicit allowance for considering the efficiency (thermal or electrical one) of different prime movers; this downside might be overcome by plotting, once decided the regulation strategy, the fuel thermal consumption duration curves for the prime movers (passing through the actual efficiency point by point).
- There is no possibility to include operating schedule issues and dynamic constraints, which basically depend upon the time evolution of the load pattern, whereas they might strongly influence the choice of the prime mover and the fuel consumption to some extent.
- There is no allowance for including the "simultaneity" of thermal and electrical load in the analysis. This is the main downside of this type of analysis when applied to cogeneration plants, since, depending on the plant sizing and on the regulation strategy adopted, the simultaneity of the load might be a key issue to enable energy saving and economic profits.
- As there is no allowance for including "time-domain" considerations within the duration curves, certain "side" conditions are hard to be taken into account, e.g., the potential drop of power and efficiency depending on outdoor temperature (in particular for MTs [5,6,34,35,49-51]).

From the above considerations, it comes out that the load duration curves are an important tool for cogeneration planning analysis. However, they are not sufficient in the bulk of the practical cases, so that other auxiliary tools are needed, such as the *lambda analysis* presented here and time-domain simulations.

3.2. The Cogeneration Ratio for Generation and Load

The planning of a CHP plant has to encompass several and manifold aspects such as, *in primis*, to take into proper account the load patterns and the characteristics of different technologies of prime movers. The subsequent outcome is the need, at the planning stage, for

having at disposal simple but sound tools and indicators able to synthetically characterize the plant loads and assist the planner in properly selecting and sizing the prime mover as well as, in case, deciding the regulation strategy. The approach followed here, although essentially based on the classical approach to cogeneration plant analysis, is aimed at pointing out the characterization of a cogeneration plant on a *time-domain* basis, as opposed to the classical approach carried out on a *rated-value* or *average-value* basis. This is aimed at avoiding some common pitfalls that might mislead the CHP evaluation [26]. In particular, considering only nominal values in the analysis might lead to overlook the potentially different energy system characteristics when operating under off-design conditions. In this sense, off-design operation could occur even frequently, above all if load-following regulation strategies are set. Since the equipment off-design performance can change consistently with respect to the full-load performance [5,6,26], apparently the actual energy evaluation, as well as the economic one, could differ even dramatically from the evaluation based upon rated characteristics. In addition, specific time-domain aspects cannot be adequately accounted for through classical tools [4,26,32], and as such need to be tackled with different approaches.

In order to characterize the user's final needs, a CHP plant is often described by the well-known *cogeneration ratio* λ [8], that is, the heat-to-electricity ratio. In particular, it is possible to define a *demand-related* cogeneration ratio λ_d as the ratio of the thermal to the electrical power demand (user's side):

$$\lambda_d = \frac{Q_d}{W_d} \tag{1}$$

Similarly, it is possible to introduce the *production-related* cogeneration ratio λ_y as the ratio of the thermal to the electrical power developed by the prime mover (generation side):

$$\lambda_y = \frac{Q_y}{W_y} \tag{2}$$

As general indications on some numerical practical aspects [3-7], ICEs feature rated cogeneration ratios from 1 to about 1.5, with the lower figure for larger machines (due to the higher electrical efficiency and subsequent lower thermal efficiency). Microturbines, instead, are characterized by higher cogeneration ratios with respect to ICEs, due to lower electrical efficiency and higher thermal one; typical values are between 1.5 and 3.

In principle, a perfectly planned CHP plant should provide the "matching" $\lambda_d = \lambda_y$ ("*matched plant*" [8]), and the problem for the planner could be seen as to best match a prime mover to the load, considering the families of machines available in the capacity range to supply the requested thermal and/or electrical demand. In this ideal case, the plant satisfies the power needs without resorting to auxiliary means (namely, auxiliary boilers for the thermal production and electricity exchanges with the power grid). Unfortunately, all real cases are made up of "*unmatched*" plants.

References such as [8] provide an insightful description of the thermodynamics aspects related to the load/plant coupling by using the parameters λ_d and λ_y. However, these considerations are mostly limited to rated or average power values.

3.3. "Unmatched" Plant and Energy Interaction Modelling

Before discussing a suitable approach to time-domain characterization of the energy system, few considerations are needed on the definitions of both demand and production cogeneration ratios. Indeed, the definitions given above include the thermal and electrical user's demand, on the one hand, and the thermal and electrical prime mover production, on the other hand. This would be sufficient only in the case of *matched* plant, but when the prime mover production and the user's load are not matched, there are four basic possibilities (or combinations of them), provided that the plant is grid-connected:

1. If the cogenerated electricity is in excess with respect to the user's electrical demand, the excess electricity is sold to the grid; of course, for stand-alone plants, or when no selling to the EDS is allowed, the plant must operate in electrical load-following mode.
2. If the cogenerator fails the electricity production, the needed additional electricity is bought from the grid; again, if the plant cannot be backed up by the EDS this situation has to be avoided in order not to bring about plant shut-down.
3. If the cogenerator produces heat exceeding the user's thermal request, the exceeding heat may be in general discarded.
4. If the cogenerator fails the user's thermal demand, the plant is generally backed up by heat produced in the CHG.

Let's now focus on how to model these plant configurations in terms of cogeneration ratio or, better, in terms of *generalized production system* and *generalized load system*. For this purpose, and from an extended point of view, the prime mover plus the CHG plus the energy networks the plant could interact with (such as EDS and DH network) can be all together seen as a *CHP generalized production system* (figure 5) with respect to the prime mover-only, featuring a *generalized production cogeneration ratio* λ_y^{CHP} expressed as

$$\lambda_y^{CHP} = \frac{Q_y^{CHP}}{W_y^{CHP}} = \frac{Q_y + Q_{CHP}^{CHG} + Q_{CHP}^{DH}}{W_y + W_{CHP}^{EDS}} \tag{3}$$

where Q_y and W_y are the energy amounts produced by the prime mover, and Q_{CHP}^{CHG} and W_{CHP}^{EDS} are the energy back-ups to the CHP system from respectively the heat generator group and the grid. The back-up heat Q_{CHP}^{DH} is possibly originated from a District Heating (DH) network. Thus, when somehow the prime mover production fails the user's demand ($\lambda_y \neq \lambda_d$) the back-up systems (EDS and/or CHG) has to balance out the energy spreads, so as to obtain

$\lambda_y^{CHP} = \lambda_d$. However, this reasoning is sound for production *failings*. When the production *exceeds* the user's demand, the exceeding heat may be discharged off or sold to a District Heating (DH) network [7,8,26], exactly as the exceeding electricity is sold to the grid. Therefore, these possibilities have to be taken into account in the formulation of the planning problem considering the presence of a "*generalized load system*", corresponding, namely, to the user plus the EDS plus the DH network. Of course, rather than *loads* or *requests*, the energy shares sold to the EDS or for DH should be seen as *opportunities*.

From a modeling viewpoint, it is possible to include the heat and electricity to be sold to external networks within the user's load model (figure 6), resulting in the *generalized demand cogeneration ratio* λ_d^{CHP}, expressed as

$$\lambda_d^{CHP} = \frac{Q_d + Q^{DH}}{W_d + W^{EDS}} \qquad\qquad (4)$$

In this way, the final match between production and load should be such that $\lambda_y^{CHP} = \lambda_d^{CHP}$. In general, there are two possible approaches, namely, to include the exceeding energy amounts into the user's demand figures, as above, or to take it into account separately. Depending on the case, either approach can be profitable over the other. The latter approach, particularly effective at a first stage of analysis, is followed in the sequel. This is done essentially because of two reasons:

1. the focus is meant to be on the user's loads, as shown in detail in Section 4 on multi-generation plant characterization, where the major purpose is to analyze the impact of the cooling/heat load over the rest of the plant;
2. from the economic point of view, the energy produced to feed the user is worth differently of the one potentially sold outside: keeping the two "shares" of produced energy vectors separate helps the planner better understand the economic potentiality of the plant [24,25].

To conclude the reasoning, a last question on the energy bought from DH or EDS arises: why are heat from district heating or electricity from the grid considered *within* the generalized production model and λ_y^{CHP} when bought, whereas when they may be sold they are treated differently? This could look "asymmetric" or "incoherent", since also electricity and heat bought from outside are worth (economically and energetically) different than the ones produced inside. The difference, however, stands in the final purpose: selling energy outside is *allowed* when the production *exceeds* the user's demand; likewise, buying energy from outside is *mandatory* when the production *fails* the user's demand; therefore, from the standpoint of "*serving the user as the first goal*" the approach followed is fully consistent.
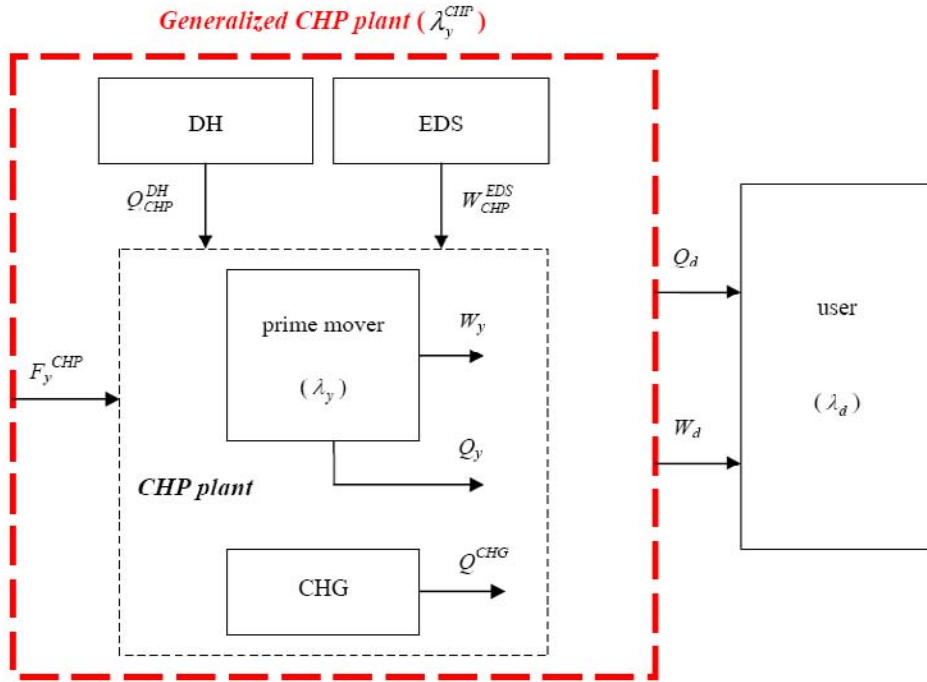
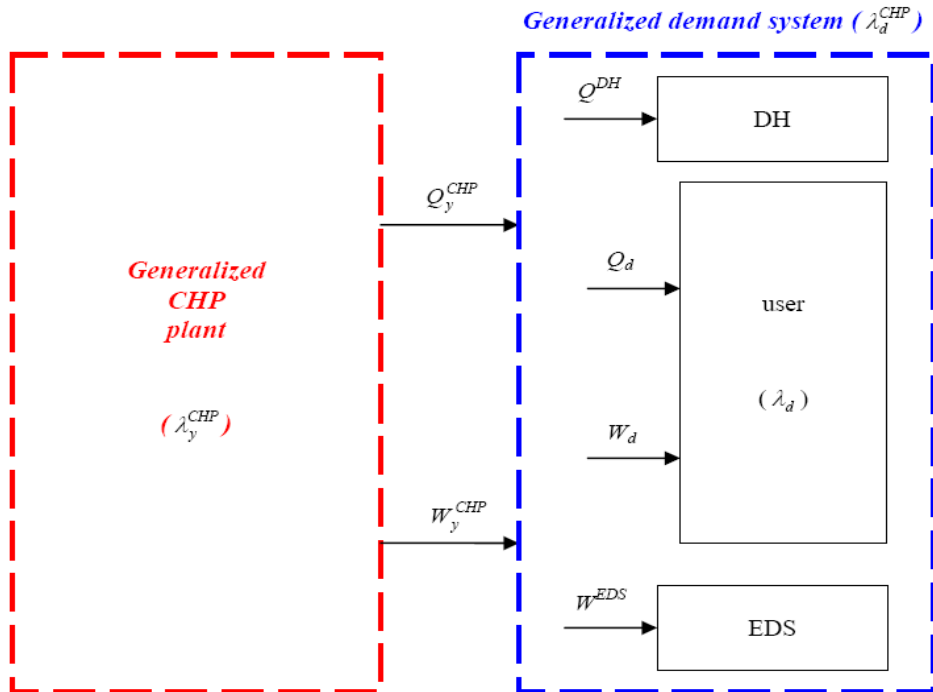Figure 5. CHP generalized production system model.

Figure 6. Generalized CHP plant and demand system model.

3.4. Time-Domain Load Characterization of a Cogeneration Plant

Let's now focus on the user's loads and the prime mover production. The definition of cogeneration ratio is completely general and can refer to any operating condition. Thus, apart from trivial cases, it is apparent that the thermal and electrical load may largely vary throughout a year (because of seasonal weather effects, for instance) or even hourly within a single day (because of the day-night cycle, at least) so as to often render it hard to simply consider rated values or to even define the rated value for λ_d, say $\hat{\lambda}_d$. Indeed, often it is not easy to figure out an "average" load, but anyway the peak load may be considered as "rated" load. On the other hand, it is tougher to choose a representative cogeneration ratio as *the* "rated" one. The best approach is therefore to plot the electrical and thermal loads on an hourly basis, together with the λ_d characteristics, so as to have a hint on what the interaction with the prime mover might be, also taking into account the time-domain cogeneration characteristics of the CHP plant, as shown below. An exemplificative numerical case study is provided in Section 8.

3.5. Time-Domain Production Characterization of a Cogeneration Plant

A cogenerative prime mover is more naturally characterized by its thermal and electrical production rated values, so that it is easier to define a rated value $\hat{\lambda}_y$. Thus, the prime mover can be selected on the basis of the electrical or thermal load to cover or follow and, at the same time, of the cogeneration ratio most suitable to follow, also depending on the regulation strategy. However, in general the CHP system is not able to follow the demand-related cogeneration ratio, since λ_d changes continuously over time. In practice, according to the classical approach to CHP planning, the equipment may be chosen in such a way that the rated cogeneration ratio is close to the demand-related cogeneration ratio values occurring in the largest part of the operating conditions. When the difference between $\hat{\lambda}_y$ and λ_d is large, it is necessary to provide energy integration by means of additional heating equipment, heat from DH networks or electricity import from the EDS.

However, in the bulk of cases this approach based *only* upon rated values of the prime mover cogeneration ratio can result poor and the actual plant performance may not correspond to the forecast one. Indeed, the reasoning developed so far is somehow incomplete, essentially due to the fact that, in load-following operation, λ_y is not constant. In fact, both thermal and electrical efficiency vary in a different fashion at partial loads, corresponding to variations of the electricity and heat cogenerated, and eventually of the cogeneration ratio. Therefore, the approach undertaken has to be improved by suitably modeling changes in λ_y during operation at partial loads, so moving on from a completely *static* point of view to a more *dynamic* one [26]. For instance, once chosen the regulation strategy, by comparing λ_y and λ_d hour by hour the planner may obtain a time-domain picture of the interaction between prime mover and load characteristics.

An effective way to describe the prime mover cogeneration characteristics is provided by exploiting its *lambda characteristics* [26], describing how λ_y changes during the actual operation and how it is suitable to someway follow the changes in λ_d.

An example of partial load performance characteristics [26] drawn from data in [52] and referring to a gas-fueled ICE is shown in figure 7. In particular, the electrical efficiency and thermal efficiency (considering separately steam and hot water production) are pictured, as well as the *EUF* (Energy Utilization Factor) [8] providing the *overall cogeneration efficiency* [5-7] (see also Section 5.2). For the same prime mover, the lambda characteristics, based on the efficiencies in figure 7, are shown in figure 8.

With reference to the figure 7 and figure 8, if only steam were profitably recovered, then the cogeneration ratio would be only about 0.5 over the modulation interval (50%-100% of the electrical output); if only the heat (from intercooler, oil coolant and jacket coolant) available in the hot water circuit were used, the cogeneration ratio would be about 0.6 at full load and would increase slightly at partial load because of the larger heat available for recovery due to the electrical efficiency reduction (prompting higher thermal discharge in the thermodynamic cycle). Finally, if heat were recovered from both sources, the cogeneration ratio would be equal to about 1.1 at full load and would increase slightly at partial load. Between the upper level (full-source recovery) and the lower level (steam-only recovery) all the states are feasible by wasting off thermal energy. The thermal output reflects this possibility of exploiting the different heat recovery techniques to modulate the prime mover thermal output. However, it would be necessary to consider the actual user's needs (steam or hot water). In addition, the possibility and the profitability of producing additional hot water instead of steam should also be evaluated given the specific case. It is therefore clear that the lambda characteristics emerge to actually be related to a region of the relevant space, more than to simple curves.

All the aspects discussed need to be analyzed already at the planning stage and, eventually, it must be taken into account that some possible regulations would be *dissipative*, and as such to avoid when possible.
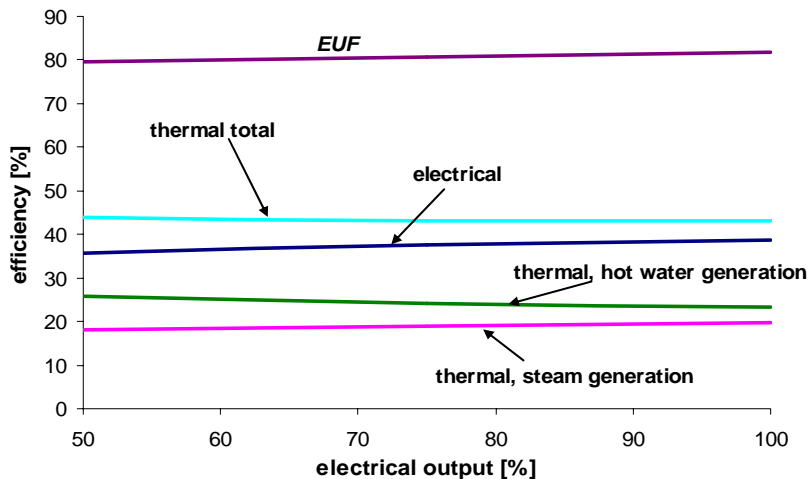
Figure 7. Partial load characteristics for a 836-kW$_e$ ICE.
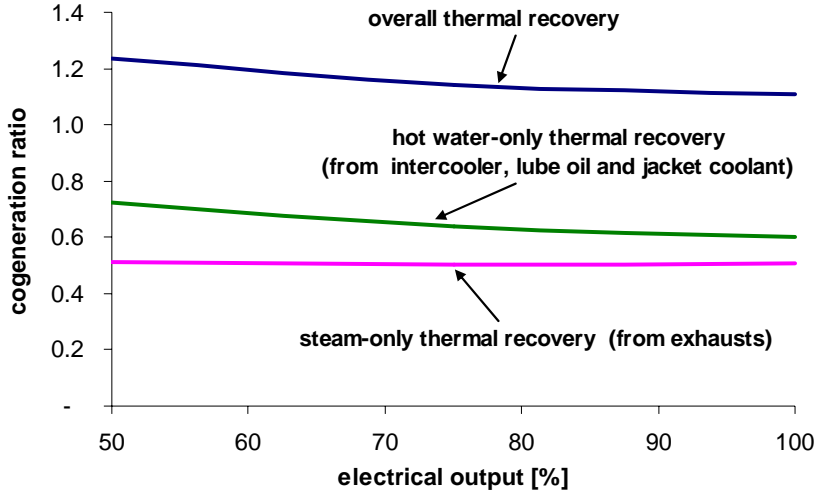
Figure 8. "*Lambda* characteristics" of a 836-kW$_e$ ICE for different levels of thermal recovery.

Besides the "intrinsic" modification of λ_y as a consequence of modifications of thermal and electrical efficiency at partial loads, it is also possible to modify *on purpose* the characteristics of the CHP system itself by means of several techniques, aimed at operating on the plant components in such a way to change either or both the thermal and electrical efficiencies [26].

A peculiar possibility to exploit in MTs would include switching the recuperator on and off [8], or adopting a *step recuperator* [53], in order to change the thermal output and the cogeneration ratio of the prime mover, a very suitable option in the presence of wide range of variation of thermal loads (especially seasonal loads) [8]. Post-combustion [5,49,54] is also possible for both microturbines and combustion engines, to raise the thermal output by exploiting additional fuel (and in case also additional air, for ICEs) injected into a HRSG. Finally, modifications of the conditions at which heat is produced may *strongly* impact on the thermal efficiency: in general, the higher the temperature of the heat to provide, the lower the amount of this heat [26,37,40], as also apparent from manufacturers' data [55].

Whatever is the technique used, it has to be highlighted that the modulation of the thermal load and the variation of the cogeneration ratio might occur at the cost of overall efficiency loss while diminishing the recuperated thermal power and/or the electrical one. Clearly, all the options presented have to be included in the formulation of the planning problem.

4. CHARACTERIZATION AND PLANNING
OF A MULTI-GENERATION PLANT

The lambda analysis approach to a multi-generation system can be effectively illustrated by means of CHCP plant analysis. In particular, the first step is to characterize the impact of the cooling power generation on the equivalent load seen from the CHP production side

(*trigeneration "lambda" analysis*), considering different technologies for the CGP, according to a broader approach to trigeneration [24-26]. Then, the analysis can be generalized to multi-generation systems (for examples of this kind of systems, see for instance [56,57]) in which an AGP is present, for the production of both heat and cooling power. This novel point of view, illustrated firstly for CHCP systems and easily extendable to multi-generation, calls for revisiting the classical approach adopted for CHP system characterization. In particular, the modelling of the impact of the AGP equipment in the interaction with the CHP side leads to the *multi-generation lambda analysis*, whose details are given in the sequel.

4.1. The Effect of Cooling Power Generation: The Trigeneration Lambda Analysis

Let's now focus on a trigeneration plant, i.e., a cogeneration plant coupled to a CGP in which only cooling power is produced, with no heat recovery. The cooling power production impacts somehow on what the cogeneration side "sees" as the load. In fact, by inspection of figure 9 it is immediate to notice that the cooling power could represent an additional load to the CHP unit, which could impact or not on either or both the thermal and electrical load, by "coupling" the cooling need to the other user's needs.

The first and straight effect on the characterization of the trigeneration CHCP system is that the load duration curves of electrical and thermal demand (as seen from the prime mover) change as well, so bringing possible modifications to the selection of the trigeneration system prime mover with respect to the cogeneration-only case. For instance, the reference [12] reports examples of planning approaches for classical trigeneration as a consequence of the so-called *aggregate thermal consumption*.

The second straight consequence on the characterization of the trigeneration plant is that the cooling plant impact on the absolute thermal and/or electrical loads to the CHP unit implies to reconsider and revise the definition of demand-related cogeneration ratio. Indeed, the definition of λ_d is aimed at the comparison with the cogeneration ratio prime mover to simplify its selection. In this case, the demand needs, as seen from the cogeneration side of the plant, should thus take into account the *overall* electricity and thermal demand. In fact, once given the user's *base* thermal load Q_d (e.g., for heating purposes) and the user's *base* electrical load W_d, the presence of different types of cooling equipment to produce the cooling power R_d can change the *overall* electricity or heat demand and subsequently the value of the demand cogeneration ratio.

From this point of view, in the CHCP plant planning the choice of the prime mover looking at the λ_d is even tougher than for CHP plants. The problem can be tackled by performing what it is possible to call *trigeneration system lambda analysis*, i.e., the analysis of the cooling power production impact over the thermal and electrical demand (as seen from the cogeneration side) and thus over the prime mover selection, sizing and regulation.

4.2. The Effect of Cooling Power Generation on the Cogeneration Ratio

The new issues brought in by the presence of the cooling side can be handled by introducing the *trigeneration demand-related cogeneration ratio*, λ_d^z in which the thermal and electrical load include the corresponding share due to cooling demand:

$$\lambda_d^z = \frac{Q_d^z}{W_d^z} = \frac{Q_d + Q_R}{W_d + W_R} \tag{5}$$

where the thermal and/or electrical loads (with respect to the cogeneration plant) Q_R and W_R, needed to produce the cooling power R_d, are added to the base Q_d and W_d loads. The terms Q_d^z and W_d^z then represent the overall thermal and electrical trigeneration loads. Likewise for CHP plants, the possibility of selling electricity to the grid or heat to a DH network might be incorporated within the definition of λ_d^z. However, this would now just complicate the concepts presented, so that the approach followed here is to consider these opportunities separately from the user's load definition.

Figure 9 and figure 10 compare the two standpoints of a *classical* demand cogeneration ratio and *trigeneration* demand-related cogeneration ratio. In the former case, the trigeneration load is seen from the "trigeneration plant" side. In the latter case, the trigeneration load is seen as a whole from the cogeneration side, and as such the demand cogeneration ratio has to include the cooling load (with respect to the CHP side) Q_R and/or W_R necessary to produce the cooling effect R_d.

As for cogeneration-only plants, λ_y^{CHP} refers to the overall CHP-side cogeneration ratio, including the heat produced by the CHG or bought from a DH network and the electricity bought from the EDS.

As a follow-up of the reasoning, the trigeneration system planning is based on the search for the best "match" between the λ_y of the prime mover and the overall trigeneration demand-related cogeneration ratio λ_d^z. When this is not possible, the prime mover cogeneration production is backed up by heat produced in boilers, electricity drawn from the EDS, and, in case, heat from a DH network (thus λ_y turns into λ_y^{CHP}), so as to reach the final match $\lambda_y^{CHP} = \lambda_d^z$.

At this stage, while formulating the planning problem, also the opportunity to sell electricity or heat to the outside should be considered *separately* (basically for economic reasons). As such, the CHP system with its λ_y^{CHP} would in practice see as possible demand interfaces the user and the electricity/district heating networks. On the contrary, the possibility of selling/purchasing cooling power to/from a District Cooling Network (DCN) should preferably be taken into account explicitly within the definition of cooling power R, as it impacts directly on the thermal and electrical load seen from the CHP system at the interface with the user.

However, in order to simplify the notation, figure 11 shows a schematic model of the trigeneration system including the input/output interface with the DCN and the output interfaces with the EDS and the DH networks; the input interface with EDS and DH is, instead, already taken into account within the definition of CHP side, as discussed above. In particular, for the sake of simplicity the subscripts *i* (input) and *o* (output) for the relevant entries are used to simply point out, respectively, the energy system incoming and outgoing flows.
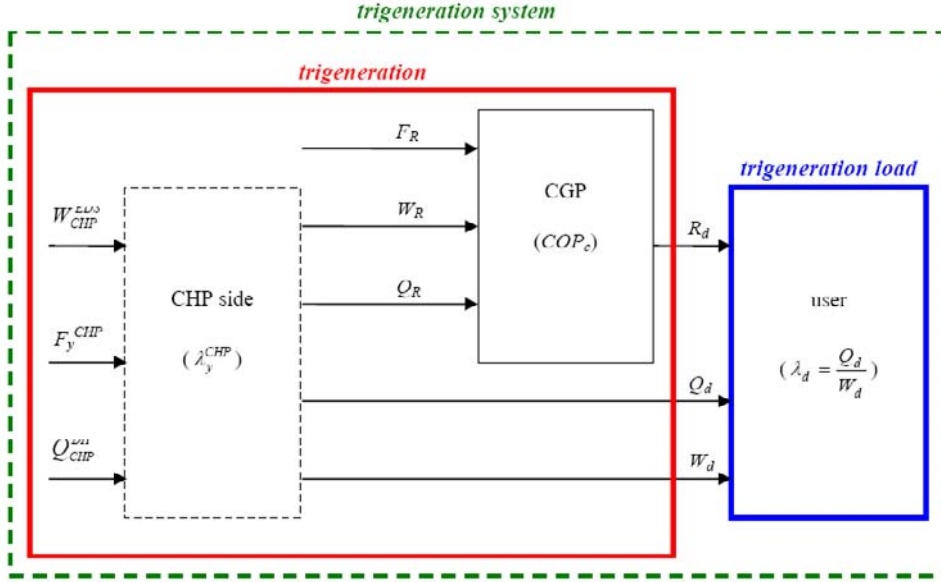
Figure 9. General trigeneration plant model with *classical* demand-related cogeneration ratio λ_d .
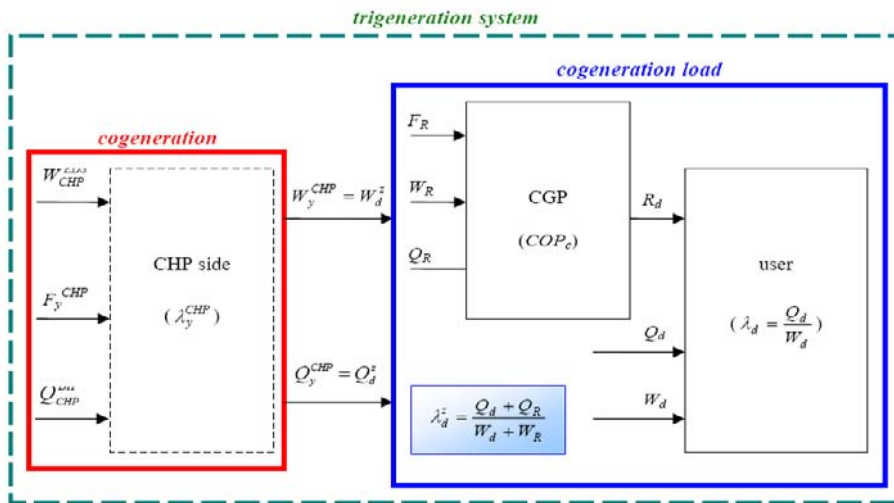
Figure 10. General trigeneration plant model with *trigeneration* demand-related cogeneration ratio λ_d^z .
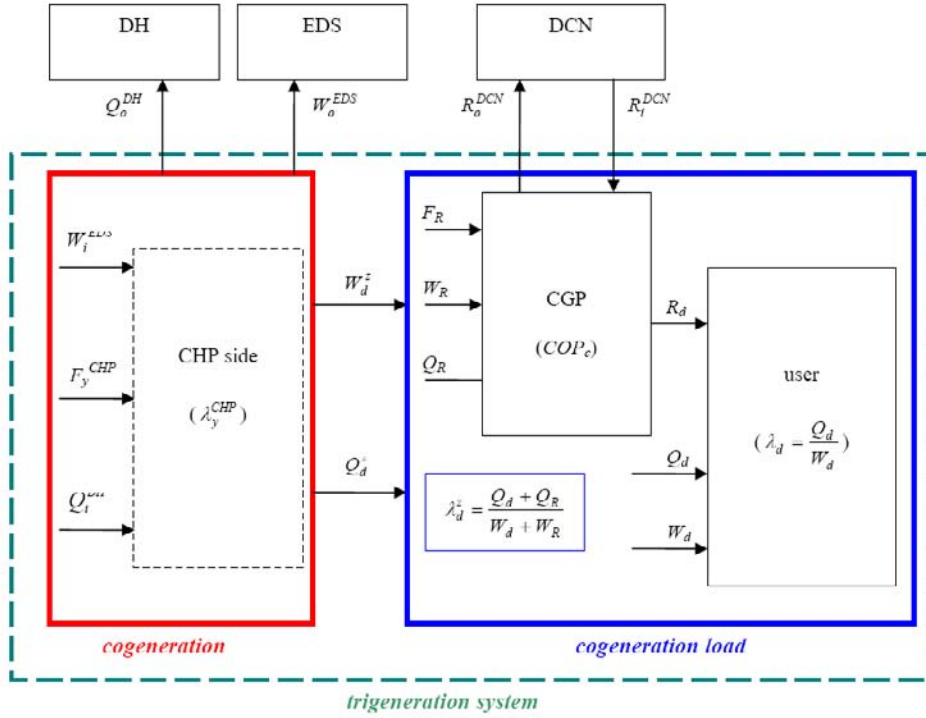
Figure 11. General trigeneration plant model with *explicit* external interfaces.

4.3. The Effect of Cooling Power Generation
on the Load Duration Curve Analysis

Planning trigeneration systems is more complicated than for cogeneration systems. The approach to follow reflects and is part itself of the trigeneration lambda analysis. Indeed, the various trigeneration alternatives, impacting differently as a thermal or electrical load, lead to change, sometimes consistently, the thermal and/or electrical load duration curves. For instance, interesting insights on the CHCP system planning are provided in [12], starting from the impact on the thermal load duration curve from adopting a WARG coupled to a CHP unit.

In order to properly evaluate the impact of the cooling load, the off-design characteristics of the CGP equipment involved are suitable to be implemented. In particular, possible seasonal effects, which can be relevant when analyzing cooling generation equipment, should be adequately accounted for by extending the time interval of observation to at least one year.

The load duration analysis carried out following these lines can help the planner get a sound view on the changes in the load absolute levels due to the different cooling equipment used. The CHP side planning is then arranged consequently, by receiving from the analysis useful hints for the prime mover sizing and suitable regulation strategy. In fact, for every different CGP considered, and thus for every different additional electrical and/or thermal load over the CHP unit, the analysis is simply developed as the one carried out for the cogeneration-only case, by comparing different prime mover technologies and sizes. Numerical examples of trigeneration load duration curve analysis are provided in Section 8.

4.4. The Effect of Heat/Cooling Power Production in the AGP: The Multi-Generation Lambda Analysis

In order to extend the *lambda analysis* developed for trigeneration systems to multi-generation systems, it is interesting to evaluate the impact of having at disposal different means to produce cooling and heating power in an AGP. Depending on the equipment used, the resulting effect from adopting specific equipment for heat generation or heat recovery can be twofold. On one side, the equivalent load seen from the cogenerator changes due to part of heat produced in the AGP. On the other side, the characteristics of the cogenerator itself can be changed, for instance due to the "heat multiplier" effect of an electric or absorption heat pump [8,26,28-31,58]. From the standpoint of the lambda analysis, this is accounted for by changing the equivalent cogeneration ratio seen by the CHP side and of the CHP side itself, as for the trigeneration case, considering also the effect of heat generation besides cooling generation. In particular, the equivalent thermal load seen by the CHP side can be reduced by adopting for instance heat recovery within the AGP, but at the same time the cogenerator could see the electrical load increased, for instance because part of the thermal power is produced by an EHP (see Section 6 for details).

The result in terms of demand-side cogeneration ratio is that the simple λ_d turns into λ_d^m, where the subscript m points out multi-generation. Similarly, the production-side cogeneration ratio λ_y^{CHP} turns into a multi-generation production-related cogeneration ratio λ_y^m, considering the effects of possible additional means to change the produced electricity and heat, such as heat pumps. Therefore, according to the *lambda analysis*, the aim of the planner is now to study the match between the generalized multi-generation production-related cogeneration ratio λ_y^m and the multi-generation demand-related cogeneration ratio λ_d^m.

The details of the lambda analysis for trigeneration and multi-generation systems, for the cases most frequently encountered (plant schemes and heat/cooling generation technologies), are exhaustively provided in Section 6.

4.5. The Lambda Transforms

From the considerations carried out so far, the multi-generation planning problem from the standpoint of lambda analysis is outlined in terms of transformation of the relevant "simple" electrical, thermal and cooling loads into "equivalent" electrical and thermal loads. The same holds true for the production side, where the cogeneration characteristics of the prime mover are changed by the CHG, external networks (DH and EDS), and equipment in the AGP (in particular, heat pumps).

In mathematical terms, the effect of the heating/cooling generation on the interaction between CHP generation side and demand side can be effectively described by means of two typologies of transforms, that we have called *lambda transforms*. The lambda transforms can be applied on the one hand to the demand side, and on the other hand to the production side. These transforms operate on a set of original energy vectors, for instance referred to a single

hour of operation, transforming them to equivalent energy vectors embedding the interactions among the various equipment in the multi-generation system. The description of the characteristics of the lambda transforms is provided here with reference to the trigeneration case. The analysis can be easily extended to other types of energy vectors (for instance a manifold level of heat production/utilization).

The first typology of lambda transforms operates on the single energy vectors (e.g., heat, electricity and cooling power), providing in general the transformation between the vector space $\Re^h$ of the original variables to the vector space $\Re^k$ of the equivalent variables, with $h \geq k$. Focusing on a trigeneration case, considering the demand side, the *lambda transform* $\Lambda_d^{WQR,WQ}$ ($\Re^3 \to \Re^2$) transforms the electrical, thermal and cooling loads into the equivalent thermal and electrical loads seen from the CHP side. From the production side, the *lambda transform* $\Lambda_y^{WQ,WQ}$ ($\Re^2 \to \Re^2$) transforms the electrical and thermal prime mover production into the equivalent thermal and electrical production after adopting additional generation means such as CHG, heat pumps, and so forth.

The second typology of lambda transforms operates on the cogeneration ratio, providing the $\Re \to \Re$ transformation from the "classical" cogeneration ratio to the equivalent cogeneration ratio. Thus, considering the demand side, the *lambda transform* $\overline{\Lambda}_d$ transforms the "simple" demand-related cogeneration ratio λ_d into the trigeneration equivalent cogeneration ratio λ_d^z seen from the CHP side. From the production side, the *lambda transform* $\overline{\Lambda}_y$ transforms the "simple" cogeneration ratio λ_y, characteristic of the prime mover generation, into the trigeneration equivalent cogeneration ratio λ_y^z.

Practical applications of the lambda transforms, referred to technologies already spread on the market, as well as to breakthrough technologies that could represent important standpoints in the future, are provided in the following sections.

5. PERFORMANCE INDICATORS FOR MULTI-GENERATION EQUIPMENT

According to the structure outlined in Section 4 for trigeneration analyses, it has been pointed out how, depending on the equipment used in the CGP, the cooling load, as seen from the CHP side, represents a further electrical or thermal load to be characterized from a time-domain point of view. In turn, the production of heating power in the AGP, as well as the heat recovery from chillers, can be seen as an equivalent change in the load characteristics or in the production characteristics for a multi-generation plant. All these issues are addressed by means of the lambda transforms, which can be expressed through the efficiency characteristics of the relevant equipment, as shown in Section 6.

In this perspective, this section describes the efficiency indicators and models referred to equipment for cogeneration, trigeneration and multi-generation systems, some of which are also detailed and adopted in the case study presented in Section 8.

5.1. Input-Output Black-Box Model Approach

The lambda analysis for multi-generation equipment is time domain-oriented, potentially accounting for the off-design equipment characteristics to be implemented in time-domain simulation codes, according to modern energy system planning techniques. From this standpoint, an effective approach to describe the characteristics and the performance of the energy system equipment is to adopt *input-output black-boxes* [26,47] describing every piece of equipment by means of the relevant efficiency indicators. In fact, no matter the physical content of the box representing the equipment, the rationale of the lambda analysis is simply to *track back*, starting from the user's needs, the input energy requested by each machine, adding it up to the overall electrical or thermal load seen from the CHP side. For instance, the cooling power, if produced by a CERG, represents an additional electrical load to the CHP side, and the amount of this load is calculated through the cooling machine *input-output performance indicator*, usually indicated as *COP* (Coefficient Of Performance) [7,30,31], as detailed in the sequel. In addition, once known the overall electricity production and the efficiency characteristics of the prime mover, it could also be possible to go further on and track back its fuel input. In this way, starting from the desired output, it is straightforward to calculate, through this simple and general approach, the input energy consumption for all the equipment and the plant (that can be in turn seen as a black-box with respect to the external networks [26]) as a whole. Moreover, it has to be underlined how this approach allows for evaluating the performance of *all* the equipment in *all* the *actual* operating points, provided that *partial-load* and in general *off-design* models of the equipment are available and implemented.

5.2. Efficiency Indicators for Black-Box Models

The efficiency indicators for energy system components are in general energetically defined as the ratio of the useful effect to the energy spent to get that effect, in practice the output-to-input energy ratio. This type of formulation makes a black-box approach of straightforward application [26,47]. The main indicators for the equipment mostly used in multi-generation systems are reported in table 1 for cogeneration systems (prime mover and CHG), and in table 2 and table 3 for AGP equipment respectively in the bottoming configuration and in the separate configuration (Section 2.2). The subscripts used highlight the "final goal" of the specific entry; in particular, *y* points out "cogeneration" production, *c* general cooling production, *t* indicates general thermal production or recovery; *Q*, *W*, and *R* refer to the final goal of producing heat, electricity, and cooling power, respectively.

Concerning the various equipment, more specifically it can be observed that:

- Cogeneration prime movers: the performance of a CHP prime mover can be described through the electrical efficiency η_W (6), electrical output to fuel thermal input ratio, and the thermal efficiency η_Q (7), thermal output to fuel thermal input ratio [3,5-8]. In addition, as an indicator of overall cogeneration performance the *EUF* (8) [3,5-7] is also often used, which gives a first idea of the actual exploitation of the fuel thermal content. The numerical values of (6) and (7)

depend upon the technology, the loading level, the outdoor conditions, the heat recovery system, and the application [3,5,6]. For instance, in figure 7 a complete partial-load model for a gas-fed CHP ICE is shown; another model is reported in table 7, referred to the engine used for the case study in Section 8.

- Combustion heat generators: the performance of the CHG units, typically industrial boilers, is normally characterized through the thermal efficiency η_t (9), thermal output to fuel thermal input ratio [7,30,31]. Also in this case, a typical efficiency model is reported in table 7.

- AGP plant equipment: the *COP* is the common way to describe the performance for the various equipment that can set up an AGP. However, the *COP* can be defined in different ways, depending on the specific machine and on the relevant input and output energy vectors. The desired output can be for instance cooling power for electric and engine driven chillers, as well as heat pumps and absorption machines operating under cooling mode, and heating power for electric heat pumps, engine-driven and absorption machines operating under heating mode; the input can be electrical power for electric chillers and electric heat pumps [7,30,31], thermal power in the form of steam or hot water for indirect-fired absorption machines [7,18,27-31], fuel for direct-fired absorption machines and engine driven equipment [7,18,27-31] (table 2 and table 3). Some typical models describing some characteristics for different equipment are reported in [7,24,26,30,31,47,48,57], as well as in table 7. In addition, heat can be sometimes recovered from chillers [26,48], which can be modeled for instance through a *thermal recovery effectiveness ε_t*, conventionally defined as ratio of the recovered thermal power to the input power (electrical power for CERG, thermal power for a WARG, fuel thermal power for a GARG and an EDC) (table 2 and table 3).

-

Table 1. Equipment and performance indicators for cogeneration plants

Equipment	Indicator	Output	Input	Equation
CHP prime mover	$\eta_W = W_y \big/ F_y$	electrical power	fuel thermal power	(6)
	$\eta_Q = Q_y \big/ F_y$	thermal power	fuel thermal power	(7)
	$EUF = \eta_W + \eta_Q$	thermal + electrical power	fuel thermal power	(8)
CHG	$\eta_t = Q \big/ F_Q$	thermal power	fuel thermal power	(9)

Table 2. Equipment and performance indicators for bottoming heat/cooling generation

Equipment	Operation mode	Indicator	Output	Input	Equation
CERG	cooling	$COP_c = R/W_R$	cooling power (heat extracted from the cold sink)	electrical power	(10)
	heat recovery	$\varepsilon_t = Q/W_R$	recovered thermal power	electrical power	(11)
EHP	heating	$COP_t = Q/W_Q$	thermal power released to the hot sink	electrical power	(12)
	cooling	$COP_c = R/W_R$	cooling power (heat extracted from the cold sink)	electrical power	(13)
WARG	cooling	$COP_c = R/Q_R$	cooling power (heat extracted from the cold sink)	thermal power	(14)
	heating	$COP_t = Q/Q_Q$	thermal power released to the hot sink	thermal power	(15)
	heat recovery	$\varepsilon_t = Q/Q_R$	recovered thermal power	thermal power	(16)
WAHP	heating	$COP_t = Q/Q_Q$	thermal power released to the hot sink	thermal power	(17)

Table 3. Equipment and performance indicators for separate heat/cooling generation

Equipment	Operation mode	Indicator	Output	Input	Equation
GARG	cooling	$COP_c = R/F_R$	cooling power (heat extracted from the cold sink)	fuel thermal power	(18)
	heating	$COP_c = Q/F_Q$	thermal power released to the hot sink	fuel thermal power	(19)
	heat recovery	$\varepsilon_t = Q/F_R$	recovered thermal power	fuel thermal power	(20)
GAHP	heating	$COP_t = Q/F_Q$	thermal power released to the hot sink	fuel thermal power	(21)
EDC	cooling	$COP_c = R/F_R$	cooling power (heat extracted from the cold sink)	fuel thermal power	(22)
	heat recovery	$\varepsilon_t = Q/F_R$	recovered thermal power	fuel thermal power	(23)
EDHP	heating	$COP_t = Q/F_Q$	thermal power released to the hot sink	fuel thermal power	(24)

6. HEAT/COOLING GENERATION IMPACT ON THE COGENERATION SIDE: EXPRESSIONS FOR THE *LAMBDA TRANSFORMS*

After introducing the performance indicators for multi-generation equipment, it is now possible to analyze in details how different cooling and heat generation equipment in an AGP can change in a different fashion the *equivalent* load seen from the CHP side. In addition, it is possible to show how the thermal power generated or recovered in the AGP can be seen as a modification of the CHP cogeneration characteristics (besides the techniques illustrated in Section 3 for the prime mover). This is the core of the lambda analysis, resulting in the expressions representing the Λ_d-transforms (see Section 4.5) reported in table 4, which exploit the performance indicators introduced above. The entries in table 4 are organized so that all the expressions relevant to the lambda analysis are clearly indicated. In particular, F_m is the overall fuel thermal input to the multi-generation system, W_d, Q_d and R_d are the user's energy demand, F_y^{CHP}, W_y^{CHP} and Q_y^{CHP} are respectively the fuel thermal input (including in case the amount for the CHG) and electricity and heat output from the CHP system (including the shares from CHG, DH and EDS). In addition, according to the rationale behind the lambda analysis, W_y^{CHP} and Q_y^{CHP} are case by case equal to the relevant expressions for the trigeneration Λ_d-transformed W_d^m and the trigeneration Λ_d-transformed Q_d^m, respectively. This section ends with showing the alternative point of view of transforming the electricity and heat generated from the CHP side into equivalent electricity and heat seen from the demand side; this is carried out by means of the Λ_y-transforms (Section 4.5) and illustrated for cogeneration prime movers coupled to heat pumps.

For the sake of completeness, the equivalent and relevant expressions for the $\overline{\Lambda}_d$-transforms introduced in Section 4.5, containing the equivalent expressions of the demand-side cogeneration ratio λ_d^m in the various cases analyzed, are summarized in table 5.

The considerations carried out in this section are supported by multi-generation plant schemes representing various equipment and interactions. Each piece of equipment is characterized by the relevant energy efficiency indicator, according to the black-box model approach presented above. In addition, although normally present and considered in the general theory, for the sake of simplicity the plant models do not envisage CHG production, nor purchase from EDS or DH networks.

Table 4. Relevant energy balances and expressions of the Λ_d-transforms for different heat/cooling generation equipment in the AGP

Equipment and operation mode		*Alternative expressions for the energy balances and Λ_d-transforms*					
		F_m		$W_y^{CHP} = W_d^m$		$Q_y^{CHP} = Q_d^m$	
		form 1	form 2	form 1	form 2	form 1	form 2
Separate cooling/heat generation							
GARG or EDC	cooling	$F_y^{CHP} + F_R$	$F_y^{CHP} + \dfrac{R_d}{COP_c}$	W_d	---	Q_d	---
	heat recovery	$F_y^{CHP} + F_R$	$F_y^{CHP} + \dfrac{R_d}{COP_c}$	W_d	---	$Q_d - Q^{AGP}$	$Q_d - \dfrac{\varepsilon_t}{COP_c} \cdot R_d$
GAHP or EDHP	heating	$F_y^{CHP} + F_Q$	---	W_d	---	$Q_d - Q^{AGP}$	---
Bottoming cooling generation							
WARG	cooling	F_y^{CHP}	---	W_d	---	$Q_d + Q_R$	$Q_d + \dfrac{R_d}{COP_c}$
	heat recovery	F_y^{CHP}	---	W_d	---	$Q_d + Q_R - Q^{AGP}$	$Q_d + \dfrac{R_d}{COP_c}(1 - \varepsilon_t)$
CERG or EHP	cooling	F_y^{CHP}	---	$W_d + W_R$	$W_d + \dfrac{R_d}{COP_c}$	Q_d	---
	heat recovery	F_y^{CHP}	---	$W_d + W_R$	$W_d + \dfrac{R_d}{COP_c}$	$Q_d - Q^{AGP}$	$Q_d - \dfrac{\varepsilon_t}{COP_c} \cdot R_d$

Table 4. (Continued).

Equipment and operation mode		Alternative expressions for the energy balances and Λ_d-transforms					
		F_m		$W_y^{CHP} = W_d^m$		$Q_y^{CHP} = Q_d^m$	
		form 1	form 2	form 1	form 2	form 1	form 2
Bottoming heat generation							
EHP	heating	F_y^{CHP}	---	$W_d + W_Q^{AG}$	---	$Q_d - Q^{AGP}$	---
WARG or WAHP	heating	F_y^{CHP}	---	W_d	---	$Q_d + Q_Q^{AGP} - Q^A$	$Q_d + (1 - COP_t) \cdot Q_Q$

Table 5. Equipment and relevant expressions for the $\overline{\Lambda}_d$ -transforms for different heat/cooling generation equipment in the AGP

Equipment	*Mode*	*Alternative expressions for λ_d^m*	
		form 1	form 2
CERG	cooling	$\dfrac{Q_d}{W_d + W_R}$	$\dfrac{Q_d}{W_d + \dfrac{R_d}{COP_c}}$
	cooling with heat recovery	$\dfrac{Q_d - Q}{W_d + W_R}$	$\dfrac{Q_d - \dfrac{\varepsilon_t}{COP_c}R_d}{W_d + \dfrac{R_d}{COP_c}}$
EHP	cooling	$\dfrac{Q_d}{W_d + W_R}$	$\dfrac{Q_d}{W_d + \dfrac{R_d}{COP_c}}$
	heating	$\dfrac{Q_d - Q}{W_d + W_Q}$	$\dfrac{Q_d - Q}{W_d + \dfrac{Q}{COP_t}}$
WARG	cooling	$\dfrac{Q_d + Q_R}{W_d}$	$\dfrac{Q_d + \dfrac{R_d}{COP_c}}{W_d}$
	heating	$\dfrac{Q_d + Q_Q - Q}{W_d}$	$\dfrac{Q_d + Q\left(\dfrac{1}{COP_t} - 1\right)}{W_d}$
	cooling with heat recovery	$\dfrac{Q_d + Q_R - Q}{W_d}$	$\dfrac{Q_d + R_d\left(\dfrac{1 - \varepsilon_t}{COP_c}\right)}{W_d}$
WAHP [a]	heating	$\dfrac{Q_Q}{W_d}$	$\dfrac{Q_d}{W_d\, COP_t}$
GARG or EDC	cooling	$\dfrac{Q_d}{W_d}$	---
	heating	$\dfrac{Q_d - Q}{W_d}$	$\dfrac{Q_d - COP_t \cdot F_Q}{W_d}$
	cooling with heat recovery	$\dfrac{Q_d - Q}{W_d}$	$\dfrac{Q_d - \dfrac{\varepsilon_t}{COP_c}R_d}{W_d}$
GAHP or EDHP	heating	$\dfrac{Q_d - Q}{W_d}$	$\dfrac{Q_d - COP_t \cdot F_Q}{W_d}$

a The expressions indicated are used if all the cogenerated heat feeds the absorption heat pump, otherwise the expression for λ_d^m is the same as for the WARG under heating mode.

6.1. Separate Cooling/Heat Generation

If a separate CGP (Section 2.2) is used to produce cooling power by means of a GARG or an EDC, the cooling power does not impact on the other loads. In fact, gas is (typically) used to supply directly the chillers, so that the input energy for cooling is completely independent of the cogeneration side operation; therefore, there is no extra electrical or thermal load. The cooling energy generation is substantially "*decoupled*" from the cogeneration.

In terms of energy balances, with reference to figure 12, the relevant expressions for a GARG or an EDC under cooling mode are reported in table 3. In particular, F_R and COP_c are respectively the fuel thermal input to the refrigerator and the cooling-mode coefficient of performance.

As a direct consequence of the separate CGP selection, the prime mover can be sized as in the classical cogeneration case (apart from possible thermal recovery from the cooling side, as shown in the sequel).
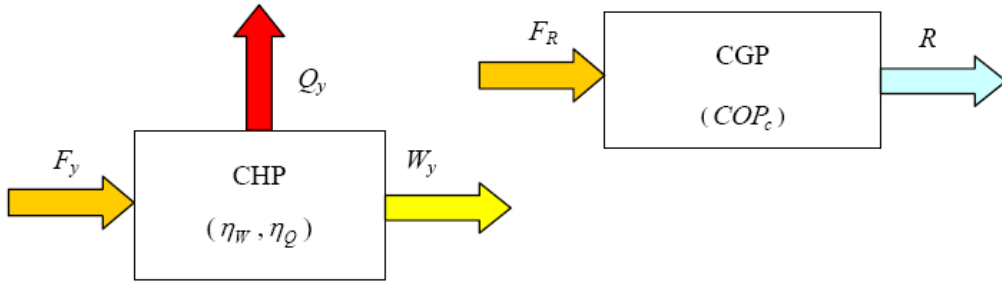
Figure 12. Black-box model for separate generation of cooling power.

Differently from the case of separate cooling production, the equipment used to produce thermal power in a separate AGP affects the production side (figure 13). In particular, if an EDHP or a GAHP are adopted, they produce an amount of thermal power equal to

$$Q^{AGP} = COP_t \cdot F_Q^{AGP} \qquad (25)$$

where Q^{AGP} and F_Q^{AGP} are respectively the heat produced in the AGP and the fuel thermal input to produce that heat. From the relevant energy balances (table 4) it can be seen that the cogenerator is "downloaded" by a certain amount of heat production to supply the thermal demand, at the cost of additional fuel input to the GAHP or to the EDHP.

$$Q_y = Q_d - Q^{AGP}$$
$$F_m = F_y + F_Q^{AGP}$$
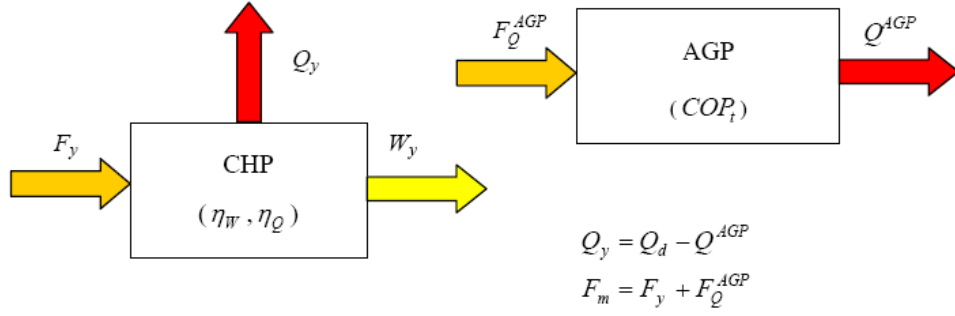
Figure 13. Black-box model for separate generation of thermal power.

6.2. Bottoming Cooling Generation

Using a bottoming CGP (Section 2.2) for cooling production, cascaded to the CHP production, the impact on the CHP plant depends on the equipment adopted.

If the CGP contains a WARG (figure 14), the input energy vector for cooling will be heat, so that the *thermal load* to the CHP unit will be increased proportionally to the cooling power required by a rate depending on the *COP* of the WARG. The cooling side is therefore *thermally-coupled* to the cogeneration side. This case represents the classical literature reference for trigeneration plants [11-14,43,59-64]. The prime mover characteristics have to take into account the increase of the thermal load due to the cooling load, according to the relations in table 4, where, in particular, the change in Q_d^m has to be pointed out, i.e., the change in the equivalent trigeneration thermal load seen by the CHP side.

$$Q_y - Q_d^m - Q_d + Q_R$$
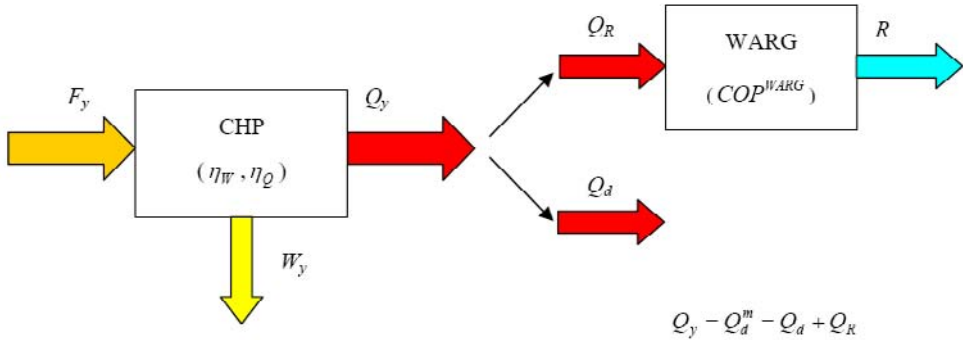
Figure 14. Black-box model for thermal bottoming generation of cooling power.

If a CERG is used as refrigerator group (figure 15), the input to the refrigerator is electricity. The cooling load therefore increases the CHP electrical load proportionally to the cooling need by a rate depending on the COP_c of the CERG. The cooling side is *electrically-coupled* to the cogeneration side. The prime mover characteristics have to take into account the rise in electrical load due to the cooling load, according to the relevant expressions in

table 4. In this case, the equivalent trigeneration electrical load W_d^m seen by the CHP side undergoes a modification with respect to the base W_d.

An EHP is usually a reversible machine. When it works under refrigeration mode, it has the same impact onto the electrical load as for the CERG, namely, the need for cooling power generation increases the electrical load by a rate depending on the refrigeration-mode COP_c.

$$W_y = W_d^m = W_d + W_R$$
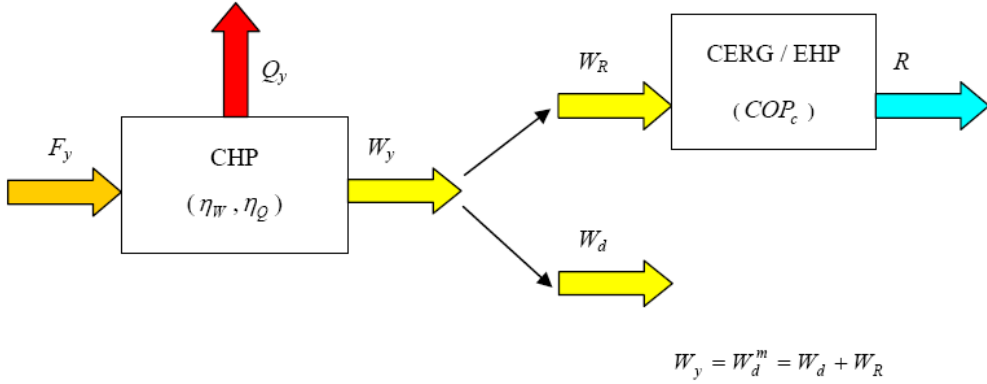
Figure 15. Black-box model for electrical bottoming generation of cooling power.

6.3. Bottoming Heat Generation

An EHP operating under heating mode (figure 16) is fed by electricity and produces heat in a proportional fashion by a rate equal to its heating-mode COP_t. In this way, the net result is to "electrically" load the prime mover to produce thermal power, so that the heat generation in the AGP is, to some extent, *electrically-coupled* to the cogeneration.

The energy balance of this process can be expressed as

$$W_Q^{EHP} = \frac{Q^{EHP}}{COP_t} \tag{26}$$

corresponding to the energy flows in the plant (excluding for simplicity the presence of simultaneous cooling load) indicated in table 4. Therefore, an EHP changes both the electrical and thermal equivalent loads seen by the cogeneration side, bringing the equivalent multi-generation electrical load W_d^m and the equivalent multi-generation thermal load Q_d^m.

The amount of heat produced by the EHP depends on the regulation strategy adopted and, as far as the machine performance is concerned, on the thermodynamic quality of the requested heat and on the cold sink temperature. In fact, the heat pump is more suitable to supply heat at temperatures below 60-70 °C in order to obtain good COP_t. Similarly, if the cold sink temperature is too low, it is not profitable to run an EHP due to its decrease in thermal capacity and efficiency [7,30,31]. On the contrary, the cogenerated heat, depending on the technology used, can be easily supplied as hot water, superheated water or steam up to

about 350 °C for recuperated-cycle MTs and about 500 °C in the case of ICEs [3,5,6,26]. In addition, the amount of thermal energy Q^{EHP} to produce in the heat pump could depend on economic factors, such as the profitability of producing "heat" from an electrical source rather than from a thermal one, depending for instance on the tariff or market conditions [65-68].

$$W_y = W_d^m = W_d + W_Q$$
$$Q_y = Q_d^m = Q_d - Q^{EHP}$$
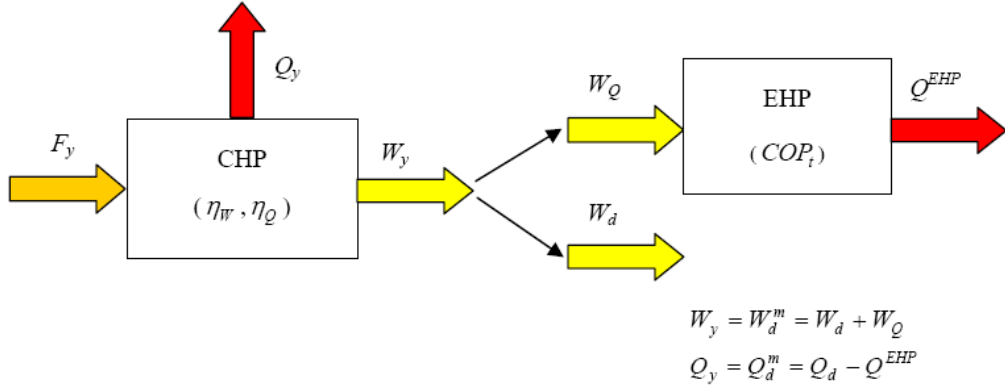
Figure 16. Black-box model for electrical bottoming generation of thermal power.

A WARG under heating mode or a WAHP (figure 17) have the effect of changing only the thermal load profile by a rate equal to the COP_t, leading to the expression

$$Q^{AGP} = COP_t \cdot Q_Q^{AGP} \tag{27}$$

and to the related expressions in table 4. In particular, Q_Q^{AGP} is the thermal input to the WAHP (or the WARG under heating mode), Q^{AGP} is the thermal output from the WAHP (or WARG under heating mode), and Q_d^m is the equivalent multi-generation thermal load seen by the CHP side. According to (27), if the COP_t of the absorption machine is greater than unity [7,31,58], the equivalent effect is to decrease the thermal load to be produced in the CHP side.

The amount of thermal heat feeding the absorption unit depends again on the regulation strategy and in case on the thermodynamic quality of the required heat, while the absorption machines are relatively less dependent on the outdoor conditions than electric heat pumps [7,29,57,69]. In particular, if all the cogenerated heat Q_y is used to feed the WAHP (or the WARG under heating mode), the expression (27) turns into

$$Q_y = Q_d^m = \frac{Q_d}{COP_t} \tag{28}$$

in which it is again possible to point out the effect of decreasing the equivalent thermal load as seen from the cogenerator if COP_t is greater than unity.
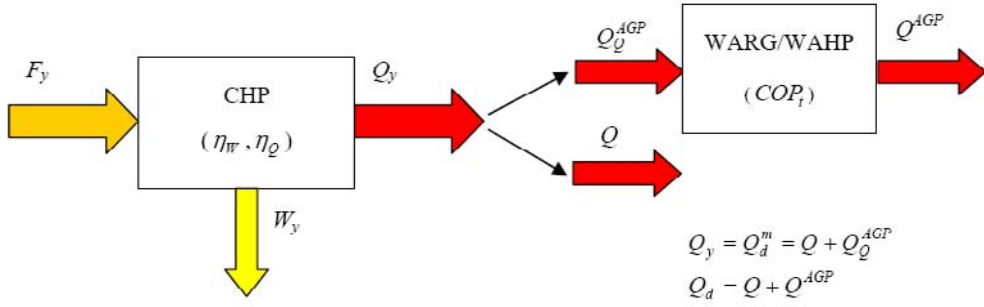
Figure 17. Black-box model for thermal bottoming generation of thermal power.

6.4. The Heat Recovery from Chillers in the AGP

Heat recovery from equipment that produces cooling power is another means to get the effect of decreasing the equivalent thermal load seen by the cogenerator. In addition, this occurs while producing cooling power, which in turns impact on the equivalent electrical or thermal load. Therefore, according to the equipment used, there might be different effects:

- if heat is recovered in a CERG by means of a Heat Recovery Condenser (HRC) [48], there is a "tri-coupling" of the loads [26], since electricity is used to produce cooling power and, through the heat recovery, also to produce thermal power; the relevant energy flows are shown in table 4, in which, in particular, Q^{AGP} is the recovered heat, COP_c is the chiller cooling-mode coefficient of performance, and ε_t is the heat recovery efficiency (11);

- if a GARG (or an EDC) is operated under heat-recovery mode, the equivalent effect is to decrease the thermal load, so that:

- $$Q_d^m = Q_d - \varepsilon_t \cdot F_R^{AGP} = Q_d - \frac{\varepsilon_t}{COP_c} \cdot R_d \qquad (29)$$

- finally, similar expressions (table 4) hold true if heat is recovered from the condenser of a WARG.

It has to be underlined that the numerical values of the cooling-mode COP_c, in case of heat recovery at the HRC, might be lower than the corresponding one without heat recovery [48]; for instance, in order to recover heat at a certain temperature useful for specific purposes, it might as well be required to increase the condensing temperature, with subsequent decrease in the COP_c. Alternatively, the heat might be more profitably brought to the desired temperature by means of an EHP [20,26,70].

6.5. An Alternative Point of View: Transformation of the Prime Mover Characteristics and Λ_y-Transforms

As generally discussed in Section 4, it might also be possible to entail the equivalent effects brought by the AGP within the production side rather than within the demand side. In terms of lambda analysis, this corresponds to exploit the Λ_y-transforms introduced in Section 4.5, from the generation side point of view rather than from the user's side. In particular, this alternative approach enables interpreting the impact of the additional heat generation as a change of the cogeneration characteristics of the prime mover (in analogy to what seen in Section 3.5) or the CHP side in general.

For instance, considering an EHP, seen from the CHP side its adoption corresponds to reduce the thermal user's request and to increase the electrical one. Now, let us consider the EHP within the CHP side, i.e., at the interface with the user: in this case, because of the same reasons as for the impact on the demand side, it is immediate to notice how the electrical heat pump *increases* the overall CHP-side cogeneration ratio λ_y^{CHP} (figure 18, in which for the sake of simplicity there is no CHG production nor electricity bought in from the EDS). In this respect, likewise a gas turbine or microturbine plant that may use fired HRSG to increase the thermal output [3,5,6], the heat pump can somehow seen as an "*electric post-combustor*".

In this case, the relevant energy flow expressions are:

$$Q_y^{CHP} = Q_y + Q^{EHP} = Q_d \tag{30}$$

$$W_y^{CHP} = W_d - W_Q^{EHP} = W_d - \frac{Q^{EHP}}{COP_t} \tag{31}$$

$$F_y^{CHP} = F_y \tag{32}$$

Similarly, the "heat multiplier" characteristics of a WAHP (as well as of a WARG in heating mode) could be seen as a means to increase the thermal production of the cogenerator, so that the WAHP could be interpreted as a "*thermal post-combustor*" (figure 19). In this case, the corresponding energy balances are:

$$Q_y^{CHP} = Q_y - Q_Q^{WAHP} + Q^{WAHP} = Q_y + Q^{WAHP}\left(1 - \frac{1}{COP_t}\right) = Q_d \tag{33}$$

$$W_y^{CHP} = W_y = W_d \tag{34}$$

$$F_y^{CHP} = F_y \tag{35}$$

Again, if all the cogenerated heat is used to feed the WAHP, the expression (33) becomes

$$Q_y^{CHP} = COP_t \cdot Q_y = Q_d \qquad\qquad (36)$$

making it apparent the "post-combustion" characteristics of the absorption heat pump if COP_t is greater than unity.

In a similar way, the heat recovered from a chiller in the AGP could be encompassed into the overall heat generated Q_y^{CHP}, so as to represent another means to change the cogeneration characteristics.

As a further comment, it has to be noticed that the heat potentially produced by the cogenerator, by heat pumps, and by chiller heat-recovery condensers, can be at different thermal levels, so as to yield actual multi-generation in the plant.

In analogy with the $\overline{\Lambda}_d$ -*transforms* reported in table 5, it is also possible to express the lambda transforms applied to the production side in terms of production side cogeneration ratio, which is carried out by means of the $\overline{\Lambda}_y$ -*transforms* (Section 4.5). For this purpose, figure 18 and figure 19 report the relevant expressions for the resulting λ_y^{CHP} in the specific case considered.
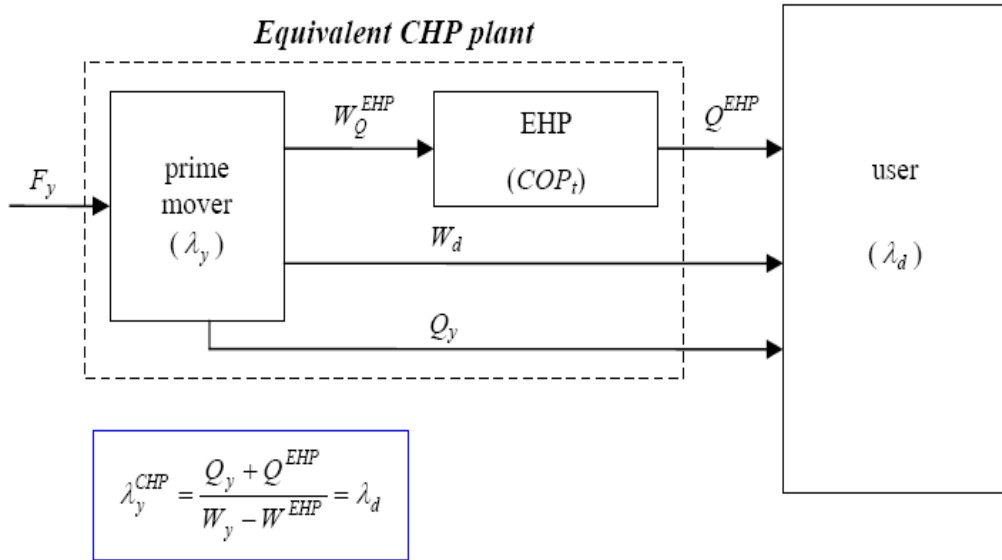
Figure 18. Multi-generation scheme with electric heat pump as an "electric post-combustor".
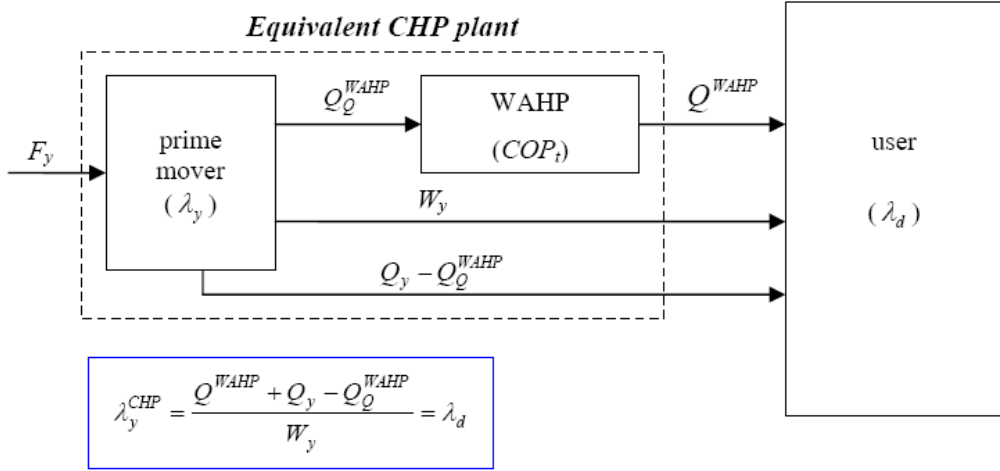
Figure 19. Multi-generation scheme with absorption heat pump as a "thermal post-combustor".

7. THE LAMBDA ANALYSIS AS A PLANNING TOOL

The effects on the multi-generation system planning brought by adopting different alternatives for the AGP can be effectively exploited at a planning stage, so that the lambda analysis can actually be seen as an additional planning tool.

7.1. The Multi-Generation Energy System Planning Process

In the light of the concepts relevant to the lambda transforms illustrated above, and developing the considerations carried out in Section 4, the steps needed while planning a multi-generation energy system, with specific reference to the production of electricity, heat and cooling power, can be summarized as follows:

- select a possible set of AGP equipment suitable to supply the cooling load and/or the thermal load, also on the basis of the units available on the market in the needed range of power and (in case) of a first rough economic analysis;
- apply the relevant lambda transforms to carry out the multi-generation lambda analysis (including the multi-generation load duration curve analysis) in order to characterize the user's needs from the CHP side perspective;
- select some of the cogenerative prime movers available on the market for application in the relevant power and cogeneration ratio ranges;
- analyze the potentiality of different prime movers to follow the load variations and the user's multi-generation lambda variation (described by the multi-generation λ_y^m after applying the relevant lambda transform) according to different regulation strategies;

- evaluate, at first approximation, the technical/economic performance of the different multi-generation systems considered as alternatives;
- run time-domain simulations to support the preliminary technical and economic analyses.

Often this process is not so clear-cut, and iterations may be needed before going on with the next step. For instance, the selection of the prime mover and/or of the regulation strategy may change after evaluating the overall performance of the system. Similarly, different options for the AGP can be considered after analyzing the prime mover alternatives. A numerical case study illustrating some of the concepts pointed out here is presented in Section 8.

7.2. AGP Selection Resorting to the Lambda Analysis

Different pieces of equipment for cooling/heat generation, with specific input/output characteristics, have a different impact on the loads seen from the CHP side. Thus, it is interesting to figure out how the production of cooling/heat power in the AGP can be interpreted as *means* to change λ_d into λ_d^m through the lambda transforms. In the same way, it is sometimes possible to modify the production λ_y^{CHP} by resorting to different prime mover configurations through electrical or thermal heat pumps operating as post-combustors, or by recovering heat from chillers in the AGP.

Therefore, to a certain extent the multi-generation planner has the *opportunity* to exploit the AGP equipment as a further variable in the selection, seeking a better match between load and production characteristics. Somehow, this opportunity is similar to the possibility of modifying the prime mover cogeneration characteristics as described in Section 3.5. Indeed, once given the manifold load (e.g., electricity, heat, and cooling), the planner carries out a first selection of prime movers, with size and characteristics suitable to supply the load, on the basis of the demand-related cogeneration ratio. Then, the AGP can be selected by taking into account the impact it may have over the cogeneration ratio in order to make λ_y (or the generalized λ_y^{CHP}) and λ_d^m better match. The more the load-production characteristics match, the better the plant is properly selected.

For instance, were the "basic" λ_y/λ_d ratio too high (without accounting for the cooling generation), the use of a WARG could "charge" the cooling load onto the thermal one, with the effect of obtaining a reduced actual λ_y/λ_d^m ratio. On the other hand, an excessively low λ_y/λ_d ratio could be tackled by adopting a CERG (whose cooling load production impacts on the electrical demand load) or a reversible EHP (with both cooling and thermal loads impacting on the electrical one, when the heat pump works respectively under chilling mode and heating mode). Similarly, if a good generation-load matching were possible without considering the cooling load, a GARG or an EDC could be a viable alternative, by decoupling the cogeneration of electricity and heat from the cooling production. In addition, the generation of heat in the AGP, directly produced (with AGP equipment working as heat

pump) or recovered from chillers, could be exploited as a further variable, thus enhancing the energy system flexibility. Finally, combinations of different cooling/heat generation equipment could be considered in some cases as viable options, bringing interesting energy saving and economic potential to the multi-generation system [20,26,58,63,64,70,71].

7.3. Suitability of Multi-Generation Solutions to Different Load Configurations

According to the considerations carried out so far, it is apparent how the possibility of exploiting different cooling and/or heat production alternatives for multi-generation in the AGP represents a high-potential opportunity. In fact, the energy recovery intrinsic to energy system cascading and equipment combination may provide better performance in terms of energy saving. At the same time, the different input/output characteristics of the equipment may provide alternative solutions to cover different plant needs in the most profitable way, also from an economic point of view [24-26].

Some general hints on possible applications of the lambda analysis, according to the considerations drawn so far, to different types of cooling/heat generators are summarized in table 6, in which equivalent multi-generation cogeneration ratios and equivalent multi-generation loads are taken into account. In particular, it is possible to synthetically notice how the main types of actions refer to situations with thermal load higher or far higher than the electrical one. This is the typical situation in the bulk of small-scale applications (residential blocks, *in primis*, along with commercial malls, offices, hospitals, schools, hotels, restaurants, and so forth [11-14,72]), as well as larger ones (for instance, airports [15,16]).

Table 6. Summary of the multi-generation lambda analysis main results

AGP equipment	Operating mode	Suitable cases	Effect on W_d^m	Effect on Q_d^m	Effect on λ_d^m
CERG	cooling	$\lambda_y < \lambda_d$	increases	---	decreases
	cooling with heat recovery	$\lambda_y \ll \lambda_d$	increases	decreases	decreases
EHP	heating	$\lambda_y \ll \lambda_d$	increases	decreases	decreases
	cooling	$\lambda_y < \lambda_d$	increases	---	decreases
WARG	cooling	$\lambda_y > \lambda_d$	---	increases	increases
	cooling with heat recovery	$\lambda_y \approx \lambda_d$	---	stationary	stationary
	heating	$\lambda_y < \lambda_d$	---	decreases	decreases
WAHP	heating	$\lambda_y < \lambda_d$	---	decreases	decreases
GARG	cooling	$\lambda_y \approx \lambda_d$	---	---	---
	cooling with heat recovery	$\lambda_y < \lambda_d$	---	decreases	decreases
GAHP	heating	$\lambda_y < \lambda_d$	---	decreases	decreases

Table 6. (Continued).

AGP equipment	Operating mode	Suitable cases	Effect on W_d^m	Effect on Q_d^m	Effect on λ_d^m
EDC	cooling	$\lambda_y \approx \lambda_d$	---	---	---
	heating	$\lambda_y \ll \lambda_d$	---	decreases	decreases
	cooling with heat recovery	$\lambda_y < \lambda_d$	---	decreases	decreases
EDHP	heating	$\lambda_y \ll \lambda_d$	---	decreases	decreases

8. CASE STUDY APPLICATION

8.1. Description of the Trigeneration User

A practical case study application referred to a CHCP system is presented. The aim is to point out the quantitative impact of different cooling generation equipment on the demand cogeneration ratio ($\overline{\Lambda}_d$-transforms) and to highlight the *net* equivalent loads seen at the interface with the CHP side (Λ_d-transforms).

A threefold load pattern (heat, electricity and cooling power) is assigned as input. In particular, the hourly thermal load is given for three reference days, related to the winter, mid season and summer seasons, respectively. The daily electrical load is assigned through the hourly pattern, that is reproduced without variations throughout the year. The thermal load pattern (hot water at 80 °C) and the electrical load pattern are shown in figure 20. In addition, figure 21 shows the hourly cooling load pattern (chilled water at 7 °C) for the same three seasons of the year as for the thermal load.

The load models presented in figure 20 and figure 21 reflect a classical approach to the planning of a combined energy system. The thermal loads (heat and cooling) are supposed to follow the night-day activity cycle as well as a seasonal behavior, essentially related to outdoor conditions and heating loads. For instance, the winter period could correspond to five months, the midseason period to four months and the summer period to three months. The electrical load, instead, is supposed to be "steady" over the year, but of course it varies throughout the day. An application with loads similar to the ones shown could be for instance a hospital or a hotel in a mild/warm climate.

From graphical inspection, it is clearly difficult to figure out an "average" load. In general, the "rated" load is considered at the maximum peak. On the other hand, it is tougher to choose a representative cogeneration ratio as the *rated* one, even though the considered load patterns are fairly simplified models. These aspects are in line with the considerations drawn in Section 3.

The duration curves for the thermal, electrical ad cooling load, together with the corresponding "basic" demand-related cogeneration ratio $\lambda_d = Q_d / W_d$ (no cooling generation impact) for the three seasons, are shown in figure 22 and figure 23, respectively.

An example of possible equipment choice to supply the user's trigenerative loads is presented in this section. The main data and the models for the equipment used are summarized in table 7.
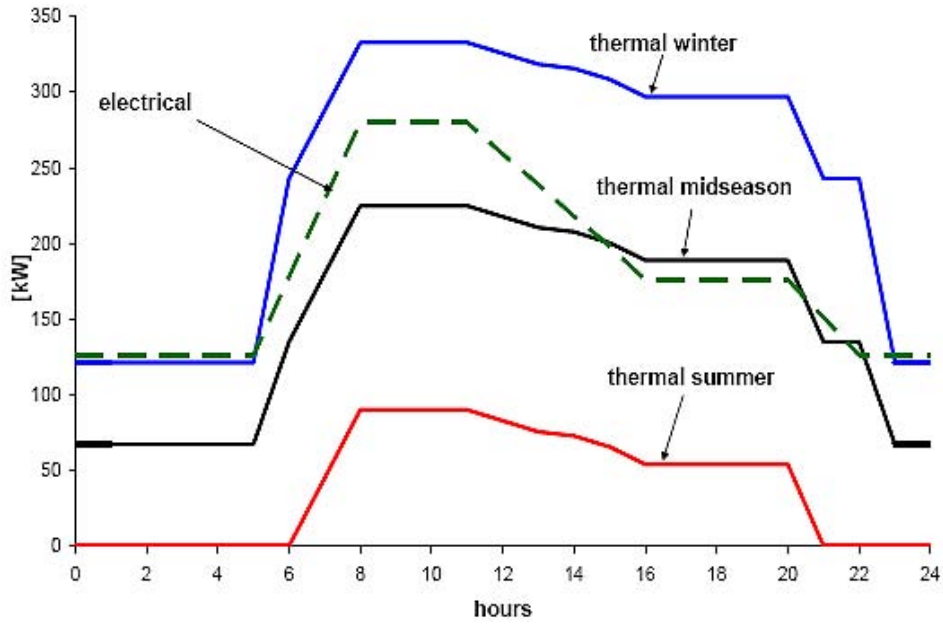
Figure 20. Hourly load patterns for the electrical power W_d and the thermal power Q_d with seasonal effects.
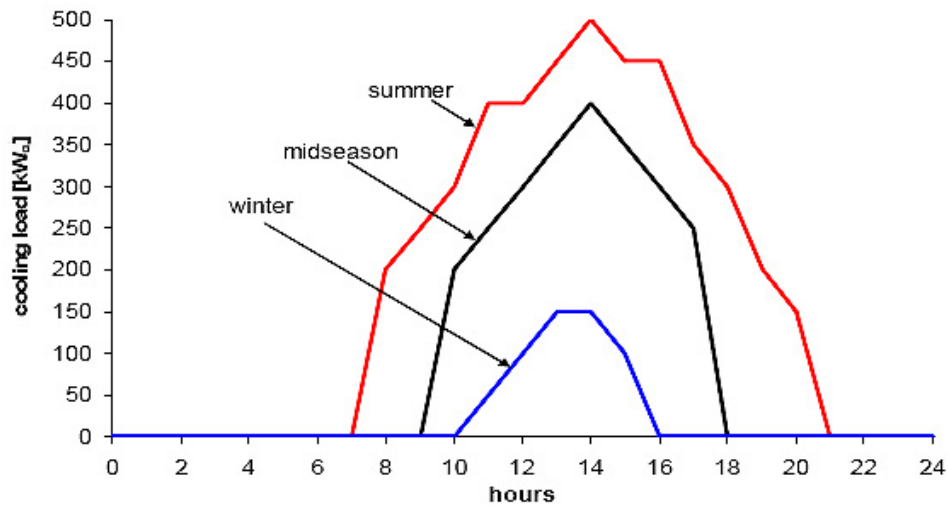
Figure 21. Hourly load patterns for the cooling power R_d with seasonal effects.
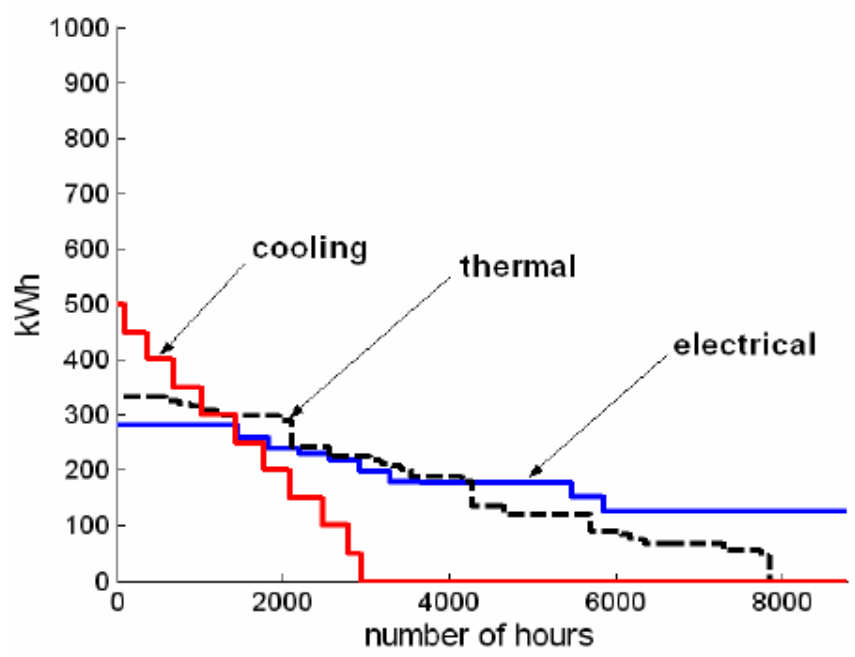
Figure 22. Threefold load duration curve for electricity, heat and cooling power.
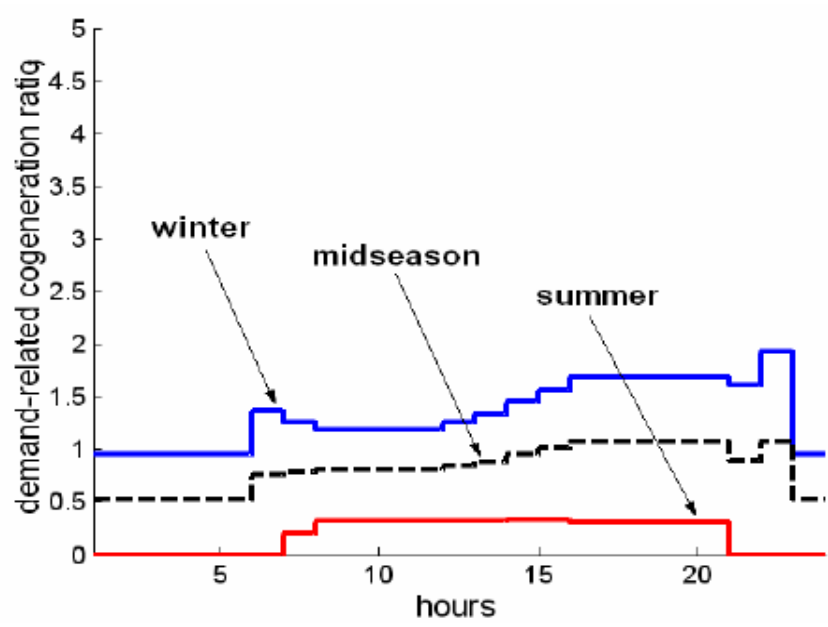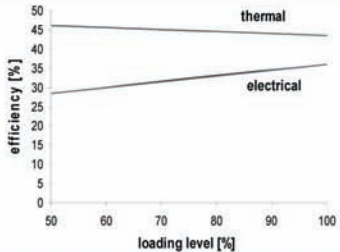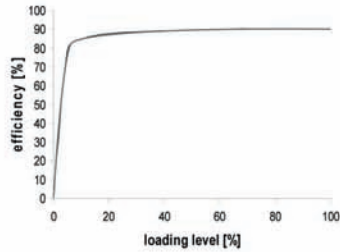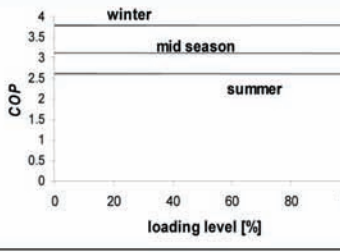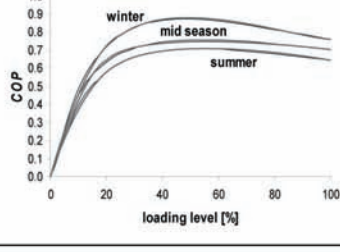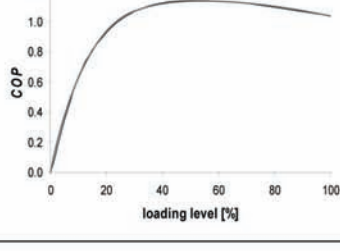
Figure 23. "Basic" hourly demand-related cogeneration ratio λ_d .

Table 7. Equipment data and models used in the case study application

Equipment	Number of units	Unit rated capacity	Unit rated efficiency	Off-design models
Prime mover (gas-fueled ICE)	1	330 kW$_e$	$w = 0.36$ $Q = 0.43$	efficiency [%] vs loading level [%]; thermal curve ~45, electrical curve rising from ~28 to ~37
CHG (gas-fueled boilers)	2 [a], 3 [c] or 4 [b]	200 kW$_t$	$t = 0.9$	efficiency [%] vs loading level [%]; rising sharply then flat at ~90
CERG	2 [a] or 1 [c]	270 kW$_c$	$COP = 3.1$	COP vs loading level [%]; winter ~3.8, mid season ~3, summer ~2.6
WARG	1	515 kW$_c$ [b] or 260 kW$_c$ [c]	$COP = 0.65$	COP vs loading level [%]; winter, mid season, summer curves rising to ~0.65–0.75
GARG	2	281 kW$_c$	$COP = 1.05$	COP vs loading level [%]; rising to ~1.1 then slightly declining

[a] Case 2; [b] Case 3; [c] Case 4 and Case 5.

8.2. The Lambda Analysis Applied to the Cooling Power Generation Equipment: Results of the Lambda Transforms

Let's now consider five cases for the CGP equipment, aimed at showing the impact of the different cooling generation characteristics on the electrical and thermal loads, that is, the results of the lambda transforms upon the duration curves and the demand-related cogeneration ratios. In particular, the impact of the cooling load is calculated by means of the expressions found in Section 6 for the relevant lambda transforms, and considering off-design *COP* characteristics for all the equipment, according to models drawn from data from equipment available in the market (table 7). The results obtained from the application of the Λ_d-transforms (table 4) and the $\overline{\Lambda}_d$-transforms (table 5) are shown in figure 24 for the five cases considered, described below.

8.2.1. Case 1: GARG

Let us consider the CGP to be composed of a GARG. In particular, the chilling group is made up of two direct-fired double-stage 281-kW$_c$ units, directly fed by gas from a GDS. The rated *COP* is about 1.05, and the off-design model takes into account partial-load operation change [26] (table 7). The two units are supposed to operate so as to equally share the load. Since a GARG "decouples" the cooling side from the cogeneration side (heat and electricity load remain unchanged by the presence of the cooling load), the duration curves and the user's lambda, shown in figure 24, are exactly the same as in figure 22 and figure 23 ($\lambda_d^z = \lambda_d$). Thus, the cogenerator sizing is independent of the cooling production. In addition, from the cooling load duration curve it is possible to estimate the gas consumption due to the GARG, as if it were a prime mover, according to what discussed in Section 3.1 and for instance in [4].

8.2.2. Case 2: CERG

In this case, the CGP plant is assumed to be composed of a CERG, made up of two 270-kW$_c$ electric chillers in parallel, fed by the CHP system and/or the EDS. The rated *COP* is about 3, and the off-design model takes into account *COP* variations with seasonal average variations of average outdoor temperature, whereas the performance is considered at first approximation constant at partial load [7,26] (table 7). Again, the two units are supposed to be regulated so as to equally share the load

The trigeneration electrical and thermal load duration curves (after applying the relevant Λ_d-transform) and the demand related cogeneration ratio λ_d^z (after applying the relevant $\overline{\Lambda}_d$-transform) for this solution are again shown in figure 24. The cooling load goes to impact on the electrical load, raising it up in the morning hours of the day, when the chillers are mostly on, thus lowering the λ_d^z. By direct comparison with *Case 1*, it is apparent that the electrical load increases in a time span equal to about one half of the yearly load duration.

8.2.3. Case 3: WARG

The classical concept of trigeneration leads to consider for the CGP the cooling production through a WARG, e.g. one 515-kW$_c$ machine, fed by hot water at 80 °C produced in cogeneration when the CHP system is switched on, otherwise by the CHG. In this case, a

single-effect unit is used, with rated *COP* equal to about 0.65. The off-design model takes into account partial-load variations and seasonal outdoor temperature-based variations of the *COP* [26] (table 7).

The lambda-transformed trigeneration load duration curves and the relevant λ_d^z are shown in figure 24. The thermal load undergoes a significant increase because of the cooling load impact, due to the relatively "low" *COP* (about 0.65 at rated conditions) and the relatively "high" cooling load. The curve profiles are completely modified, if compared to the other cooling machines considered.

8.2.4. Case 4: WARG (Base-Load) + CERG (Modulation)

In this case, the CGP is composed of two different equipment, respectively a 260-kW$_c$ unit for a WARG, and 270-kW$_c$ unit for a CERG. The off-design characteristics of the equipment are set as in *Case 2* and *Case 3*. Being double the equipment used to supply the cooling load, a better defined regulation strategy for the cooling plant needs to be set up. Thus, in this case the WARG is used to cover the base-cooling load (so as to better exploit the exceeding cogenerated heat), whereas the CERG is regulated to follow the peak cooling load (modulation). This represents, actually, a case that can be encountered quite often in existing plants, whose performance is thoroughly analysed from a technical point of view, for instance in [63,64].

The duration curves and the user's trigeneration lambda for this case are shown in figure 24. The lambda transforms for both the WARG and the CERG (table 7) are applied to the duration curves and the cogeneration ratio accounting for the considered regulation strategy. The results, as expected, represent something in the middle between *Case 2* and *Case 3*. In practice, the thermal curve is somehow "flattened" and "regularized" with respect to the case of thermal-only equipment (WARG, *Case 3*); at the same time, also the electrical load is modified with respect to the case with cooling production through electrical-only equipment (CERG, *Case 2*). This intermediate situation could enable to better exploit the prime mover to cover either or both of the aggregated electrical and aggregated thermal loads.

8.2.5. Case 5: CERG (Base-Load) + WARG (Modulation)

The same cooling machines as in *Case 4* are used. However, in this case the cooling-side regulation strategy is such that the CERG follows the base-cooling load, whereas the WARG is used for the cooling load modulation. The duration curves and user's trigeneration lambda are shown in figure 24, after applying the relevant lambda transforms as in *Case 4*, but with "opposite" regulation strategy. The results, as expected, represent again something in the middle between *Case 2* and *Case 3*. Indeed, as in *Case 4*, both the effects of the additional thermal and electrical loads due to the cooling generation are mitigated with respect to the WARG-only and CERG-only cases, respectively, as apparent from visual inspection of figure 24.

8.3. Discussion on the Prime Mover Selection

The five cases illustrated above provide in practice the results of the lambda analysis as a means to characterize the user's need by applying the relevant lambda transforms (figure 24).

Following the steps indicated in Section 7, it is further possible to attempt to work out a possible choice for the cogenerator and its performance, also according to different regulation strategies. In this respect, for the sake of simplicity, and as in most applications, in the following considerations the possibility of incorporating the heat production in a district heating network is not included, whereas connection to the electrical grid is explicitly taken into account.

Hence, in general the thermal demand can be satisfied in part by the heat produced in cogeneration and in part, if necessary, by a CHG. For instance, the CHG can be composed, in *Case 1* and *Case 2*, of two 200 kW$_t$ boilers in parallel with rated efficiency η_t =0.9, which ensure also adequate back-up reserve (suitable in applications such as hospitals or hotels). The part-load model for this equipment is shown in table 7. However, if WARG units are used in the CGP (*Case 3*, *Case 4* and *Case 5*), the number of boilers needs to be adequately increased in order to guarantee complete back-up for the cooling production as well (table 7). At this point, looking up to the different cases, it could be possible to search for a suitable prime mover from commercial catalogues. In this outlook, the approach illustrated is consistent with classical methods based on load duration curve analyses [4], as well as with newer techniques, such as the one based on aggregated thermal loads suggested for planning WARG-based trigeneration systems in [12], on whose basis the cogenerator selection could be readily carried out.

An alternative approach, shown in the sequel, consists of selecting a certain prime mover roughly suitable for all the different CGP cases, and to work out the potential plant performance for the different cases. For instance, let's consider a 330 kW$_e$ gas-fed ICE, whose technology guarantees high electric efficiency also at partial loads and good dynamic performance. Its size is a good choice for matching the *electrical* demand (the electricity-only load) among the alternatives commercially available. The partial load characteristics of the selected ICE are shown in table 7. In particular, the prime mover λ_y is about 1.2 at the rated power, and increases at partial loads (see also Section 3.5).

In general, at first approximation the following considerations can be carried out, case by case, about the possible prime mover behavior, the suitable regulation strategy, and the estimated plant performance:

- GARG (*Case 1*): in this case, being $\lambda_y > \lambda_d$ in the summertime and in the midseason, the engine would mostly find itself operating, if run under electrical load-following mode, wasting part of the thermal power cogenerated; the alternative would be to adopt a heat-following strategy, which would imply electricity sale to the grid, to be further evaluated.
- CERG (*Case 2*): changing the cooling equipment to a CERG would bring about an additional electrical load, as discussed above. Therefore, holding the same engine, there would be an increase in the number of operation hours (under electrical load-following mode), and the engine would also work closer to its rated conditions (at higher electrical efficiency) for longer time, thanks to the higher electrical load. This would mean in general overall performance better than with a GARG, although still with thermal waste.

- WARG (*Case 3*): a hot-water absorption machine could allow for a reduction of the thermal waste, thanks to the increase in the thermal load. However, in this case, the regulation strategy could rather be turned into thermal load-following and the electricity overproduced could be again sold to the grid. In case, it might also be viable to increase the size of the engine in the light of a thermal, rather than electrical, load-following strategy, provided that it is profitable to sell the overproduced electricity. Of course, a longer time interval with partial load operation should be taken into account, as well as the risk to switch off the engine more often and for longer time, if subject to constraints on the minimum load level.
- WARG (base load) + CERG (modulation) (*Case 4*) and CERG (base load) + WARG (modulation) (*Case 5*): in these combined solutions, the overall performance should improve, no matter whether adopting electrical or thermal load-following strategies; in fact, both the equivalent load duration and lambda curves are flattened with respect to the cases with cooling production only based on electrically-fed or thermally-fed equipment, thus allowing better exploitation of the prime mover sizing. In fact, balancing somehow the thermal load and the electrical load (which is carried out through combination of cooling generation equipment) enables to supply the load more effectively by avoiding waste of cogenerated thermal power and thus improving the prime mover operation for longer time.

Other alternatives could take into account the possibility of splitting the prime mover into more units, for example a cluster [45,53,72] of 100-kW$_e$ microturbines. Operating simultaneously the MTs, with the possibility of exploiting them in different partial load conditions, could enable a better "match" between the overall λ_y^{CHP} (MT cluster + CHG) and λ_d^z. Of course, the λ_y^{CHP} would not depend only upon the machine characteristics any longer, but it would also be related to the number of MTs used and on their regulation strategy. These options should be analyzed in further details through time-domain simulations.

CONCLUSION

The diffusion of distributed multi-generation systems for combined production of manifold energy vectors (such as electricity, heat and cooling power) on a small scale is continuously increasing, owing to the excellent energy, environmental and economic performance of these systems. At the same time, the adequate planning of a multi-generation system represents both a challenging task as well as a modern issue to cope with in the worldwide energy scenario.
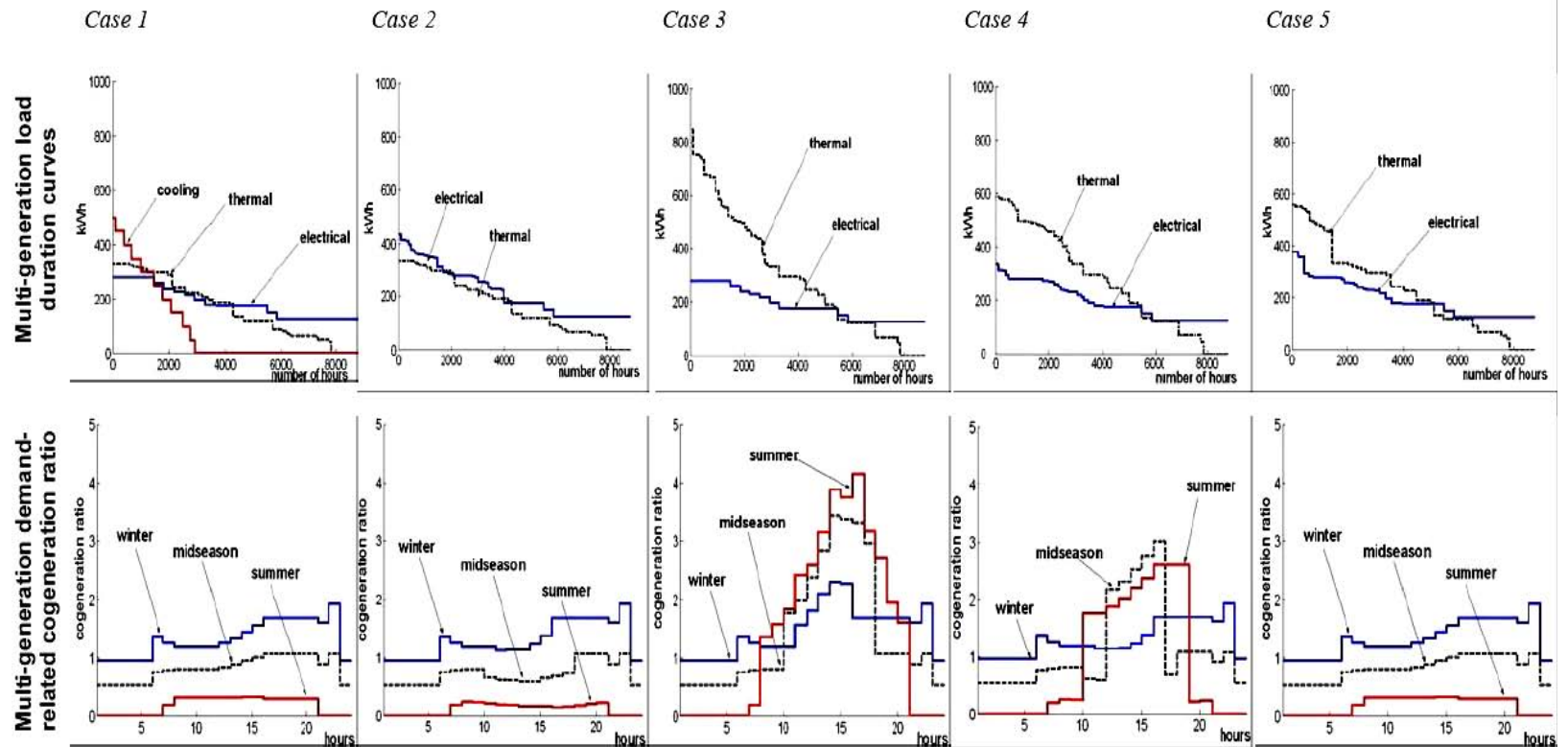
Figure 24. Multi-generation load duration curves and cogeneration ratios after applying the relevant lambda transforms for the case study application.

In this chapter, a novel and comprehensive approach to multi-generation system characterization and planning, with specific reference to cogeneration and trigeneration, has been presented. This approach, called *lambda analysis*, is formulated in mathematical terms by introducing the so-called *lambda transforms*, that is, transformations of the relevant loads into equivalent loads, with the objective of facilitating the selection and sizing of the equipment involved in the analysis.

More specifically, the *cogeneration lambda analysis* focuses on characterizing the cogeneration loads and the cogeneration production taking into account time-domain aspects (for instance, seasonal, daily, and hourly load variations), as well as off-design aspects (for instance, variation of the prime mover cogeneration characteristics at partial load), thus revisiting from a modern standpoint "classical" tools such as the cogeneration ratio indicator widely adopted for cogeneration studies. In addition, the interactions with energy networks (and, in case, related markets) such as the electrical grid, district heating and district cooling networks, whose presence is intrinsic in the concept of "distributed" energy system, are adequately modeled in the novel approach formulated.

The *trigeneration lambda analysis* is a further extension including the new aspects brought about by the need for cooling power generation. In particular, if different alternative pieces of equipment are considered within the CGP, each of these pieces can impact differently on the thermal and/or electrical load, so as to make it necessary to build new models for the energy system characterization.

In a more general case, the CGP turns into an AGP that also envisages adopting equipment for heat generation, so that the interaction between the production side and the demand side (both including flow exchanges with energy networks) becomes even more intertwined, and can be adequately described within the so-called *multi-generation lambda analysis*.

After introducing the theoretical framework, the relevant formulas to apply the lambda transforms have been presented for several cooling and heat generation equipment and plant schemes. In addition, different multi-generation cases have been analyzed, pointing out the suitability of the specific cases to particular load configurations, and thus the possibility of exploiting the multi-generation lambda analysis as a planning tool. From a more practical standpoint, some numerical applications exemplificative of the concepts presented have been provided in the case study application, in order to underline, also from a graphical point of view, the use and the results of the lambda transforms.

The lambda analysis, as a first step in the energy system characterization and planning procedure, provides a useful representation of the energy flow interactions within the plant and with the external energy networks in a synthetic and at the same time comprehensive way. As such, it represents a powerful tool to be exploited at a planning stage and has the upside of taking into account aspects such as time-domain variations of the loads and off-design characteristics of the equipment, paving the way to more detailed analyses, based for instance upon time-domain simulations, also facing further issues related to energy saving and economic aspects [23-26,71].

One of the most promising directions for future research consists of including within the lambda analysis approach the possibility of shifting the electrical, thermal and cooling loads by means of *energy storage* devices [7,30,31,43,48,73,74], evaluating the related impact in terms of energy efficiency and of economic profitability.

REFERENCES

[1]   The IEEE. IEEE Standard 1547: IEEE Standard for Interconnecting Distributed Resources with Electric Power Systems. New York, 28th July 2003.

[2]   Jenkins, N; Allan, R; Crossley, P; Kirschen, D; Strbac, G. Embedded generation. *IEE Power and Energy Series* No.31, The IEE, London, 2000

[3]   Borbely, AM; Kreider, JF (Eds.). *Distributed generation: the power paradigm of the new millennium.* Boca Raton, Florida, USA: CRC Press; 2001.

[4]   Willis, HL; Scott, WG. *Distributed power generation: planning and evaluation.* New York: Dekker; 2000.

[5]   EDUCOGEN. The European Educational Tool on Cogeneration; December 2001. Available at www.cogen.org/projects/educogen.htm.

[6]   US EPA. Catalogue of CHP Technologies. Available www.epa.gov.

[7]   Danny Harvey, LD. A handbook on low energy buildings and district energy systems. UK: James and James; 2006.

[8]   Horlock, JH. Cogeneration-Combined Heat and Power (CHP). Malabar, FL: Krieger; 1997.

[9]   Cogeneration technologies website: www.cogeneration.com.

[10]   Martens, A. The energetic feasibility of CHP compared to the separate production of heat and power. *Applied Thermal Engineering,* 1998; 18: 935-946.

[11]   Trigeneration technologies website: www.trigeneration.com.

[12]   Cardona, E; Piacentino, A. A method for sizing a trigeneration plant in Mediterranean areas. *Applied Thermal Engineering,* 2003; 23: 1665-1680.

[13]   Ziher, D; Poredos, A. Economics of a trigeneration system in a hospital. *Applied Thermal Engineering,* 2006; 26: 680-687.

[14]   Maidment, GG; Tozer, RM. Combined cooling heat and power in supermarkets, *Applied Thermal Engineering,* 2002; 22: 653-665.

[15]   Cardona, E; Piacentino, A; Cardona F. Energy saving in airports by trigeneration. Part I: Assessing economic and technical potential. *Applied Thermal Engineering,* 2006; 26: 1427-1436.

[16]   Cardona, E; Sannino, P; Piacentino, A; Cardona, F. Energy saving in airports by trigeneration. Part II: Short and long term planning for the Malpensa 2000 CHCP plant. *Applied Thermal Engineering, 2006;* 26: 1437-1447.

[17]   Bassols, J; Kuckelkorn, B; Langreck, J; Schneider, R; Veelken, H. Trigeneration in food industry. *Applied Thermal Engineering,* 2002; 22: 595-602.

[18]   Resource Dynamics Corporation. *Cooling, Heating, and Power for Industry: A Market Assessment.* Final Report for U.S. Department of Energy, August 2003. Available at: www.cogeneration.net.

[19]   Colonna, P; Gabrielli, S. Industrial trigeneration using ammonia-water absorption refrigeration systems (AAR). *Applied Thermal Engineering,* 2003; 23: 381-396.

[20]   Havelsky, V. Energetic efficiency of cogeneration systems for combined heat, cold and power production, *International Journal of Refrigeration,* 1999; 22: 479-485.

[21]   Maidment, GG; Prosser, G. The use of CHP and absorption cooling in cold storage. *Applied Thermal Engineering,* 2000; 20: 1059-1073.

[22] Hernandez-Santoyo, J; Sanchez-Cifuentes, A. Trigeneration: an alternative for energy savings. *Applied Energy,* 2003; 76: 219-227.

[23] Chicco, G; Mancarella, P. Planning Aspects and Performance Indicators for Small-scale Trigeneration Plants. *Proc. Future Power Systems* (FPS) 2005 Conference. Amsterdam, The Netherlands; 2005; paper No. O01-05.

[24] Chicco, G; Mancarella, P. From cogeneration to trigeneration: profitable alternatives in a competitive market. *IEEE Transactions on Energy Conversion,* 2006; 21: 265-272.

[25] Chicco, G; Mancarella, P. Planning Evaluation and Economic Assessment of the Electricity Production from Small-scale Trigeneration Plants. *WSEAS Transactions on Power Systems,* 2006; 1: 393-400.

[26] Mancarella, P. From cogeneration to trigeneration: energy planning and evaluation in a competitive market framework. Doctoral Thesis. Politecnico di Torino, Torino, Italy; 2006.

[27] Tozer, R; James, RW. Heat powered refrigeration cycles. *Applied Thermal Engineering,* 1998; 18: 731-743.

[28] Foley, G; De Vault, R; Sweetser, R. Future of Absorption Technology in America: A Critical Look at the Impact of BCHP and Innovation. Oak Ridge National Laboratory, June 2000. Available at http://www.eere.energy.gov/de/publications.html.

[29] US DOE. Review of Thermally activated technologies. Distributed energy program report, July 2004. Available at http://www.eere.energy.gov/de/publications.html.

[30] Kreider, JF (ed.). *Handbook of Heating, Ventilation and Air Conditioning.* Boca Raton, Florida, USA: CRC Press; 2001.

[31] American Society of Heating, Refrigerating and Air-conditioning Engineers. *ASHRAE HVAC Systems and Equipment Handbook,* SI Edition. USA: ASHRAE; 2000.

[32] Voorspools, KR; D'haeseleer WD. The impact of the implementation of cogeneration in a given energetic context. *IEEE Transactions. on Energy Conversion,* 2003; 18: 135-141.

[33] Chevalier, C; Meunier, F. Environmental assessment of biogas co- or tri-generation units by life cycle analysis methodology. *Applied Thermal Engineering,* 2005; 25: 3025-3041.

[34] Ameri, M; Hejazi, SH. The study of capacity enhancement of the Chabahar gas turbine installation using an absorption chiller. *Applied Thermal Engineering,* 2004; 24: 59-68.

[35] Hwang, Y. Potential energy benefits of integrated refrigeration system with microturbine and absorption chiller. *International Journal of Refrigeration,* 2004; 27: 816-829.

[36] Onovwiona, HI; Ugursal, VI. Residential cogeneration systems: review of the current technologies. *Renewable and Sustainable Energy Reviews,* 2006; 10: 389-431.

[37] Hanssen, JH et al.. Operating experiences from 18 microturbine applications for CHP and industrial purposes. Danish Gas Technology Centre, Report C0404. Av. http://www.dgc.dk/publikationer/konference/jdw_igrc04.pdf.

[38] Pilavachi, PA. Mini- and micro-gas turbines for combined heat and power. *Applied Thermal Engineering,* 2002; 22: 2003-2014.

[39] Kaikko, J; Backman, J. Technical and economic performance analysis for a microturbine in combined heat and power generation. Energy, 2007; 32: 378-387.

[40] Colombo, PML; Armanasco, F; Perego, O. Experimentation on a cogenerative system based on a microturbine. *Applied Thermal Engineering,* 2007; 27: 705-711.

[41] De Ruyck, J; Delattin, F; Bram, S. Co-utilization of biomass and natural gas in combined cycles through primary steam reforming of the natural gas. *Energy,* 2007; 32: 371-377.

[42] Fiaschi, D; Carta, R. CO_2 abatement by co-firing of natural gas and biomass-derived gas in a gas turbine. *Energy,* 2007; 32: 549-567.

[43] Wu, DW; Wang, RZ. Combined cooling, heating and power: A review. *Progress in Energy and Combustion Science*, 2006; 32: 459-495.

[44] Zhang, RR; Lu, XS; Li, SZ; Lin, WS; Gu, AZ. Analysis on the heating performance of a gas engine driven air to water heat pump based on a steady-state model. *Energy Conversion and Management,* 2005; 46: 1714-1730.

[45] Lazzarin, R; Noro, M. District heating and gas engine heat pump: Economic analysis based on a case study. *Applied Thermal Engineering,* 2006; 26: 193-199.

[46] Lazzarin, R; Noro, M. Local or district heating by natural gas: Which is better from energetic, environmental and economic point of views? *Applied Thermal Engineering,* 2006; 26: 244-250.

[47] Mancarella, P. Small-scale trigeneration systems: components, models and applications. *Proc. VI World Energy System Conference.* Torino, Italy, July 10-12, 2006; 234-241.

[48] Wulfinghoff, DR. *Energy Efficiency Manual*. Wheaton, Maryland, USA: Energy Institute Press; 1999.

[49] Lozza, G. Gas Turbines and combined cycles. Bologna, Italy: Progetto Leonardo; April 2000 (in Italian).

[50] Gomatom, P; Jewell, W. Fuel parameter and quality constraints for microturbine distributed generators. Proc. 2003 *IEEE/PES Transmission and Distribution Conference and Exposition*, Dallas, Texas, USA, 7-12 Sept. 2003; 1: 409-412.

[51] Zaltash, A; Petrov, AY; Rizy, DT; Labinov, SD; Vineyard, EA; Linkous, RL. Laboratory R and D on integrated energy systems (IES). *Applied Thermal Engineering,* 2006; 26: 18-35.

[52] Bua, F; Calderazzi, L. Design and building of an industrial cogeneration plant. *Power Technology,* 2003; 4 (in Italian).

[53] Barsali, S; Ceraolo, M; Figlioli, R; Pelacchi, P. Microturbines for dispersed generation. *Proc. CIRED* 1999. Nice, France, June 1-4, 1999.

[54] Bruno, JC; Valero, A; Coronas, A. Performance analysis of combined microgas turbines and gas fired water/LiBr absorption chillers with post-combustion. *Applied Thermal Engineering,* 2005; 25: 87-99.

[55] Internal Combustion Engines, Technical data from commercial catalogues. March 2006.

[56] Chicco, G; Mancarella P. Beyond the Electricity-only Production: Towards a Distributed Multi-Generation World. *Proc. International Conference on Power Engineering, Energy and Electrical Drives* (POWERENG 2007). Setúbal, Portugal, 12-14 April 2007; 219-224.

[57] Li, H; Nalim, R; Haldi, PA. Thermal-economic optimization of a distributed multi-generation energy system – A case study of Beijing. *Applied Thermal Engineering,* 2006; 26: 709-719.

[58] Meunier, F. Co- and tri-generation contribution to climate change control. *Applied Thermal Engineering,* 2003; 22: 703-718.

[59] Lin, L et al. An experimental investigation of a household size trigeneration. *Applied Thermal Engineering*, 2007; 27: 576-585.

[60] Heteu, PMT; Bolle, L. Economie d'énergie en trigenération. *International Journal of Thermal Sciences,* 2002; 41: 1151-115.

[61] Macchi, E; Campanari, S. Potential development for gas microturbines. Hybrid cycles and trigeneration. *Cogeneration and On-Site Power Production,* 2001; 2: 41-49.

[62] Rong, A; Lahdelma, R. An efficient linear programming model and optimization algorithm for trigeneration. *Applied Energy,* 2005; 82: 40-63.

[63] Minciuc, E; Le Corre, O; Athanasovici, V; Tazerout, M; Bitir, I. Thermodynamic analysis of tri-generation with absorption chilling machine. *Applied Thermal Engineering,* 2003; 23: 1391-1405.

[64] Minciuc, E; Le Corre, O; Athanasovici, V; Tazerout, M. Fuel savings and CO_2 emissions for trigeneration systems. *Applied Thermal Engineering,* 2003; 23:1333-46.

[65] Lucas, K. Efficient energy systems on the basis of cogeneration and heat pump technology. *International Journal of Thermal Sciences,* 2001; 40: 338-343.

[66] Berntsson, T. Heat sources – technology, economy and environment. *International Journal of Refrigeration,* 2002; 25: 428-438.

[67] Marcic, M. Long-term performance of central heat pumps in Slovenian homes. *Energy and Buildings,* 2004; 36: 185-193.

[68] Ossebaard, ME; Van Wijk, AJM; Van Wees, MT. Heat supply in the Netherlands: a systems analysis of costs, exergy efficiency, CO_2 and NO_x emissions. *Energy,* 1997; 22: 1087-1098.

[69] Kaynakli, O; Kilic, M. Theoretical study on the effect of operating conditions on performance of absorption refrigeration system. *Energy Conversion and Management,* 2007; 48: 599-607.

[70] Burer, M; Tanaka, K; Favrat, D; Yamada, K. Multi-criteria optimization of a district cogeneration plant integrating a solid oxide fuel cell-gas turbine combined cycle, heat pumps and chillers. *Energy,* 2003; 28: 497-518.

[71] Chicco, G; Mancarella, P. Enhanced energy saving performance in composite trigeneration systems. *IEEE Power Tech 2007 Conference.* Lausanne, Switzerland, 1-5 July 2007, paper No.526.

[72] Nesreddine, H; Dostie, M; Millette, J; Poulin, A; Handfield, L. Economic assessment of microturbine implementation in commercial and institutional buildings. *WSEAS Transactions on Power Systems,* 2006; 1: 1563-1572.

[73] Hasnain, SM. Review on sustainable thermal energy storage technologies, Part I: heat storage materials and techniques. *Energy Conversion and Management*, 1998; 39: 1127-1138.

[74] Hasnain, SM. Review on sustainable thermal energy storage technologies, Part II: cool thermal storage. *Energy Conversion and Management,* 1998; 39: 1139-1153.

Chapter 2

GENETIC ALGORITHM APPLICATIONS TO THE SOLUTION OF ELECTRIC POWER SYSTEM PROBLEMS

G.P. Granelli and M. Montagna

University of Pavia, Department of Electrical Engineering
Via Ferrata, 1 - Pavia, Italy

ABSTRACT

Genetic algorithms, proposed about 40 years ago, have been used as a general purpose optimization technique. In this work, the authors' experience with genetic algorithm based optimization is presented with reference to some electric power system analysis problems.

At first, the fundamentals of genetic algorithms are described: the basic genetic algorithm by John Holland is presented and the function of the three genetic operators of selection, crossover and mutation is discussed. Among the more recent theoretical developments, the micro-genetic approach by K. Krishnakumar and the algorithm of Chu and Beasley are considered. The former takes advantage from operating with small-sized populations and the latter proposes an effective technique to deal with functional constraints.

The second part of this work is concerned with the description of some applications of the above mentioned genetic algorithm based procedures to power system problems. The topics that are considered in detail are:

- allocation of compensating capacitors in high and medium voltage networks to improve voltage regulation;
- optimization of the topology of EHV networks with the aim of improving security and to overcome the problem of parallel or loop flows; control measures include switching of substation breakers as well as deployment and operation of phase shifting transformers;
- identification of multiple interacting bad data in state estimation, formulated as a combinatorial optimization problem.

The above mentioned items represent a clearly non-exhaustive list of the many application fields where genetic algorithms have been profitably employed, but we feel they demonstrate the main advantage of using genetic algorithms with respect to other search methods. Indeed no present day optimizer is so general, robust and flexible to deal with problems so different from each other as the ones considered here.

1. INTRODUCTION

Genetic algorithms (GAs) were invented by John Holland about 40 years ago. The aim was twofold: first, the creation of an artificial environment for the study of evolution in nature; secondly, the identification of techniques capable of including the mechanisms of natural adaptation within information systems. In his book "Adaptation in Natural and Artificial Life" [1], genetic algorithms are presented as a metaphor of the adaptation occurring in nature by which a population of individuals evolves in time thanks to a sort of "natural selection" obtained by means of suitable genetic operators.

Each individual is provided its own artificial DNA consisting of genes, each taking a particular value (for example 0 or 1 in case of binary coding).

A "selection" operator chooses the individuals for reproduction among those having the highest fitness as it happens in nature for those organisms which are best fitted to the natural conditions.

For each couple of selected individuals, a "crossover" operator produces the exchange of part of their DNA to create the next generation of population individuals, imitating the DNA recombination of "haploid" organisms.

Then, a "mutation" operator can modify the value of one or more DNA elements of any individual in a random fashion, but with a very low probability.

To protect the peculiar characteristics of any species, the information necessary to replicate an individual in its integrity is memorized within its DNA. Through generation, this information is inherited thus allowing children to have characteristics similar or possibly superior to those of their parents. By this mechanism, nature provides all living species with the ability to safeguard the characteristics of those individuals which, by natural selection, have developed the capacity to best suit the features of the natural environment.

The same way as nature codes "life" into a DNA molecule, genetic algorithms require problem variables to be suitably coded; the features of this artificial coding strictly depend on the characteristics of the problem at hand. Binary and decimal coding are the most used techniques.

A great variety of modifications and extensions were introduced into the original proposal of Holland; anyway, most present day genetic algorithms maintain the three basic operators of selection, crossover and mutation.

To understand the behavior of genetic algorithms, it is necessary to clearly define what is meant by "fitness" of an individual.

In the case of living organisms, fitness can be defined as the attitude of a given individual to face the survival problems imposed by the environment. The same definition can be extended to the theory of genetic algorithms, where each individual in a population is denoted by a score assigned to it on the basis of the corresponding value of an objective function to be optimized.

For example, if a function $f(x)$ is to be maximized with respect to a bounded variable x, the genetic algorithm will start by creating a random initial population consisting of a pre-assigned number of bit strings each corresponding to the binary coding of a value of the real variable x. Then, to evaluate the fitness of an individual, it is necessary to decode the corresponding binary string to obtain the value x which will be substituted into $f(x)$. The fitness of the individual is the value taken by the function to be maximized for the value of the x variable corresponding to that particular binary string. The closer the x value to the maximum of $f(x)$, the grater will be the fitness of the corresponding string.

Nowadays, the applications of genetic algorithms are countless and there is almost no field of engineering and applied science where they are not in use.

With reference to electric power system analysis problems, only a significant even though non-exhaustive selection of papers on the applications of GAs is briefly described in the sequel.

GAs have been applied to almost all fields of power system planning including transmission and distribution network planning [2–4] and generation planning [5,6]. A consistent number of works has been also devoted to the optimal allocation of compensating devices (capacitors in most cases), with the aim of improving voltage profiles and reactive power balance [7–11].

The unit commitment and the short term hydro-thermal coordination problems have been favourite topics in the application of GAs. In these applications, GAs can be considered as a valuable alternative to mixed integer or dynamic programming for the solution of large-scale problems with both real and integer variables [12–16].

Coming closer to the real time environment, GAs have been proposed as a tool for a more realistic economic dispatch; here GAs have been demonstrated capable of correctly taking into account the problem of generator valve point loading and multi-fuel unit dispatching [17–20].

Finally, GAs have found many applications to the modelling of energy market behavior and GENCO strategies for maximum revenue bidding [21–26].

In this work, the experience of the authors in the application of GA optimization techniques is presented.

After a description of the basic GA and of some recent variants, the application of GAs to optimal capacitor allocation [27] and to the topological rescheduling of large interconnected systems [28] will be considered. Then some GA based techniques for the identification of bad data in the framework of the static state estimation problem [29,30] will be presented. Results relative to IEEE test systems and to actual networks of Italian and European origin will be shown and commented.

2. THE ROLE OF GENETIC OPERATORS

In this Section the basic role of the three operators of selection, crossover and mutation is discussed highlighting the impact they have on the genetic algorithm behavior.

The selection operator has the task of selecting those individuals of a given population which will take part in reproduction; the higher the fitness of each individual, the larger the

probability of being selected. The "roulette wheel" selection method is the one proposed by Holland: a sector of an imaginary roulette wheel is assigned to each individual with a size proportional to its fitness. This is still the most used approach; however, the so-called "tournament" based selection strategy is also widely employed. According to the tournament method, selection is carried out by randomly choosing a number of individuals; the one selected for reproduction is the best in this reduced set. Compared with the roulette wheel method, tournament selection has the advantage of being more easily adapted to function minimization.

The crossover operator is based on the concept that the parent solutions, coded by the strings of a couple of individuals selected for reproduction, may be combined aiming at producing children solutions possibly with a higher fitness. The crossover operation simulates the process of reproduction between haploid organisms, although with some simplifications. The two most widely adopted versions of the crossover function are single-point and uniform crossover. According to single-point crossover, the two parent individuals exchange the sub-string lying left of a randomly chosen crossover point. One of the two children obtained will contain the entries of the first parent up to the crossover point and those of the second from the crossover point onward; the other children will contain the two complementary sub-strings. The uniform crossover scheme implies that string entries can be exchanged, according to a pre-assigned probability, in any location of the strings belonging to the parent individuals.

The mutation operator slightly alters genetic coding of randomly chosen individuals in the population to prevent premature convergence toward a sub-optimal point which may be far from the true optimal solution. The initial population of a genetic algorithm generally represents only a small fraction of the variety of possible combinations corresponding to the whole search space. By means of mutation, the algorithm succeeds in exploring different parts of the search space since the gene pool of the current population is always different from that of preceding generations. Mutation also provides a way to avoid the so-called "founder effect", a sort of genetic drift that occurs in small isolated populations. In terms of fitness, the result of mutation on a single individual may be favourable or not. If mutation has the effect of creating a new individual with a fitness value higher than average, that individual will probably be rewarded by selection, otherwise it will be rejected and it will have no off-springs.

3. OPTIMIZATION BY GENETIC ALGORITHMS

Genetic algorithms differ from deterministic optimization procedures for the following characteristics:

- they operate on codes representing variables, not on the corresponding values;
- they search for the optimal solution by using a population of individuals, corresponding to different points of the feasibility domain, instead of operating on a single point at a time;

- they use only the information available from the value of the objective function relative to each member of the population; in particular, they do not require the computation of derivatives;
- they employ probabilistic instead of deterministic transition rules.

Moreover, GAs differ from random search methods (and notably the Monte Carlo algorithm) in that probabilistic transition rules are employed as a means to guide the search and not just to provide a random walk through the feasibility domain.

The above considerations explain the robustness of genetic algorithms for what concerns the search for the global optimum point. Deterministic methods are not equally robust since they often look for the optimal solution by exploiting information derived from the local gradient vector. As an example, the function $Z = f(X,Y)$ shown in figure 1 presents some local optima but only one global optimum; it can be easily realized that, by starting search in the vicinity of the lower peak, a deterministic algorithm is likely to converge toward a local sub-optimal solution thus missing the global optimum. For global optimization, the procedure should be iterated each time starting from a different initial point. Moreover many deterministic methods require the objective function to be derivable; however, quite often the objective function happens to be neither continuous nor derivable and its feasibility domain may lack the convexity requirements needed for global optimization.

Genetic algorithms are among the most general search methods now available; this, however, may penalize their performance in comparison with deterministic methods specifically designed for some class of problems.

Apart from their excellent robustness, it must be noted that the convergence of genetic algorithms toward the global optimum may be obtained only after a thorough exploration of the space of feasible solutions and, therefore, at the price of a remarkable computation and CPU time burden. This makes genetic algorithms suited for tough combinatorial problems arising in planning and operational planning studies relative to electrical power systems; they are seldom proposed for on-line applications.

Another limit of genetic algorithms is their inability to handle functional constraints directly; while upper-lower bounds can be accounted for implicitly by the coding mechanism,
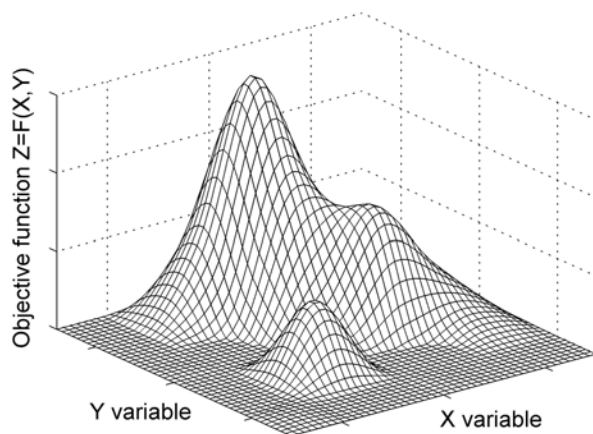
Figure 1. Local and global optima.

functional constraints are generally handled by penalization of the objective function.

However, some recent GA variants have been proposed to overcome the above mentioned drawbacks. In particular, the "micro-genetic" algorithm [31] works with a small population to gain solution speed; the method proposed by Chu and Beasley [32] is designed to cope with functional constraints by introducing the concept of solution "unfitness".

3.1. The Basic Genetic Algorithm

The above general considerations can be summarized in the concise algorithmic form reported in the sequel. This basic genetic algorithm (BGA) may be viewed as the simplest implementation of the original ideas of Holland and is presented in details in [33] together with the corresponding programming code (in the Pascal language).

BGA, with binary coding, can be stated in the following form:

1. Select an initial population of randomly generated individuals each represented by a string of binary digits. Let s be the population size.
2. Evaluate fitness of all members of the population by the objective function. This allows estimation of the probability of each individual to be selected for reproduction.
3. Select individuals for reproduction (by the roulette wheel or by the tournament method). Apply crossover (with a probability p_c) and mutation (with a probability p_m). Create the new population and go back to Step 2.

The choice of the population size s largely depends on the characteristics of the problem at hand; larger populations lead to better ultimate convergence to the optimum because of the larger genetic diversity of the population. On the other hand, the genetic inertia of a large population leads to a poorer initial performance of the algorithm. Smaller populations generally exhibit a faster initial convergence, but may rapidly loose their genetic diversity and therefore the ability to find the global optimum. Population sizes in the range of a few tens are common.

The crossover probability p_c establishes the expected number of crossover operations to be performed. At each generation, an average of $p_c \cdot s$ individuals in the new population will receive their genes from crossover of the artificial DNA of their parents. Crossover probability may range between 0.25 and 1 with a preference for higher values which guarantee a better spread of genetic information among the population members.

The mutation probability p_m establishes the expected number of mutation operations occurring at each generation. It is commonplace to take $p_m = k_m / s$, thereby assuming that, on average, k_m individuals (no more than two or three) of the new population undergo mutation. It is not infrequent to adopt larger mutation probabilities ($p_m \cong 0.1$) to maintain sufficient diversity in small populations (in the order of a few tens) such as those generally considered in genetic algorithms.

It must be noted that there are no stopping rules in BGA; once more, this is an effect of the biological metaphor underlying genetic algorithms. Indeed, the biology of species does not show anything similar to algorithmic convergence; living organisms, even when perfectly adapted, always maintain the capacity to change their gene pool to react against changes occurring in the natural environment.

Generally, it is practical to set a limit to the maximum number of generations or to the CPU time used in the computation. Adaptive stopping rules have also been proposed and used by some Authors; these convergence criteria set an upper bound on the number of generations spent with no improvement in the fitness of the best individual of the population.

Several modifications have been proposed aiming at improving the characteristics of the simple BGA and, in particular, its speed and convergence reliability. One of the strategies often used for this purpose is the so-called "elitist" strategy: at each iteration of BGA, the best individual is identified and it is transferred to the next generation. This has no counterpart in biology, but it guarantees against the loss of good information embedded in the genes of the best individual produced. Generally the elitist strategy results in an increased efficiency of the optimization process.

3.2. Micro-Genetic Algorithm

Many variants of the simple BGA were proposed to suit the needs of the researchers working in the field of optimization; one of the most successful is the micro-genetic algorithm (μGA) [31]. The main features of μGA are briefly presented.

Most of the genetic algorithms perform poorly with small sized populations because of insufficient information processing and premature convergence toward non-optimal results. For this reason, the population size is generally assumed in the range of few tens at least. In contrast, μGA explores the possibility of employing smaller populations with a potential advantage deriving from the reduction of the computation burden compared with the BGA procedure. A population of 5 individuals is typically used according to the μGA approach. From the genetics viewpoint, it is known that frequent interbreeding within a small population may favor the spread of hereditary diseases rarely found in larger populations. On the other hand, small populations sometimes act as natural laboratories where desirable genetic characteristics may rapidly emerge. The adoption of some suitable strategy to prevent the loss of population diversity is crucial to the success of μGA. The procedure proposed in [31] consists in a periodic refreshment of the population whenever the genetic diversity of the population has decreased below a given threshold.

The following algorithm presents the basic steps of μGA:

1. Select a population of randomly generated individuals. Alternatively, individuals may be generated randomly together with one good individual obtained from previous search.
2. Evaluate fitness and determine the best individual which is always transferred to the next generation, i.e. the elitist strategy is adopted to guarantee against the loss of good information embedded in the best individual produced thus far.
3. Select individuals for reproduction by the tournament selection strategy.

4. Apply crossover with probability equal to 1 to favor exchange of genetic information among the population members.
5. Check for convergence by measuring the amount of diversity left in the population (by counting the total number of bits which are unlike those possessed by the best individual). If population diversity has fallen under a pre-selected threshold, go to Step 1; otherwise, go to Step 2.

Mutation is not needed according to μGA since enough genetic diversity is introduced after every convergence when all the population (with the exception of the best individual) is randomly recreated afresh.

3.3. Chu-Beasley Algorithm

The genetic algorithm proposed by Chu and Beasley [32] improves the performance of BGA with reference to those problems where functional constraints are present.

As previously stated, functional constraints cannot be easily handled by BGA other than by the penalization method. However, penalization is never simple to implement because both the penalization function and its parameters need a careful and often adaptive tuning to be successful in practice.

The approach of Chu and Beasley (CB-GA), deals with functional constraints in a simpler way which is also more consistent with the principles of optimization by genetic algorithms. The idea of the method is quite simple; instead of employing penalization, two objective functions are associated with each individual. The former, is properly the fitness defined by the objective function; the latter, named the "unfitness", is a measure of each individual's infeasibility expressed as the sum of functional constraint violations.

CB-GA systematically exploits the elitist strategy by imposing very stringent conditions for the acceptance of children solutions; indeed, each newly created individual becomes a member of the current population only if it outperforms the worst individual (which is replaced by it). In compliance with the elitist strategy, the worst individual is the one with the largest unfitness. If all members of the population are feasible, the worst individual is the one with the lowest fitness value. Duplicated individuals are never accepted in the population and therefore need not be tested.

CB-GA can be stated algorithmically as follows:

1. Select an initial population of randomly generated individuals each represented by a string of binary digits.
2. For each individual of the population, evaluate both fitness (by the objective function) and unfitness (by the sum of constraint violations).
3. Select individuals for reproduction by the tournament method; only the corresponding fitness values are involved in this phase.
4. Apply crossover (with a probability $p_c = 1$) and mutation (with a probability p_m).
5. If the newly created individual is a duplicate of another string in the population, it is discarded; otherwise its fitness and unfitness are evaluated. The child solution is accepted only if it outperforms the worst individual in the current population, i.e. the

individual with the highest unfitness or, if all members are feasible, the one with the lowest fitness. The worst individual is replaced by the new one. Go back to Step 3.

4. PLANNING STUDIES

Planning studies are those to which genetic algorithms have been most often applied since the less stringent computation time requirements allow to fully exploit their characteristics of robustness and flexibility.

In particular, the reactive power compensation problem and the topological rescheduling of interconnected power systems are considered in this Section.

The reactive power compensation problem plays an important role in the planning of an electrical system; its aim is principally to provide an appropriate placement of the compensation devices to ensure a satisfactory voltage profile while minimizing the relevant costs.

Topological rescheduling tries to minimize parallel flows through interconnection corridors in large power systems.

4.1. Reactive Power Compensation

The problem of reactive power compensation must take into account generator voltages, on-load tap changing transformer (OLTC) ratios and network topology in addition to localization and sizing of reactive compensation devices. Uncertainties in load and generation increase have a negligible impact on the planning scenario since the time horizon considered is typically only 3–4 years ahead.

The installation of capacitor banks offer a simple and cheap approach to reactive power compensation. Linear and nonlinear programming methods [34] have been proposed to solve the placement problem. In most approaches, the size of capacitor banks has been handled as a continuous variable; however, a realistic formulation of the problem requires the discrete nature of capacitor banks to be accounted for.

The resulting nonlinear programming problem, with mixed (discrete and continuous) variables could be solved by implicit enumeration algorithms or by the nonlinear branch-and-bound technique. Due to the difficulties intrinsic to this kind of approach, a simplified method is considered by exploiting the sequential linear programming (SLP) method. At each iteration, the problem model is linearized; the linear programming problem to be solved presents both continuous variables (voltage magnitudes at PV buses, transformer ratios, reactive generations at PQ buses) and discrete variables (capacitor bank location and relative size). This mixed integer linear programming (MILP) problem can be tackled by deterministic algorithms such as branch-and-bound; it is solved here by BGA with the possible application of the micro-genetic approach. A hybrid procedure is also considered exploiting the synergy between the genetic and the branch-and-bound algorithms.

4.1.1 Mathematical Model

The mathematical model of the capacitor placement (CP) problem can be expressed as follows:

$$\min \sum_{i \in I_5} \left(\chi_i C_{0i} + C_{1i} q_{Ci} \right) \qquad \text{subject to:} \tag{1}$$

$$\underline{Q}_{Gi} \le Q_{Gi} \left(\mathbf{q}_C, \mathbf{q}_G, \mathbf{v}, \mathbf{r} \right) \le \overline{Q}_{Gi} \qquad i \in I_2 \tag{2}$$

$$\underline{V}_i \le V_i \left(\mathbf{q}_C, \mathbf{q}_G, \mathbf{v}, \mathbf{r} \right) \le \overline{V}_i \qquad i \in I_1 \cup I_3 \tag{3}$$

$$\underline{q}_{Gi} \le q_{Gi} \le \overline{q}_{Gi} \qquad i \in I_1 \tag{4}$$

$$\underline{v}_i \le v_i \le \overline{v}_i \qquad i \in I_2 \tag{5}$$

$$\underline{r}_i \le r_i \le \overline{r}_i \qquad i \in I_4 \tag{6}$$

$$0 \le q_{Ci} \le \chi_i \overline{q}_{Ci} \qquad i \in I_5 \tag{7}$$

The value of binary variables $\boldsymbol{\chi}$ depends on whether capacitors are installed ($\chi_i = 1$) or not ($\chi_i = 0$). The size of capacitor banks can be modelled either as a discrete or a continuous variable. The content of sets $I_1, \ldots, I_5$ is specified as follows:

$I_1 = \{\text{PQ type generation buses}\}$

$I_2 = \{\text{PV type generation buses}\}$

$I_3 = \{\text{PQ type sentinel buses}\}$

$I_4 = \{\text{OTLC transformer}\}$

$I_5 = \{\text{buses candidate to reactive compensation}\}$

The objective function in (1) is the overall cost of reactive compensation. For each candidate bus, the cost of reactive compensation consists of the sum of the two following terms:

- A fixed installation cost C_{0i} to be accounted for only when $\chi_i = 1$.
- A cost proportional to the reactive power q_{Ci}, C_{1i} being the per Mvar cost of capacitor banks.

Besides the reactive power compensations q_{Ci} and the binary switches χ_i, the control variables of the problem include the reactive power generations q_{Gi} of PQ type generation

buses, the voltage magnitudes v_i of PV type generation buses and the ratios r_i of the OLTC transformers.

Constraints (2) impose limits on the reactive power generation Q_{Gi} at the PV generation buses; constraints (3) impose limits on the voltage magnitudes V_i with reference to PQ type generation buses and to load buses whose voltage is placed under surveillance ("sentinel buses"); constraints (4)–(6) set proper upper and lower bounds on the variables $\mathbf{q}_G, \mathbf{v}, \mathbf{r}$ respectively; constraints (7) force q_{Ci} to 0 when no capacitor installation allowed; otherwise, an upper bound $\overline{q}_{Ci}$ is set on the capacitor size.

4.1.2. Solution Algorithm

CP is a nonlinear programming problem with mixed variables. According to the SLP approach, the CP model is linearized in correspondence to a load flow solution and the resulting MILP problem is solved.

The adopted SLP algorithm consists of the following steps:

1. Initial load flow calculation; the optimal value F_{best} of the objective function is set to an arbitrarily large number.
2. Linearization of the CP problem; solution of the resulting MILP problem.
3. Update of the control variables $\boldsymbol{\chi}, \mathbf{q}_C, \mathbf{q}_G, \mathbf{v}, \mathbf{r}$.
4. Load flow calculation corresponding to the updated working point.
5. Check of the nonlinear constraints (2) and (3); in case of violation, go to Step 2; otherwise, go to Step 6.
6. Store the current working point and the corresponding value F of the objective function.
7. Check exiting conditions. Compare F with F_{best}: if a satisfactory decrease in the objective function value is found, set $F_{best} = F$ and go to Step 2; otherwise, the algorithm is over.

The MILP problem of Step 2 has the following form:

$$\min\left(\mathbf{C}_c^T \mathbf{x}_c + \mathbf{C}_d^T \mathbf{x}_d \right) \qquad \text{subject to:}$$
$$\mathbf{A}_c \mathbf{x}_c + \mathbf{A}_d \mathbf{x}_d \leq \mathbf{b}$$
$$\underline{\mathbf{x}}_c \leq \mathbf{x}_c \leq \overline{\mathbf{x}}_c$$
$$\underline{\mathbf{x}}_d \leq \mathbf{x}_d \leq \overline{\mathbf{x}}_d$$

where the control variable vector $\mathbf{x}$ is partitioned into the subsets of continuous $\mathbf{x}_c = \left[\mathbf{q}_G, \mathbf{v}, \mathbf{r} \right]^T$ and discrete $\mathbf{x}_d = \left[\boldsymbol{\chi}, \mathbf{q}_C \right]^T$ variables.

To solve problem MILP, BGA or µGA are used. The application of genetic algorithms would require the solution of MILP to evaluate the fitness of each individual. In our

implementation, however, only the discrete variable vector is directly operated on by genetic algorithm techniques. In particular, the variables are coded in binary form, with the relative upper and lower bound constraints implicitly accounted for by the "coding" mechanism.

Each individual (corresponding to a possible choice of the discrete variable values) is evaluated by solving the following linear programming problem (LPP) obtained from MILP by setting $\mathbf{x}_d = \tilde{\mathbf{x}}_d$:

$$\min \left[\mathbf{C}_c^T \mathbf{x}_c + \mathbf{C}_d^T \tilde{\mathbf{x}}_d - C_p \left(\varphi - 1 \right) \right] \quad \text{subject to:}$$

$$\mathbf{A}_c \mathbf{x}_c - \left(\mathbf{b} - \mathbf{A}_d \tilde{\mathbf{x}}_d \right) \varphi \le 0$$

$$\underline{\mathbf{x}}_c \le \mathbf{x}_c \le \overline{\mathbf{x}}_c$$

$$0 \le \varphi \le 1$$

Each iteration of the genetic algorithm requires a number of LPP's equal to the population size to be solved.

In the model above, C_p is a large positive penalization constant intended to drive the feasibility variable φ toward 1 since this value corresponds to the optimal feasible solution. When constraints are infeasible, the solution obtained (with $\varphi < 1$) is as close as possible to constraint enforcement.

The fitness of feasible individuals ($\varphi = 1$) is equal to the value of the objective function of MILP. On the other hand, infeasible individuals are provided with a direct measure of their fitness by means of the penalized objective function of LPP.

The above model allows ranking of infeasible individuals according to the objective function of LPP and succeeds in avoiding the loss of the genetic material embedded in highly promising nearly feasible individuals.

Regarding the implementation of both genetic algorithms, it was found profitable to include a "smarter" individual in the initial population. The discrete variables $\tilde{\mathbf{x}}_d$ relative to this individual are obtained by solving a linear programming problem (derived from MILP by relaxing the integer variable restrictions) and successively rounding the solution to the nearest discrete values.

Finally, satisfying constraints (7) may become troublesome for discrete type capacitor banks since both the reactive power q_{Ci} and the corresponding binary switch χ_i belong to the discrete variable vector directly handled by the genetic algorithm. For example, individuals with a nonzero q_{Ci} and the corresponding $\chi_i = 0$ may appear in the population as a result of crossover or mutation. To enforce feasibility of constraints (7) a sort of character dominance strategy was devised. The χ_i character is declared dominant with respect to q_{Ci} meaning that a nonzero q_{Ci} shows up only if the corresponding χ_i is 1; otherwise it is set to 0. In this case, however, the binary coding q_{Ci} is left unchanged to allow the recessive character to show up again in successive generations.

4.2. Topological Rescheduling

The aim of topological rescheduling is that of minimizing parallel flows which consist in the undesired circulation of power flows through certain corridors in a large interconnected network. Whatever their origin may be, parallel flows are synonymous with heavy loading of transmission lines, increased losses and sometimes the need to curtail certain transactions to account for the undesired flows along the parallel paths.

The countermeasures available to transmission system operators and planners generally include network topology reconfiguration by corrective switching of substation breakers, installation and operation of phase-shifting transformers (PST) and generation rescheduling to alleviate possible overloads.

The problem of parallel flow control can be formulated as that of minimizing the "power transfer distribution factors" (PTDFs) of certain transmission lines belonging to an interconnection corridor, with reference to a set of given power transactions. Constraints to be accounted for include N and N-1 security levels that have to be enforced after the topological layout of the system has been changed to improve the PTDFs.

The resulting large–scale mixed integer programming problem is solved by a genetic algorithm based procedure. This choice is made because of the discrete nature of most of the control variables (such as the status of substation breakers and location of PSTs) and also because the minimum PTDF objective function cannot be expressed analytically with respect to the control variables.

The original constrained problem is reformulated as a multiobjective optimization. Beside the PTDF of a line (or the sum of PTDFs associated to certain lines in an interconnection corridor), the objective functions considered are the sum of the violations to N and N-1 security constraints. The multiobjective formulation allows an easier evaluation of the impact that different topology and PST operations may have on power transfer and security levels. Moreover, it is sometimes convenient to allow a possibly small amount of constraint infeasibility to get a significant improvement in some of the other objective functions.

4.2.1. Generalities

Owing to the large amount of computation necessary to solve the problem, the approximate dc load flow model is adopted. The power flow p_{ij} through a branch (line or transformer) connecting nodes i and j, is expressed as:

$$p_{ij} = \left(\theta_i - \theta_j\right)\big/ x_{ij} \tag{8}$$

or, if a PST is used to control the power flow through line $i - j$, as:

$$p_{ij} = \left(\theta_i - \theta_j - \alpha_{ij}\right)\big/ x_{ij} \tag{9}$$

In (8) and (9), x_{ij} is the branch longitudinal reactance; θ_i and θ_j are the phase angles of the corresponding nodes; α_{ij} is the angle shift imposed by the PST.

The real power balance equations, at all nodes except the slack bus, can be written as:

$$\mathbf{B}\boldsymbol{\theta} + \mathbf{C}\boldsymbol{\alpha} = \mathbf{A} \tag{10}$$

In (10), $\mathbf{B}$ is the dc load flow Jacobian matrix, $\boldsymbol{\alpha}$ is the PST angle shift vector, $\mathbf{A}$ is the vector of bus power injections, matrix $\mathbf{C}$ takes the PST contributions into account. Each PST contributes to one column of matrix $\mathbf{C}$; for example, the column corresponding to the PST located on line $i - j$ takes the following form:

$$\mathbf{c}_{ij}^T = \begin{bmatrix} 0 \cdots 0 - 1/x_{ij} \; 0 \cdots 1/x_{ij} \cdots 0 \end{bmatrix}^T$$

with the two nonzero entries located in positions i and j.

For a given transaction, the PTDF corresponding to a certain line is evaluated by the following simple computations.

The source node h and the sink node k are identified and the corresponding transaction incidence column vector $\mathbf{e}_{hk}$ is formed ($\mathbf{e}_{hk}$ is a null vector except for a $+1$ entry in position h and a -1 entry in position k).

The change in power flow through line $i - j$, as an effect of both the traded energy and the operation of PSTs, is given by:

$$\Delta p_{ij} = \frac{1}{x_{ij}} \mathbf{e}_{ij}^T \mathbf{B}^{-1} \left(p_{tr} \mathbf{e}_{hk} - \mathbf{C}\boldsymbol{\alpha} \right)$$

where $\mathbf{e}_{ij}$ is a column vector with all zero entries except for a $+1$ in position i and a -1 in position j and p_{tr} is the transaction power.

By dividing the above equation by p_{tr}, a generalized expression of the PTDFs, retaining the effects of both transaction and PST operations, is obtained:

$$\mu_{ij,hk} = \frac{1}{x_{ij}} \mathbf{e}_{ij}^T \mathbf{B}^{-1} \left(\mathbf{e}_{hk} - \frac{\mathbf{C}\boldsymbol{\alpha}}{p_{tr}} \right) \tag{11}$$

when more PTDFs, associated to a given set of lines in a power transfer corridor, have to be evaluated, (11) can be expressed in matrix form.

The constraints to be taken into account include N and N-1 security limits that must be satisfied also after network topology and PST shift angles were modified to minimize the parallel flows.

N security constraints consist of a possibly large number of inequalities of the form:

$$-p_{hk}^{\max} \leq p_{hk}(\theta) \leq p_{hk}^{\max} \tag{12}$$

involving all the lines and transformers of the network.

N-1 security limits account for the power flows in the network branches as a consequence of a line outage occurring in the system. N-1 security constraints are much more numerous than N security ones and consist of a large set of inequalities of the form:

$$-p_{hk}^{\max} \le p_{hk,rs}\left(\theta\right) \le p_{hk}^{\max}$$

where hk is the index of a network branch (line or transformer) whose power flow is being monitored and rs is the index of the outaged network element.

The control actions considered for the minimization of parallel flows with N and N-1 security constraints are the following: disconnecting an operating line from the network; changing over a line or a transformer from one substation busbar to another; operating a PST including the possibility of placing a new PST in a given site selected from a set of candidate locations.

The above mentioned operations have the effect of changing matrix $\mathbf{B}$; PST operation or placement also changes vector $\boldsymbol{\alpha}$ and matrix $\mathbf{C}$; as a result, the PTDFs computed according to (11) are significantly modified.

Disconnecting a line can be modeled as a rank one modification of matrix $\mathbf{B}$, whereby the change in the PTDFs can be computed by means of the "inverse matrix modification lemma" (IMML). However, opening a line is a somewhat dangerous operation since it reduces the meshing degree of the system and may have a negative impact on the N and N-1 security levels.

A change over operation is represented in figure 2. A line, initially connected to substation busbar $sb1$ (figure 2a) is connected to the busbar $sb2$ by the operation of breakers $br1$, $br2$ and $br3$ (figure 2b).

The operation of disconnecting the line from $sb1$ and reconnecting it to $sb2$ can be easily handled, from the computation viewpoint, as a rank-two update of matrix $\mathbf{B}$.

The possible opening of breaker $br3$ however has the effect of creating an extra node that would increase the size of matrix $\mathbf{B}$. Instead of working with a variable size matrix $\mathbf{B}$, it is found expedient including in $\mathbf{B}$ all the extra nodes that would be created as an effect of bus splitting operations. This can be done by modeling busbar breakers like $br3$ explicitly, leaving them in the closed position until otherwise required. A closed breaker is represented as two reactances of opposite values connected in series. As a consequence, beside the rank-two update corresponding to the line change over operation, a further rank-one modification of matrix $\mathbf{B}$ is required to simulate the opening of breaker $br3$.

Finally, PST operations modify the elements of vector $\boldsymbol{\alpha}$ and, if a new device is placed, also the columns of matrix $\mathbf{C}$. Matrix $\mathbf{C}$ is built only once at the beginning of the procedure with its columns in one-to-one correspondence with all the PSTs both existing and prospective. Decommitting a PST as a consequence of a given planning choice is carried out by setting to zero the corresponding entry in vector $\boldsymbol{\alpha}$. In (9), the PST reactance is actually added to that of the line where it is installed. This means that, in case the PST in not installed, its reactance should be eliminated by a further rank-one update of matrix $\mathbf{B}$.
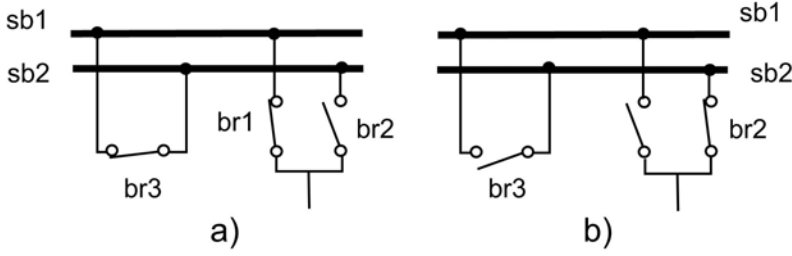
Figure 2. Example of line change over.

4.2.2. Solution Approach

PTDFs cannot be expressed analytically as a function of the status of substation breakers and of the setting of PSTs; they can be evaluated only numerically according to (11). This is the reason why the optimization of PTDFs, based on the system topology and PST operation, cannot be carried out by means of standard optimization techniques.

An approach based on BGA is used to find a solution to the problem of minimizing parallel flows with N and N-1 security constraints.

The problem of minimizing PTDFs with security constraints is reformulated as a multiobjective optimization by converting N and N-1 security levels into subsidiary objective functions. Two are the reasons behind this choice. First, there is no simple way to deal with N and N-1 security constraints by BGA except by the penalization method: constraint violations are added to the original objective function of the problem, weighting them with some suitable penalization parameters. This approach closely resembles the weighting method used to solve multiobjective optimization problems. The second reason is that, owing to the discrete nature of the control actions, the different Pareto optima are stable with respect to small changes in the weighting coefficients; as a consequence only a few points on the Pareto optimal front will need to be generated.

The GA fitness function is computed as follows:

$$F = w_1 \phi_1 + w_2 \phi_2 + w_3 \phi_3$$

where w_1, w_2 and w_3 are non negative weighting coefficients which satisfy the condition $w_1 + w_2 + w_3 = 1$.

Objective function ϕ_1 is the one related to the PTDFs of a given parallel path. Since the aim of GAs is that of maximizing the population fitness, function ϕ_1 is taken as:

$$\phi_1 = \frac{1}{\sum_{i \in I} |PTDF_i| + \varepsilon}$$

where I is the set of lines whose PTDF should be minimized and ε is a small number (say 10^{-5}) introduced to avoid overflows due to possible divide-by-zero operations.

Objective function ϕ_2 is the one related to N security constraints. A non-negative violation function v_{hk} is computed for each network branch as follows:

$$\begin{cases} v_{hk} = p_{hk}(\theta) - p_{hk}^{\max} & p_{hk}(\theta) > p_{hk}^{\max} \\ v_{hk} = 0 & \text{if} \quad -p_{hk}^{\max} \le p_{hk}(\theta) \le p_{hk}^{\max} \\ v_{hk} = -p_{hk}(\theta) - p_{hk}^{\max} & p_{hk}(\theta) < -p_{hk}^{\max} \end{cases} \tag{13}$$

Function ϕ_2 is then defined as:

$$\phi_2 = \frac{1}{\sum\limits_{hk} v_{hk} + \varepsilon}$$

Objective function ϕ_3 is the one related to N-1 security constraints. A violation function $v_{hk,rs}$ is computed similarly to (13), so ϕ_3 is defined as:

$$\phi_3 = \frac{1}{\sum\limits_{hk,rs} v_{hk,rs} + \varepsilon}$$

An interesting feature of the multiobjective formulation is that the optimization procedure can be readily adapted for different purposes by simply changing the values of the weights. For example, by taking $w_1 = w_3 = 0$ and $w_2 = 1$, the multiobjective optimization procedure turns into a corrective switching method. When small or no violations of N and N-1 security limits are present in base case conditions, a good choice for the weighting coefficients is: $w_1 = 0.8$ and $w_2 = w_3 = 0.1$; these values were used in the tests shown in Section 6.3.

Similarly to some existing multiobjective optimization programs, the procedure is provided with an interactive user interface designed to enable the decision maker to control the optimum seeking process by modifying the values of the weighting coefficients w_1, w_2 and w_3. After each BGA optimization, the decision maker is asked about the suitability of the current solution; if the solution is not acceptable, the weights can be modified to look for an improved optimum.

4.2.3. Implementation Issues

Evaluating the fitness of a member of the population involves the computation of functions ϕ_1, ϕ_2 and ϕ_3 after applying the control actions specified by its binary coding. The core of the evaluation procedure is the application of IMML to matrix $\mathbf{B}^{-1}$ to evaluate PTDFs according to (11). This allows the computation of function ϕ_1.

The computation of ϕ_2 is a simple byproduct of the evaluation of ϕ_1. Indeed, the IMML can be stated as follows:

$$\left(\mathbf{B} + \mathbf{M\delta M}^T\right)^{-1} = \mathbf{B}^{-1} - \mathbf{B}^{-1}\mathbf{M\gamma M}^T\mathbf{B}^{-1}$$

where $\boldsymbol{\delta}$ is a square modification matrix, $\mathbf{M}$ is a connection matrix and $\gamma = \left(\boldsymbol{\delta}^{-1} + \mathbf{z}\right)^{-1}$ with $\mathbf{z} = \mathbf{M}^T\mathbf{B}^{-1}\mathbf{M}$.

The N security limits are checked by first computing the phase angle vector $\boldsymbol{\theta}$ as follows:

$$\boldsymbol{\theta} = \left(\mathbf{B}^{-1} - \mathbf{B}^{-1}\mathbf{M\gamma M}^T\mathbf{B}^{-1}\right)\left(\mathbf{A} - \mathbf{C\alpha}\right) \tag{14}$$

Matrices $\mathbf{B}^{-1}\mathbf{M}$ and γ are the same used in the evaluation of PTDFs and need not to be recomputed. Once $\boldsymbol{\theta}$ is known, it is possible to check constraints (12) and compute ϕ_2.

Regarding the N-1 security constraints, the computation steps are similar to those required for the N security check. At first it is necessary to compute vector $\boldsymbol{\theta}$ as in (14). The difference with respect to N security is that the outage of one line at a time is to be considered; as a result, matrix $\mathbf{B}^{-1}\mathbf{M}$ need to be supplemented by an extra column required to account for the outage. Matrices $\mathbf{z}$ and γ need to be partially updated accordingly. When the outaged line is equipped with a PST, the corresponding value of the shift angle must be set to 0.

5. BAD DATA IDENTIFICATION IN STATIC STATE ESTIMATION

5.1. Post-Processing of the WLS Solution

The procedures for bad data identification, in the framework of the weighted least square (WLS) state estimation, rely on successive elimination of suspect bad data or on the use of non-quadratic criteria in carrying out the state estimation.

The commonly used largest normalized residual (LNR) criterion exploits the successive elimination strategy. LNR may have problems in correctly identifying multiple interacting bad data especially when they are of the conforming type. In such a case, repeated application of the LNR criterion sometimes results in the suppression of correct measurements.

Among the non-quadratic state estimators, the weighted least absolute value approach has gained widespread popularity thanks to the automatic bad data rejection property of this method. However, non-quadratic state estimators may suffer from convergence problems and are more computationally demanding with respect to least square estimators.

The problem of identifying multiple bad data can be considered as a combinatorial decision procedure. From the mathematical point of view, this leads to the formulation of a nonlinear minimization problem with mixed (integer and real) variables. The problem may be tackled by the branch-and-bound method; however, genetic algorithms (in particular, the CB-GA approach) and tabu search (TS) techniques can be profitably employed.

Each of these approaches requires the execution of three steps: generation of a tentative bad data identification, solution of the corresponding state estimation problem (by the WLS method) and the memorization (either explicit or implicit) of already considered cases. The

state estimation solution is the most computationally demanding task; to reduce the computation burden, the possibility of replacing the WLS solution with an approximate one based on sensitivity methods is investigated.

5.1.1. Problem Statement

The general state-estimation problem consists in solving an over-determined set of nonlinear equations as follows:

$$\mathbf{h}(\mathbf{x}) + \mathbf{e} = \mathbf{z} \tag{15}$$

where $\mathbf{x}$ is the n-dimensional state vector, $\mathbf{z}$ is the m-dimensional measurement vector, $\mathbf{h}$ is a vector function relating the measured quantities to the state vector and $\mathbf{e}$ is the m-dimensional vector of measurement errors.

Equation (15) can be solved by the WLS method which consists in minimizing the following objective function with respect to $\mathbf{x}$:

$$J(\mathbf{x}) = \left[\mathbf{z} - \mathbf{h}(\mathbf{x})\right]^T \mathbf{R}^{-1} \left[\mathbf{z} - \mathbf{h}(\mathbf{x})\right]$$

where $\mathbf{R}$ is the diagonal matrix of the measurement covariances.

The results obtained by a WLS estimation are corrupted if bad data are present within the measurement set; this can be checked, for example, by means of the χ^2 test. To accept a WLS solution, it is sufficient to verify that, corresponding to the state estimate $\mathbf{x}$, it results: $J(\mathbf{x}) < C$, being C the χ^2 detection threshold.

Alternatively, the maximum (absolute) value of the normalized residuals could be used as the detection criterion as in the LNR test. The identification and elimination of single or multiple bad data could be carried out by repeated application of the LNR test. This procedure works satisfactorily even in the case of multiple interacting bad data, but it may lead to a wrong identification and therefore to the elimination of good measurements when the bad data are mutually consistent (i.e. when they are of the conforming type).

In the formulation adopted for bad data identification, any combination of good and bad measurements is associated with an m-dimensional decision vector $\mathbf{b}$ in which:

$b_i = 1$ if the $i - $ th measurement is bad

$b_i = 0$ if the $i - $ th measurement is good

Hence, for an observable system with m measurements there are 2^m possible decision vectors $\mathbf{b}$ each corresponding to a set of suspect bad data (measurements with $b_i = 1$). The combination of good and bad measurements in a given vector $\mathbf{b}$ provides an explanation for the data inconsistency if, after the removal or the correction of the suspect data, no further bad data are detected.

Since different explanations for the measurement inconsistency are likely to exist, the next step is to find the set of most plausible explanations, according to some pre-specified

rule. Following the conservative hypothesis that the occurrence of bad data is rather unlikely, the solution with the least number of bad data is assumed to be the one with the highest probability.

For any given decision vector $\mathbf{b}$, $S(\mathbf{b})$ denotes the corresponding measurement set assuming that only the "good" data are taken into account and the suspect bad data have been eliminated. The $S(\mathbf{b})$ measurement set must satisfy the requirement of observability for the system. After re-estimation according to $S(\mathbf{b})$, the possible presence of bad data is to be checked either by the χ^2 or by the LNR test.

If the χ^2 test is used, the bad data identification can be formulated as the following combinatorial problem:

$$\min F(\mathbf{b}) \qquad \text{subject to:}$$
$$S(\mathbf{b}) \text{ is observable} \tag{16}$$
$$J\big[\mathbf{x}(\mathbf{b})\big] < C(\mathbf{b})$$

where $\mathbf{x}(\mathbf{b})$ is the new state vector, $J\big[\mathbf{x}(\mathbf{b})\big]$ is the corresponding value of the objective function, and $C(\mathbf{b})$ is the updated χ^2 detection threshold. The objective function is equal to the total number of suspect bad data: $F(\mathbf{b}) = \sum_{i=1}^{m} b_i$.

5.1.2. Solution by CB-GA

The formulation of the bad data identification problem is well suited to the characteristics of genetic algorithms, since the unknown vector $\mathbf{b}$ is already coded in binary form.

However, it must be noted that the constraints of problem (16) cannot be easily handled by BGA other than by the penalization method.

The GA variant proposed by Chu and Beasley manages to find the constrained optimum without employing penalization. Two objective functions are associated with each individual. The former, coincident with $F(\mathbf{b})$, is properly the fitness; the latter, named the "unfitness", is a measure of each individual's infeasibility.

With reference to problem (16), unfitness is defined as:

$$U(\mathbf{b}) = J\big[\mathbf{x}(\mathbf{b})\big] + K(\mathbf{b})$$

where parameter K introduces a large additional term whenever the measurement layout, corresponding to the decision vector $\mathbf{b}$, makes the system unobservable.

It is convenient to add a branch-and-bound flavor to CB-GA by employing a method for systematically filtering out large portions of the search space. Whenever a feasible solution of problem (16) is found with a number of bad data equal to N_b, all solution cases which present more than N_b suspect measurements are excluded from the set of possible solutions

without performing the state estimation computation; in fact, all solutions which eliminate more than N_b bad data would be worse than the current best solution.

5.1.3. Solution by Tabu Search

Tabu search is a memory driven search procedure that relies on intelligent use of both short and long term memory to find its way toward the optimum.

TS works with a single tentative solution, represented by the decision vector $\mathbf{b}$, which is modified during the search by effect of "add" or "drop" moves.

The add move consists in finding a measurement to be declared bad, for example according to the LNR test; the corresponding element of vector $\mathbf{b}$ is set to 1.

The drop move consists in finding a measurement that was previously declared bad, and switch the corresponding element of vector $\mathbf{b}$ back to 0. A simple dropping rule is that of selecting the measurement (among the bad ones) which presents the minimum absolute value of the residual.

After a move is made, the corresponding measurement is declared "tabu", that is, if the measurement was added, it cannot be dropped for a fixed number of iterations (called "tabu tenure") and vice versa if the measurement was dropped. The enforcement of tabu prescriptions is performed thanks to the short term "recency" based memory which keeps a record of the latest moves made by TS.

Starting from the initial condition with $\mathbf{b} = [0,\ldots,0]^T$, repeated application of the add move eventually leads to a bad data free solution (the same obtained by the repeated application of the LNR elimination criterion). All bad data free solutions are stored in the long term "quality" based memory.

After a solution with no bad data was obtained, TS enters the intensification phase during which the neighborhood of the solution is examined more thoroughly to look for possibly better solutions located in the vicinity of the one just obtained. Intensification usually entails a sequence of drop and add moves.

When the neighborhood of a solution has been explored unsuccessfully, it becomes convenient to examine a different region of the solution space by applying diversification. The ordinary recency based transition rules are temporarily modified to force the TS algorithm to leave the vicinity of the last solution found and to resume search in some distant region of the solution space. To this purpose, it is convenient to exploit the long term "frequency" based memory which keeps a record of the number of times a given measurement was declared bad.

The TS steps are outlined by the following algorithm:

1. Initialization
    a. set the initial decision vector $\mathbf{b} = [0,\ldots,0]^T$;
    b. define the maximum number of iterations M_{TS} and set iteration counter k_{TS} to 0;
    c. define the tabu tenure value T .

2. Solution

   a. increase iteration counter: $k_{TS} \leftarrow k_{TS} + 1$; if $k_{TS} > M_{TS}$, the TS algorithm is over;
   b. solve the state estimation problem corresponding to the current decision vector **b** and test the solution for bad data;
   c. if no bad data are detected, memorize the solution in the long term quality based memory and go to Step 3;
   d. if bad data are detected:
   - let i be the measurement with the largest normalized residual;
   - set $b_i = 1$ and declare measurement i tabu for the next T iterations and go to Step 2.

3. Intensification

   Examine the neighborhood of the last solution found:

   - look for a measurement candidate for a drop move among those declared bad, according to the minimum absolute value of the residual and to the tabu prescriptions; if no candidate is available, go to Step 4;
   - being j the measurement candidate to dropping, set $b_j = 0$ and declare j tabu for the next T iterations;
   - go to Step 2.

4. Diversification

   Perform search in a different region of the solution space:

   - look at the frequency based memory for those measurements that were most often declared bad;
   - make some of these measurements un-eligible for add moves (for a number of iterations $\gg T$) and zeroise the corresponding elements of vector **b**;
   - go to Sep 2.

In the diversification step, the measurements to be declared un-eligible are chosen randomly with a pre-assigned probability φ. By taking $\varphi = 1$, the TS algorithm can be forced to assume that all measurements often considered bad are actually good, at least for a number of iterations sufficient to depart from already explored solutions.

The progress of the TS algorithm in the search for a bad data free solution can be represented by the binary tree shown in figure 3 where each node corresponds to a state estimation problem. If the current solution contains bad data, the decision relative to the next move is driven by the LNR test. The measurement with the largest normalized residual (say the i-th) is taken as the candidate for the next solution operation (Step 2.d); it is declared bad for testing purposes ($b_i = 1$) and the corresponding tentative solution is obtained by solving the ensuing state estimation problem.
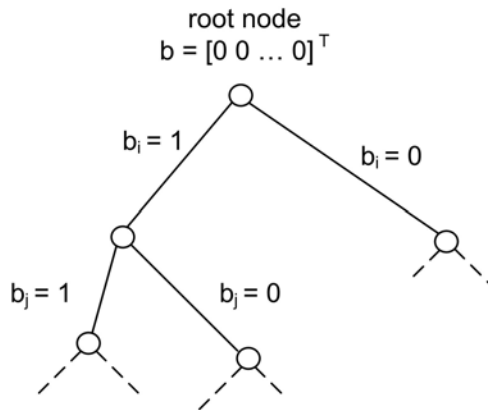
Figure 3. Binary tree of solution nodes in Tabu search

5.1.4. Search Depth and Search Breadth

The concepts of depth and breadth are widely used in relation to search problems. With reference to the tree shown in figure 3, the depth of a node can be defined as its distance from the root in the vertical direction. The number of nodes having the same depth defines the breadth of the solution tree at a given stage of the search.

The choice to privilege depth or breadth in performing the search depends on the characteristics of the problem at hand. For bad data identification, depth first search appears more appealing than breadth first, since the maximum number of expected bad data is not usually large. The TS algorithm described above belongs to the depth first search category.

In particular, when a solution free from bad data is found (a case that happens quite often), the limit value of expected bad data M is updated and all non-promising alternative solutions (i.e. those with $F(\mathbf{b}) < M$) can be discarded. The number of possible solutions that would have to be examined remains very large even after pruning. Indeed, the solutions with a number of bad data less than or equal to M can be as many as:

$$S = \sum_{i=2}^{M} \frac{m!}{(m-i)!\,i!} \tag{17}$$

Each term of the summation is the number of different ways i bad data can be picked from m measurements.

The contributions of terms $i = 0$ (no bad data) and $i = 1$ (a single bad data) are ignored; the former case is trivial while, in the latter, the LNR criterion would suffice to identify the bad data.

A measure of the quality of the search can be given by the percent ratio δ between the number ν of valid test cases (i.e. those with $F(\mathbf{b}) > M$) evaluated by the procedure and the number of possible alternative solutions S : $\delta = (\nu/S) \cdot 100$. Parameter δ will be used to define, in a practical and measurable way, the breadth of the search achieved by the particular procedure employed.

5.1.5. Implementation Issues

In the present Section some characteristics of the proposed search approaches will be considered with reference to the three basic steps of tentative bad data identification, state estimation solution and memorization of results.

A. Tentative Bad Data Identification

TS exploits the LNR test as a local optimization criterion to select the next move; for the subsequent iterations, TS employs intensification and diversification strategies.

The creation of new solution cases according to CB-GA is completely non-deterministic. The population of tentative solutions is obtained by a recombination of the most promising solutions under the pressure of a selection mechanism imitating that occurring in nature.

B. State Estimation Solution

The computation of the state estimation solution corresponding to each tentative bad data identification is the most computationally demanding phase of any of the proposed procedures. In addition to the classical solution of the WLS problem by the Gauss-Newton iterative scheme, a sensitivity based approach was implemented to improve computational efficiency.

The linearization of the residual equation $\mathbf{r} = \mathbf{z} - \mathbf{h}(\mathbf{x})$ leads to the following sensitivity relation:

$$\Delta\mathbf{r} = \left(\mathbf{I} - \mathbf{H}\mathbf{G}^{-1}\mathbf{H}^T\mathbf{R}^{-1}\right)\Delta\mathbf{z}$$

where $\mathbf{I}$ is the $(m \times m)$ identity matrix, $\mathbf{H}$ is the $(m \times n)$ measurement Jacobian and $\mathbf{G} = \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ is the WLS gain matrix.

Instead of deleting suspect bad data from the measurement set, it is convenient to correct the entries of vector $\mathbf{z}$ relative to the bad data so as to zeroise the corresponding residuals. The correction vector $\mathbf{c}_b$ to be added to the proper $\mathbf{z}$ elements are obtained as:

$$\mathbf{c}_b = -\mathbf{W}_{bb}^{-1}\mathbf{r}_b \tag{18}$$

where $\mathbf{W}_{bb}$ is the submatrix of $\mathbf{W} = \mathbf{I} - \mathbf{H}\mathbf{G}^{-1}\mathbf{H}^T\mathbf{R}^{-1}$ relative to the bad data to be corrected and $\mathbf{r}_b$ is the corresponding vector of residuals computed at the end of the base case WLS solution.

In carrying out the computation, $\mathbf{W}$ is replaced by the weighted sensitivity matrix $\tilde{\mathbf{W}} = \mathbf{R}^{-1/2}\mathbf{W}\mathbf{R}^{1/2}$ which is cheaper to evaluate thanks to its symmetry; linear equation (18) still holds as long as $\mathbf{c}_b$, $\mathbf{W}_{bb}$ and $\mathbf{r}_b$ are replaced by the corresponding weighted quantities $\tilde{\mathbf{c}}_b$, $\tilde{\mathbf{W}}_{bb}$ and $\tilde{\mathbf{r}}_b$. The WLS objective function J can be easily updated to account for the correction of the bad data without carrying out a new WLS computation.

C. Memorization of Results

Memorization of results is intended to avoid that the same solution case is considered more than once. Two types of memorization (explicit or implicit) may be considered.

Explicit memorization implies the use of a sort of dictionary and completely prevents re-computation of already solved cases.

The main drawback of this approach consists in the dimension of the dictionary; even for medium-sized systems it is not practical to handle a dictionary with 2^m items (all possible determinations of vector **b**).

The explicit memory strategy is used for CB-GA. The dimensionality problem is tackled by reducing the search space suitably. After carrying out the first state estimation with the original data set, the vector of normalized residuals is evaluated and re-arranged in decreasing order of absolute values. The bad data are searched within the subset of measurements with the p largest normalized residuals: the dimension of the dictionary is therefore reduced to 2^p with $p < m$.

TS makes use of an implicit memory approach, which aims at avoiding recalculation without implementing a dictionary. A proper choice of the tabu tenure and a careful implementation of the intensification and diversification procedures can make unwanted evaluation of already considered test cases very improbable although not impossible.

5.2. The Maximum Agreement Algorithm

As shown in Section 5.1, one of the drawbacks of WLS state estimation is that it lacks the necessary robustness in case numerous bad data are present within the set of available measurement. Instead of post-processing of the WLS solution, it would be highly desirable to employ a robust state estimation algorithm capable of automatically rejecting the bad data.

The search for robust estimators led to the proposal of non-quadratic estimators such as WLAV and to the least median of the squares (LMS) method, which is shown to possess the best characteristics of bad data rejection with respect to other approaches. The LMS method is also immune to effects of leverage points which plague state estimators of the WLS or WLAV type. However, LMS requires a sort of combinatorial optimization which originates form statistical considerations on the breakdown point of linear regression models [35].

In this Section, a robust state-estimation procedure is presented. It is based on the idea of selecting subsets of the measurements, called "essential sets", with size equal to the number of states. Corresponding to each subset, a tentative solution is evaluated by solving the nonlinear model equations of the measurements belonging to the essential set; the remaining redundant measurements are employed to evaluate the quality of the solution. With a terminology borrowed from politics, a solution is considered acceptable if it satisfies the majority of the remaining measurements which act as voters in an election. Among acceptable solutions, the best one is that supported by the largest majority of the electorate. The number of different essential sets to be evaluated can be determined according to the expected number of bad data. BGA is employed instead of the random selection approach of [35], to favor rapid appearance of the optimal solution.

5.2.1. General Considerations

The WLS method is the most popular approach to the solution of the over-determined set of nonlinear equations (15). Other methods, such as WLAV or LMS, exploit the idea of exactly solving a subset of n (essential) measurements, while the remaining $m-n$ are not directly involved in the estimation process. If bad data are present among the measurements, the WLAV or LMS methods have better chances of obtaining a correct state vector as long as no bad data are included in the subset of the essential measurements. Our approach belongs to this category.

Let $\bar{E}$ be an essential measurement set, that is a subset of n measurements which fulfils the observability requirement for the whole network. The number of these sets is, at most, equal to the number of combinations $\dfrac{m!}{(m-n)!n!}$, but it is likely to be much less, because many measurement combinations do not observe the network completely. Corresponding to each $\bar{E}$, a state vector $\bar{\mathbf{x}}$ can be determined by solving the following set of nonlinear equations by the Newton method:

$$h_j(\bar{\mathbf{x}}) = z_j \qquad \forall\, j \in \bar{E} \tag{19}$$

where, for explanatory purposes, it will be assumed that all measurements are error free ($e_i = 0$) except for the bad data.

Each of the remaining $m-n$ measurements presents a residual defined as:

$$r_k = z_k - h_k(\bar{\mathbf{x}}) \qquad \forall\, k \notin \bar{E}$$

A measurement $k \notin \bar{E}$ is said to agree with the given essential set if $r_k = 0$, otherwise measurement k is said to disagree with $\bar{E}$. The percentage of agreement corresponding to $\bar{E}$ can be defined as $a = 100\, n_a/(m-n)$ where n_a is the number of agreeing measurements.

The state-estimation procedure proposed aims at finding the essential set which presents the maximum agreement. Formally the "maximum agreement algorithm" (MAA) is based on the solution of the following maximization problem:

$$\max_{\Omega} f \tag{20}$$

where f is the function that relates each essential set $\bar{E}$ with the corresponding agreement percentage, that is $f(\bar{E}) = a$, while $\Omega = \{E_1, E_2, \dots E_T\}$ is the set of all essential sets.

The solution of the state estimation is the vector $\mathbf{x}$ that exactly satisfies (19) for the essential set with the maximum agreement.

Since the image of function is bounded between 0 and 100, a maximum of f necessarily exists, which means that a solution of the maximum agreement state estimation can always be found. However, in order to be accepted, it is necessary that the solution is supported by the

majority of the $m - n$ redundant measurements, that is its percentage of agreement must be over 50%.

From the above consideration, it could be argued that for a solution to be accepted, the number of disagreeing measurements should be less than $[(m - n)/2]$, where the square brackets denote the integer part of the number. However, as shown in [36], a number of bad data much smaller than that is sufficient to make any state-estimation procedure break down according to how bad data are localized in the system.

5.2.2. Relations with other Methods

The maximum agreement state-estimation procedure presents significant analogies with the LMS approach. To compare the two methods, the 5-bus test system of [36] is considered.

Using the dc load-flow approximation, assuming all reactances equal to 1 p.u. and bus 5 as the reference node ($\theta_5 = 0$), the measurement Jacobian is:

$$\mathbf{H}^T = \begin{bmatrix} 1 & -1 & 0 & 0 & 1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & -1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & 2 \end{bmatrix}$$

The following definitions introduced in [36] will be used. The "fundamental set" S_j associated to a state variable x_j is the set of all measurements for which $H_{ij} \neq 0$ ($\partial h_i / \partial x_j \neq 0$ in the nonlinear case).

The "surplus" of a fundamental set is the smallest number of measurements whose removal from a fundamental set turns at least one of the remaining measurements into a critical one. With reference to the system in figure 4, the fundamental set S_1 associated to the state variable θ_1 consists of the measurements 1, 2, 5, 6, and 7 and the corresponding surplus s_1 is equal to 4 because the elimination of any 4 of these measurements leaves the remaining one critical. The surplus of S_2 is $s_2 = 2$ because deletion of measurements 7 and 8 makes measurements 9 and 10 critical. Note that it may happen that elimination of measurements in a fundamental set makes some measurements (belonging to other fundamental sets) critical, as is the case here for S_2 (see [36]). By similar reasoning it can be seen that the surplus of the fundamental sets S_3 and S_4 is equal to 2.

As an example, it is assumed here that measurements 1–6 are exact and that two of them were selected to enter a tentative essential set, thus allowing state variables θ_1 and θ_2 only to be observed. It is also assumed that bad data are present in the set $S_3 \cup S_4 = \{7, 8, 9, 10\}$; two of these measurements have to be chosen to complete the essential set.

The possible outcomes which can be expected by applying MAA are shown in table 1 and discussed in the sequel. It must be noted that the choice to be made has no effect on the

agreement status of measurements 1–6 because of the structure of **H** ; contrary votes will necessarily belong to set $\{7, 8, 9, 10\}$ only.

If only one bad data were present (first row in table 1), the two measurements needed to complete the essential set may be picked in two different ways: either one bad and one good measurement or two good data. In the first case the two contrary votes of the remaining two
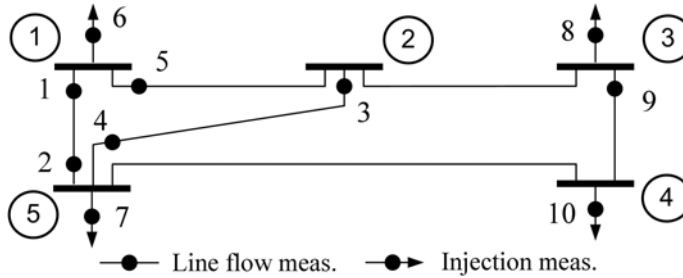
Figure 4. 5-bus test system.

Table 1. Expected MAA results for the 5-bus test system

No. of bad data	No. of good data	Possible combinations	Contrary votes
1	3	1 bad+1 good	2
		2 good	1
2	2	2 bad	2
		1 bad+1 good	1 or 2
		2 good	2

good measurements are to be expected. In the latter case only one contrary vote from the bad datum will be obtained. Therefore MAA can discriminate between the right solution and the corrupted one in this case.

If, on the contrary, two bad data were present, there are three possible ways to select two measurements out of the set $\{7, 8, 9, 10\}$. If either the two bad or the two good data are selected the same number of contrary votes, namely 2, will be found. If one good measurement is selected together with a bad data, the number of expected contrary votes will be equal to one if the bad data are conforming or equal to two if they are not. In any case MAA is not able to find the correct solution and therefore it fails. Of course it is impossible to find the correct solution if three or four bad data are present in the set $\{7, 8, 9, 10\}$. More generally, if s_i is the surplus associated with the fundamental set S_i, it will be possible to identify no more than $[s_i/2]$ bad data by application of MAA. If $s^* = \min s_i$ is, by definition, the surplus of the whole system, it can be concluded that, in the worst case, no more than $[s^*/2]$ bad data can be tolerated by MAA without breaking down.

It is not surprising that this result is the same found in [36]; indeed both the MAA and the LMS method are based on the requirement that the estimated state should satisfy the majority of the measurements. Unlike LMS, MAA does not require that the system surplus s^* is

determined in advance; the objective function in (18) can be readily evaluated once the state-estimation solution, relative to any essential set, is found and the corresponding agreement percentage is computed.

5.2.3. MAA Solution

Problem (20) is of a combinatorial nature and would require a prohibitive computation time if all the essential sets in were to be examined. Following [35], it is sufficient to carry out a limited number of state estimation computations to obtain a reasonable probability that at least one of the sets considered is free from bad data. If ε is the fraction of expected bad data, the probability that at least one out of z essential sets, chosen at random, is free from bad data is:

$$P = 1 - (1 - (1 - \varepsilon)^{n})^{z} \tag{21}$$

Equation (21) allows determining the number of different sets that have to be tested to achieve a preselected probability P (say 95%) that at least one of them in not contaminated.

The solution of (20) could be found by a Monte Carlo-like technique, that is by randomly selecting the z essential sets for which the state estimation is computed and the objective function is evaluated. This approach was used in [35].

However, GA-based search methods are generally preferable because they guarantee faster convergence toward the optimum with respect to random search techniques.

A procedure based on BGA was adopted for the implementation of MAA.

5.2.4. Implementation Issues

Because of the random nature of the crossover and mutation operators, the new solution strings are not guaranteed to possess the observability property even when parent solutions are observable. To avoid the presence of large fractions of unobservable solutions within the current population, a specialized mutation operator was devised.

Each new solution is tested for numerical observability according to the method described in [37]. The Jacobian rows corresponding to the measurements in the essential set are eliminated according to Crout's method. Unobservability is detected when one of these rows becomes null during elimination, thus showing that the corresponding row is linearly dependent from previously eliminated rows. In this case, the measurement associated with the vanishing row is replaced by another measurement from the same fundamental set, provided that it passes the observability check.

Whenever the essential set is modified to enforce observability, the corresponding solution string is updated accordingly; the entering measurement index is over-written on the string location corresponding by the replaced measurement. This operation resembles the mutation process, but there is nothing casual in it, since it is suitably designed to restore the observabilty of an essential set.

Most of the computation effort required by MAA is devoted to the computation of state estimation solutions corresponding to the essential sets produced by the BGA. Since new individuals are generated according to nondeterministic rules, it is possible that the same individual appears in successive generations. Unwanted evaluation of already considered cases is avoided by saving state estimation results in a suitable database of solved cases from

which it is possible to retrieve the value of the objective function. Strings corresponding to already evaluated individuals are stored together with the values of the objective function. A suitable ordering vector allows efficient retrieval of stored information by binary search.

6. VALIDATION OF THE COMPUTATION PROCEDURES

6.1. Test Networks

The GA based procedures were tested on some small sample systems including the IEEE systems with 14, 30 and 118 buses and two CIGRE networks with 33 or 41 buses.

The CIGRE 33 bus system (CIGRE1) consist of 23 lines, 18 two-winding transformers and 1 three-winding transformer with a total load of 1980 MW. The voltage levels are 225, 150 and 60 kV for the transmission and sub-transmission systems; generators are connected to the HV grid by means of 15/225 kV step-up transformers. The CIGRE 41-bus system (CIGRE2) was obtained by explicit modeling of some medium voltage buses.

More significant tests were carried out with reference to actual networks of Italian and European origin; these include the regional system of the Sicily island with about 200 buses, two representation of the Italian network with about 400 (ITALY1) and 500 buses (ITALY2), and the European (UCTE) EHV system with 4531 buses, 5898 lines, 1413 transformers and 467 generators with a total real load of about 265 GW.

The procedures were implemented in the MATLAB programming language, with the exception of reactive power compensation which was programmed in Fortran. Tests were carried out on a 3.2 GHz Pentium based personal computer.

6.2. Reactive Power Compensation Results

The procedures considered in Section 4.1 were tried on three electrical systems: CIGRE2, the Sicilian regional network and the Italy2.

The main characteristics of the capacitor placement problem are summarized in table 2. The computational complexity implied by the discrete modelization of capacitor banks is related to the number of steps corresponding to the discrete values of the installed reactive power. The total number of available discrete capacitor sizes is also reported in the table.

With reference to the smaller test cases of the CIGRE and Sicily networks, four iterations of the sequential linear programming method are needed to reach the solution. Table 3 shows the overall number of simplex iterations (in millions) needed to solve the whole problem using BGA and μGA based procedures. The larger figure refers to the number of simplex iterations required by the sequential linear programming procedure. The smaller value (between brackets) considers only the number of simplex iterations actually required to reach the actual optimal solution.

It was found necessary to run BGA and μGA for 2000 generations to guarantee a sufficiently reliable exploration of the solution space.

CPU times relative to BGA are greater than those provided by μGA with a slow-down ratio equal to the ratio of overall simplex iterations.

Table 2. Capacitor placement problem characteristics

Number of	Network		
	CIGRE2	Sicily	ITALY2
Problem variables	55	108	147
Functional constraints	84	164	315
Generators (PQ and PV)	10	14	77
OLTC transformers	5	10	0
Candidate sites	20	42	35
Continuous capacitors	16	11	0
Discrete capacitors	4	31	35
Discrete capacitor sizes	21	400	475

Table 3. Number of simplex iterations required by the GAs

	Network	
	CIGRE2	Sicily
BGA (population size: 30)	21.86 (4.86)	17.99 (4.27)
µGA (population size: 5)	2.98 (1.01)	3.73 (0.33)

Figure 5 shows the behavior of the MILP solution procedure obtained by µGA for the CIGRE2 network. Reference is made to the third iteration of the sequential linear programming procedure which results in the optimal solution of the capacitor placement problem. Indeed, the objective function is not reduced further by the subsequent fourth iteration. The objective function value corresponding to the fittest individual of the population is plotted versus the number of iterations. The actual MILP optimum occurs after 1400 iterations.

Regarding the Sicilian system, the optimal solution (with an objective function value of 255.09) is obtained at the third iteration of the sequential linear programming procedure. The corresponding MILP optimum is found in the course of the 9th iteration.

In the test case regarding ITALY2, all capacitor banks are modelled as discrete variables with a number of available sizes ranging from 5–25 for each bank; this leads to a large number of discrete solution possibilities to be examined.

The genetic algorithm based approach required carrying out at least 3000 iterations to solve each MILP problem, leading to very long computation times.

For the sake of a comparison, the same sequential linear programming procedure was
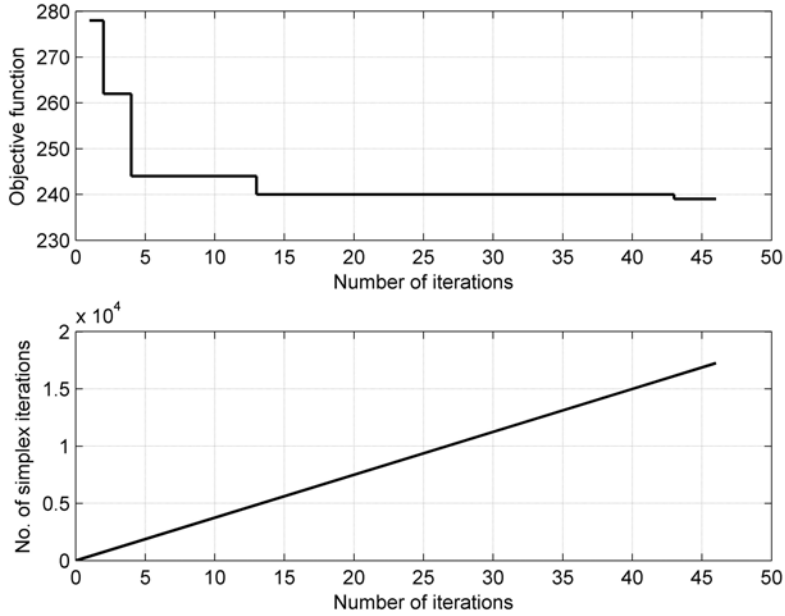
Figure 5. CIGRE2 network: MILP solution by µGA.

carried out by replacing the GAs with a linear branch-and-bound algorithm [38] in the solution of each MILP sub-problem. The branch-and-bound procedure obtains the same optimal solution of the GAs methods with reference to the CIGRE2 and Sicily systems. In the case of ITALY2, however, it was impossible to exhaust the whole solution tree by the branch-and-bound procedure because of numerical problems encountered in computation. The branch-and-bound algorithm performs a periodic check on the accuracy of the inverse of the basis matrix. If the accuracy check is not passed, the basis matrix is re-inverted as many times as needed to regain the desired accuracy. In this case, it was impossible to maintain the desired accuracy of the inverse basis matrix; suppression of the accuracy check only led to unrecoverable numerical efforts at a later stage of computation. Therefore, only a sub-optimal solution to the capacitor placement problem was obtained by taking the best discrete solution achieved before the occurrence of the numerical problems in the solution of each MILP. A value of 3875 was obtained for the objective function after three sequential linear programming iterations requiring a total of 3412 simplex iterations.

To overcome the drawbacks of both GA and branch-and-bound algorithms, a hybrid procedure was designed to exploit the computational efficiency of branch-and-bound and the robustness of genetic algorithms. For each MILP, the procedure consists in carrying out a limited (500 generations) GA search including the incomplete branch-and-bound solution in the initial population. The hybrid procedure requires three sequential linear programming iterations and achieves a substantial saving in installation cost (3268 instead of 3875), deriving from the reduced amount of capacitor power installed (565 instead of 670 Mvar).

Figure 6 shows the different installation patterns given by the incomplete branch-and-bound and by the hybrid procedure. The figure reports only those candidate locations involved in the capacitor bank placement.
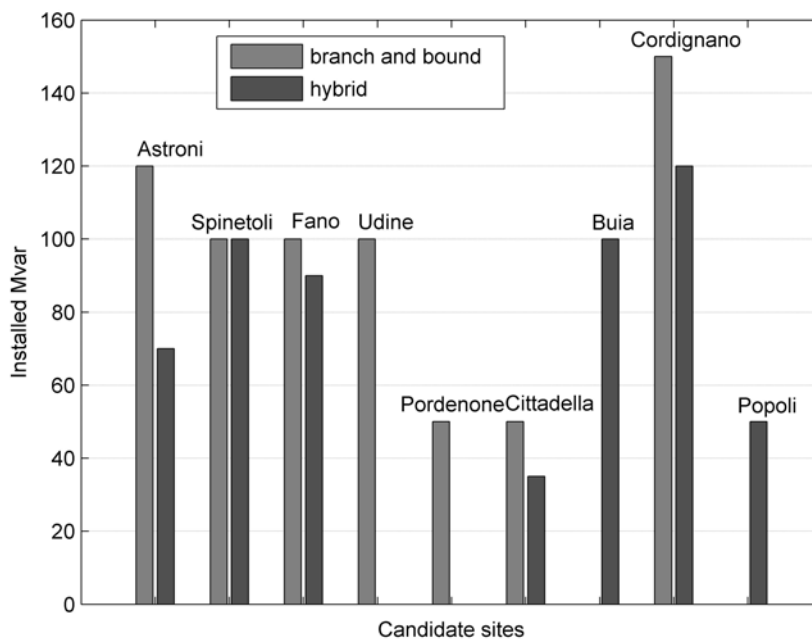
Figure 6. System ITALY1: installed capacitor patterns.

It is interesting to note that the GA search performed during the first MILP only slightly reduces the objective function compared to the best branch and bound solution found (3850 instead of 3875). Starting from this point, a much greater saving in the objective function is obtained after solving the second MILP computation.

The success of the hybrid procedure mainly depends on the robustness of the genetic algorithm approach. Genetic algorithms can tolerate that a few individuals in the population give rise to numerical problems in the evaluation phase. These individuals are given a low fitness value so that the relevant genetic material is not lost and contributes to some extent in building the genes of successive generations. In contrast, a deterministic search procedure may be subject to failure whenever even a single step of the computation chain cannot be carried out up to normal completion because of numerical problems.

6.3. Topological Rescheduling Results

Tests were carried out on the CIGRE1 network and on the UCTE system. With reference to CIGRE1 network, real power generations, loads and branch flows in MW (computed according to the dc load flow) are shown in figure 7.

The following choice of the BGA parameters was adopted: population size: 20, cross-over probability: 1, mutation probability: 0.1. The limited population size is intended to

reduce computation times; the cross-over and mutation probability values were taken higher than is usually recommended to favor the spread of promising genetic material among candidate solutions.

Regarding N-1 security analysis, a complete scan of all line outages is carried out. Outages splitting the network into separate parts are identified by a numerical test and are not simulated.

Tests were carried out with reference to a transaction between a source node located at bus 8 and a sink node located at bus 19. The corridor composed by the three lines 1–3, 4–3 and 6–5 is monitored; the PTDF to be minimized is the one relative to the first of these lines. It must be noted that there are violations of the N and N-1 security constraints in the base case solution: the genetic procedure must cope with these violated constraints beside looking for the minimum PTDF.

The following situations were considered:

1) Line opening and change over operations allowed at bus 1 only; installing and operation of PSTs are not allowed.
2) Only PSTs are employed.
3) Topology reconfiguration and PST installing both allowed.

Figure 7. Layout of the CIGRE1 network.

The topology reconfiguration obtained in the case 1 is shown in figure 8. It consists in opening line 1–2 and splitting the original bus 1; the source node 8 is connected to the newly created bus 1' together with part of the generation available via the step-up transformer.

Table 4 shows the value of the PTDF corresponding to the monitored three line corridor before and after the topological reconfiguration. If line opening is prohibited, the procedure ends by giving the same change over solution and the PTDF values in table 4 but there remains a 26 MW violation of the N-1 security limit on line 1–4 as an effect of the outage of line 2–4.

The results obtained are quite typical of the behavior of the genetic algorithm procedure: it tries to rearrange the system topology in such a way that the line whose PTDF is to be minimized is no longer on the path joining the source and the sink node.
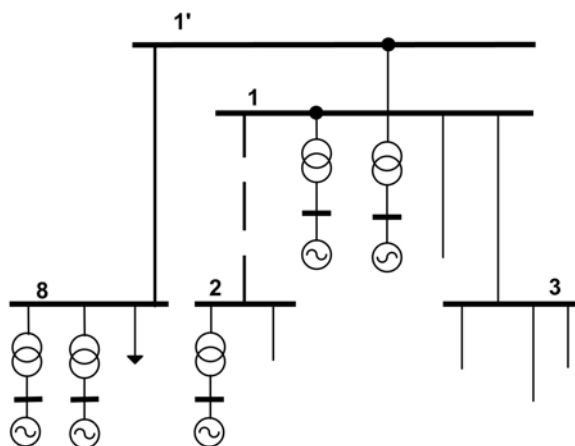
Figure 8. Topological reconfiguration of network CIGRE1.

**Table 4. CIGRE1 network: PTDF values before
and after topological reconfiguration**

Line	Base case	After reconfiguration
1–3	0.5087	0
4–3	0.1506	0.4249
6–5	0.3407	0.5751

The solution obtained is not completely satisfactory because the reliability of the system appears to be reduced by effect of the network reconfiguration; indeed not only one line was put out of service, but also the newly created bus 1' remains connected to the rest of the network by a single line.

A better result is obtained in case 2. By installing a single PST at the beginning of the line 4–3 it is possible to reduce the PTDF on line 1–3 from 0.51 to 0.09. N security constraints are satisfied and only a small 1.2 MW violation occurs on line 2–4 after the outage of line 1–4.

A further reduction of the PTF (up to about 0.01) is possible by installing another PST on line 1–3, but N-1 security cannot be satisfied anyway.

The results obtained in case 3 show that the procedure finds it more convenient to operate on breaker status than on PSTs because it can reduce the PTDF to zero even though this always results in creating antenna-like sub-networks.

To account for the risk of obtaining solutions with a weak topology, the objective function ϕ_1 defined in Section 4.2.2 can be modified as follows:

$$\phi_1 = \frac{1}{\sum_{i \in I} |PTDF_i| + \sum \psi_i + \varepsilon}$$

where ψ_i is a term penalizing the occurrence of antenna lines and is computed as follows:

$$\psi_i(d_i) = 0 \quad \text{if} \quad d_i \geq 2 \quad \text{or} \quad \psi_i(d_i) = M \quad \text{if} \quad d_i = 1$$

being d_i the number of lines connected to bus i and M any positive constant. The summation $\sum \psi_i$ is extended to those buses which are actually involved in the topology rescheduling operations; a nonzero contribution occurs only when bus i remains connected to the network by a single line.

More significant tests were carried out with reference to the large-scale UCTE system. A transaction relative to a nuclear power generation in Northern France with the corresponding consumption in Northern Italy is simulated. It can be considered as a typical transaction in the light of the parallel flow problems involving the interconnection between France, Switzerland, Austria and Italy. The corridors to be monitored, beside France–Italy, are the following: France–Switzerland, Switzerland–Italy, Austria–Italy, Slovenia–Italy, France–Belgium and France–Germany.

In base case condition, only about 40% of the contracted power crosses the French–Italian border while consistent parallel flows involve Switzerland and, to a lesser extent, Belgium, Germany and Austria. In the simulations, the PTDF relative to the France–Switzerland interconnection was minimized while satisfying the N security limits and the N-1 security constraints relative to a small subset of pre-selected outages corresponding to lines located in Northern Italy.

The control actions available include line and transformer change over operations at seven substations, located in France, Switzerland and Northern Italy along the expected path of the transaction power flow. The operation of three PSTs located on the France to Italy interconnection lines and of another one between Italy and Slovenia is also considered. No line opening is allowed.

Three different choices (labeled C1, C2, C3) were considered regarding the set of available control actions. In case C1, only substation breakers are employed to carry out all the profitable change over operations. In case C2, PTDF minimization is carried out by operating on PSTs only. Both change over and PST operations are allowed according to the third case C3. Line opening is never allowed. Results are collected in table 5 with reference to the most important interconnection corridors.

Best results were achieved in case C3, i.e. by exploiting the synergy between topology reconfiguration and PST installation and operation: the PTDF across the France–Switzerland

interconnection, the one being minimized, can be decreased from 0.2471 to 0.0106. Correspondingly, the PTDF relative to the France–Italy interconnection increases from 0.3895 to 0.8075.

It must be pointed out that topology reconfiguration alone (case C1) does have a non-negligible impact anyway with a PTDF reduction from 0.2471 to 0.1637 on the France–Switzerland interconnection. In case C2 a consistent reduction of the PTDF is also achieved; the solution obtained is not as good as in case C3 since the PTDF corresponding to the French–Italian interconnection is only slightly increased.

The results obtained were checked by means of an ac load flow calculation. An additional injection of 100 MW was considered at the source node and the same power was withdrawn at the sink node. The power flow through the contract corridor (France–Italy) and through the most important parallel paths (France–Switzerland and Switzerland–Italy) are shown in table

Table 5. UCTE System: PTDF values for different optimization choices

Interconnection	Base case	C1	C2	C3
France–Switzerland	0.2471	0.1637	0.0691	0.0106
France–Italy	0.3895	0.5031	0.3964	0.8075
Switzerland–Italy	0.4869	0.3866	0.4623	0.1354
Austria–Italy	0.0369	0.0332	0.0417	0.0166
Slovenia–Italy	0.0868	0.0771	0.0995	0.0405
France–Belgium	0.1920	0.1808	0.2723	0.1273
France–Germany	0.1714	0.1525	0.2621	0.0546

5 with reference to the base case conditions and to the best solution obtained C3.

Table 6. UCTE System: AC load flow calculation results (MW)

Interconnection	Base case		C3	
	Initial LF	LF with transaction	Initial LF	LF with transaction
France–Switzerland	613.3	638.5	551.7	547.2
France–Italy	2342.9	2381.4	2329.0	2411.5
Switzerland–Italy	2743.1	2794.7	2720.0	2732.5

In table 6, the initial power flows (MW) through the three interconnection are shown together with their values modified by effect of the transaction. The PTDFs evaluated from the ac load flow results are remarkably close to those predicted by the BGA based procedure; their values are: 0.2519, 0.3856, 0.5164 for the base case and: -0.0435, 0.8252, 0.1247 for solution C3.

The procedure is general enough to handle cases where several source and sink nodes are present and many PTDFs have to be minimized.

CPU times required by the GA based optimization are presented with reference to the more significant test case of the UCTE system. A complete study takes from about 40

minutes (case C2) to 3.5 hours for case C3. A restart tool is provided which allows stopping the simulation after a pre-selected number of GA iterations and restarting it later.

A careful timing analysis was carried out by means of the "profiler" available in the MATLAB environment. It resulted that over 90% of the computation time was spent in the evaluation of fitness and, in particular, in carrying out the PTDF calculation and security limit check by means of the IMML. A consistent speed improvement could be obtained by replacing IMML with partial matrix re-factorization or sparse vector solution.

Owing to the long computation times, the procedure is to be considered as a planning tool to assess the effects of various topology changes and PST operation on PTDFs and parallel flows. The flexibility of the multi-objective BGA approach also allows planners to easily account for other important aspects of the problem such as system reliability and short circuit power limitation.

6.4. State Estimation Results

6.4.1. WLS Post-Processing

Test cases include the IEEE test systems (with 14, 30 and 118 nodes) and network ITALY1. For the IEEE 14-bus system of figure 9, the set of multiple bad data shown in table 7 was tested. The repeated LNR method fails in this case because of the conforming nature of the four bad data. Repeated LNR starts by removing the correct measurement P_{21} which looks bad because the bad data on P_{12} and P_1 are consistent. Correct measurements on real power flows P_{15}, P_{51} are similarly eliminated by the LNR algorithm together with the bad data on P_{12} and P_1. Reactive power flows Q_{21}, Q_{15} and Q_{51} are removed together with the bad data on Q_{12} while the bad data on Q_1 is not recognized by this algorithm.

—●— Line flow meas. —●► Injection meas. ○ Voltage magn. meas.
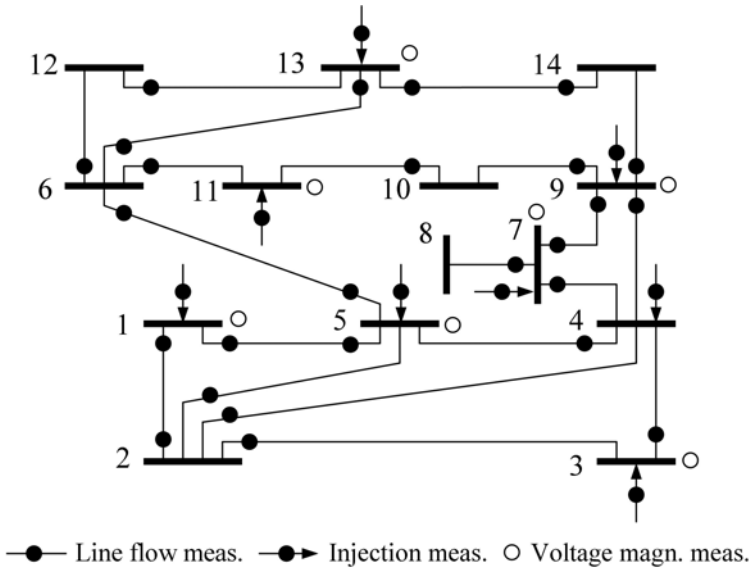
Figure 9. IEEE 14-bus system: network and measurement layout.

For what concerns the CB-GA procedure, the following values of the algorithm parameters were adopted: cross-over probability 1, mutation probability 0.08, length of artificial chromosome 16, number of individuals in the population 16, number of iterations 100. Regarding TS, the following choice of parameters was made: short term tabu tenure (for both add and drop move) 5, maximum number of iterations 200.

Table 8 shows the total number of non-linear state estimation computations carried out for a run of the CB-GA and TS procedures; the number ν of valid tested cases and the search breadth percentage δ are also reported.

The two procedures were also successfully applied to the 8 bad data test case of [39]

Table 7. IEEE 14-bus system: conforming bad data

Bad data type	Good measurements (p.u.)	Bad data values (p.u.)
P_{12}	1.55	0.85
Q_{12}	-0.20	0
P_1	2.32	1.62
Q_1	-0.30	-0.10

Table 8. IEEE 14-bus system: state estimation computations and search breadth

Method	No. of state estimations	ν	δ (%)
CB-GA	151	34	0.003
TS	192	87	0.008
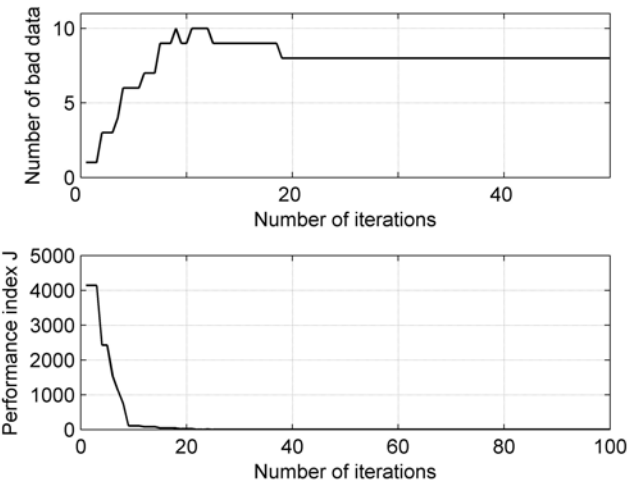
Figure. 10. CB-GA: identified bad data and performance index.

(relative to the IEEE 30-bus system) which the repeated LNR criterion fails to solve. In this case, it was necessary to broaden the range of the search with respect to the 14-bus system; for CB-GA the population size was increased to 36 members, while the maximum number of iterations of TS was raised to 1000. Figure 10, relative to the CB-GA method, shows the behavior of the individual with the minimum number of bad data and the corresponding value of the state estimation performance index $J\left[\mathbf{x}(\mathbf{b})\right]$. A solution free from bad data appears quite early (often within the first tens iterations) as shown in figure 10 by the sharp decrease in the value of $J\left[\mathbf{x}(\mathbf{b})\right]$. The actual optimum solution generally appears a few iterations later as shown by the first plot in the figure. Similarly, TS succeeds in finding the optimal solution within the first 100 iterations.

The tests performed with the IEEE 118-bus test system, with 16 interacting bad data, basically confirm the results discussed with reference to the IEEE 30-bus system. For these tests, the following parameters were chosen: for CB-GA the population size and the maximum number of iterations were taken equal to 40 and 100 respectively; for TS the maximum number of iteration was set to 1000.

The results obtained by the sensitivity based approach of Section 5.1.5 are shown in table 9. For the sake of comparison, the results obtained by employing non linear state estimation are presented, between brackets, in the same table. Both the exact and the sensitivity based version of each procedure are successful in identifying the bad data, but the CPU times are about two orders of magnitude smaller for the sensitivity based approach.

Table 9. IEEE 118-bus system: 16 bad data case

Method	No. of state estimations		ν		CPU time (s)	
CB-GA	6992	(10840)	195	(221)	16.8	(1565.5)
TS	978	(981)	672	(665)	1.9	(189.5)

Table 10. ITALY1 system: 30 bad data case

Method	No. of state estimations	ν	CPU time (s)
CB-GA	12815	521	25.7
TS	997	476	2.3

The sensitivity based procedures were also tested with Italy1; results relative to a 30 bad data case are shown in table 10.

Both the CB-GA and TS methods succeed in finding the solution within a reasonable CPU time, but they lack in a reliable method to control the breadth of the search. The small number of state estimations carried out by CB-GA depends on systematic elimination of non-promising test cases and on explicit memorization of already computed results. Regarding TS, the number of computed state estimations depends on the number of iterations, but there is no guarantee that all tested cases are different from each other.

From the timing results, it is apparent that the methods above and, in particular, those variants based on full state estimation, do not lend themselves to be applied in a real time environment. Their use appears to be limited to after-the-fact analysis of troublesome cases of multiple interacting bad data.

The situation is somewhat different with reference the variants employing the sensitivity based method. In particular, the TS version appears well suited for on line applications as is demonstrated by the timing results reported in table 10.

An improvement in the performance of the procedures could be obtained by exploiting the poor propagation of the bad data effects through large-scale systems. A measurement "pocket" is defined as a set of suspect bad data together with the state variables involved. Pockets can be examined one at a time looking for the actual bad data [40].

A different localization strategy was implemented, with reference to the sensitivity based methods only. Suspect measurements are subdivided into clusters, trying to put measurements with a remarkable mutual influence (evaluated by the entries of the sensitivity matrix $\tilde{\mathbf{W}}$) into the same group. The efficiency of this strategy depends on the number of clusters produced. The 421-bus ITALY1 system with 12 bad data (belonging to 3 separate groups) was taken as an example. By employing TS and the proposed clustering technique, it was possible to speed up the solution by a factor of two; all bad data were detected, in 0.89 CPU seconds compared with 1.76 CPU seconds of the standard procedure. However, when many bad data are present, as in the 30 bad data case of table 10, the time overhead deriving from the sequential solution of all the 15 clusters produced in that case, makes the clustering method slower than the standard one.

The branch-and-bound procedure of [39] was also implemented; tests carried out with reference to the above mentioned IEEE and Italian systems lead to the same results of CB-GA and TS. The CB-GA and TS procedures present a consistent advantage over branch-and-bound in terms of CPU times as a result of the reduced number of computed test cases. Indeed, the convergence of branch-and-bound strictly depends on the search breadth parameter which establishes an upper bound on the number of alternative solutions to be examined. If search breadth is taken too small, there is a non-negligible risk of missing the actual optimal solution. On the contrary, if search breadth is too large, computation times become overlong, specially when many bad are present. Anyway, a large number of test case evaluations cannot be avoided because of the exponential growth in the size of the binary tree of possible solutions with the number of bad data.

6.4.2. Maximum Agreement Algorithm

In the implementation of MAA, it was found expedient to code some computation intensive functions which could not perform efficiently in MATLAB by using the FORTRAN language. To this purpose, the external interfacing facilities of MATLAB were exploited; in particular the observability enforcing mechanism and the access to the database of solved cases were implemented in this way.

The MAA procedure stopping criterion was set, according to (21), by assuming a fraction of bad data $\varepsilon = 5\%$. It was required anyway that at least 50 iterations of BGA were performed in order to guarantee a sufficient exploration of the solution space. A crossover probability of 0.9 and a mutation probability of 0.02 were used.

Removing the assumption that all errors are zero except for the bad data, a less stringent condition has to be determined to ascertain whether a measurement agrees with the tentative solution corresponding to a given essential set. An agreeing measurement is one that satisfies the condition

$$\left(\left|r_k\right|/\sigma_k\right) \le t \qquad \forall \, k \notin \bar{E}$$

being σ_k the standard deviation of measurement and a given threshold. In the tests, the agreement threshold t was taken equal to 2.5.

The MAA procedure has no problems in finding the correct solution to the four bad data case of table 7 relative to the 14-bus IEEE test system shown in figure 9. A population of 4, 8, and 16 individuals was adopted. GA runs were executed; it was found that, although bad data free solutions were achieved in all cases, a faster convergence is obtained when the largest population is adopted. This behavior can be explained with the implicit parallelism mechanism of GAs, described in [33]. Larger populations allow a better circulation of the most promising substrings and therefore a rapid appearance of the best solution. In contrast, smaller populations, require more iterations to converge to the solution, even though they have the advantage of requiring a reduced computation effort.

The MAA procedure was successfully applied to the 8 bad data test case of [39] relative to the IEEE 30-bus system. The maximum agreement state-estimation procedure takes between 6 and 8 s CPU time. A larger number of state estimations is required in this case due to the increased number of bad data.

A CPU time of 10 to 12 s is required to obtain the correct solution and to identify four interacting bad data with reference to the IEEE 118-bus system. The average CPU time T_{tot} required to carry out the whole computation is reported in table 11 with reference to the test cases considered. The GA parameters used for these runs are the same mentioned above and an 8 individual population was adopted.

A nice feature of the MAA procedure is that a feasible solution of the state-estimation problem, if not the actual optimum, is generally found after very few generations. The

Table 11. Timing results for the three test systems

IEEE test system	No. of bad data	T_{tot} (s)	T_{first} (s)
14-bus	4	5.4	0.4
30-bus	8	7.5	0.7
118-bus	4	10.8	0.9

average time of the first appearance of the correct solution is shown as T_{first} in table 11.

Some tests were carried out to compare the MAA procedure with the LMS method. By adopting GA based optimization in both cases, the two approaches behave similarly, with regard to both results and computation timings. However, LMS requires that the value of the

surplus must be known beforehand. Evaluating the surplus is computationally demanding since many possible combinations of measurement removal have to be checked. This task was performed by a separate program requiring 0.4 s, 3 s and 19 s for the 14, 30, and 118-bus IEEE test systems.

The main obstacle to the application of MAA to large-scale systems is the growth of the search space dimensions. Techniques to overcome this drawback include problem partitioning or using MAA as a post processing tool combined with some faster state estimator.

7. CONCLUSION

In this work, the authors' experience in GA application to power system problem is illustrated with particular reference to planning problems and state estimation.

In Section 4.1 two versions of GA were employed to solve a mixed integer optimization problem in the framework of reactive compensation planning. In particular, the capacitor placement problem is solved by sequential linear programming which requires a MILP problem to be solved at each iteration. BGA and μGA based procedures were validated with reference to a CIGRE network and to actual systems representing the Sicilian regional network and continental Italy. The results were compared with those obtained using the branch-and-bound method. A hybrid procedure was also designed to exploit the best characteristics of the branch-and-bound and genetic algorithm approaches. This procedure consists in carrying out a GA search including the incomplete branch-and-bound solution in the initial population. The procedure achieved saving in installation cost of about 16% with respect to incomplete branch and bound solution.

In Section 4.2, a procedure based on BGA was employed to minimize the "power transfer distribution factor" (PTDF) relative to a given transmission corridor by topological reconfiguration and phase-shifting transformer (PST) operation. PTDF optimization should be employed as a tool for the control of undesired parallel flows through interconnection corridors. The constraints of N and N-1 security are accounted for by means of subsidiary objective functions in the framework of multi-objective optimization. The choice of adopting a solution approach based on GAs was made because of the discrete nature of most of the control variables and also because the minimum PTDF objective function cannot be expressed analytically with respect to the control variables. The results, obtained with reference to a CIGRE system and to the large-scale UCTE network, confirm the effectiveness of topological reconfiguration and PST operation as a means to reduce the PTDF corresponding to parallel flows which should be minimized.

In Section 5.1 the problem of correctly identifying multiple interacting bad data in the framework of the WLS method was considered by employing a GA variant proposed by Chu-Beasley (CB-GA) and tabu search (TS). CB-GA is generally quite effective, especially when use is made of explicit storage of past solutions to avoid re-computation of already solved cases. TS relies on the use of short and long term memory to carry out its search; it is very efficient, but it cannot guarantee that duplicated solution cases occur during the search. Both these methods behave satisfactorily in the identification of bad data and are remarkably faster than a branch-and-bound method implemented for reference purposes. Tests were carried out

with reference to the IEEE sample systems with 14, 30 and 118 buses and to network ITALY1.

Another promising approach to state estimation is that described in Section 5.2; instead of post-processing the WLS solution, use is made of the maximum agreement algorithm (MAA) which is capable of rejecting a possibly large number of bad data automatically. The idea behind MAA is that the best state estimate should satisfy the majority of measurements acting as voters in a general election. BGA was tailored to the characteristics of MAA by adopting a specialized mutation operator designed to enforce observability of tentative state estimation cases. MAA was tested with reference to well known IEEE test systems. Despite of a non-optimized implementation, timing results are interesting. The application of GA techniques to MAA often provides the optimal solution within the first few iterations and with a much reduced time.

In conclusion, our experience with GAs in the solution of power system analysis problems highlights the fundamental characteristics of GAs, namely their flexibility and generality. Indeed, no deterministic optimization algorithm can be shown to be flexible enough to adapt to completely different problems such as reactive compensation planning, topology rescheduling and state estimation.

GAs appear one of the most useful general tools available to anyone involved in the solution of optimization problems and it is conceivable that GAs will still play this role for many years in the future.

REFERENCES

[1]   Holland, JH. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. Cambridge, USA: MIT Press; 1992.

[2]   Miranda, V; Ranito, JV; Proença, LM. Genetic algorithms in multistage distribution network planning. IEEE Trans. Power Systems, November 1994,Vol. 9, No. 4, 1927–1933.

[3]   Gallego, RA; Monticelli, A; Romero, R. Transmision system expansion planning by an extended genetic algorithm. *IEE Proc.-GTD,* May 1998, Vol. 145, No. 3, 329–335.

[4]   de Silva, I; Rider, MJ; Romero, R; Murari, CA. Genetic algorithm of Chu and Beasley for static and multistage transmission expansion planning", *IEEE 2006 PES General Meeting,* June 2006, Montréal, Canada, 1–7.

[5]   Park, JB; Park, YM; Won, JR; Lee, KY. An improved genetic algorithm for generation expansion planning. *IEEE Trans. Power Systems,* August 2000, Vol. 15, No. 3, 916–922.

[6]   Teixeira Firmo, H; Loureiro Legey, LF. Generation expansion planning: an iterative genetic algorithm approach. *IEEE Trans. Power Systems,* August 2002, Vol. 17, No. 3, 901–906.

[7]   Levitin, G; Kalyuzhny, A; Shenkman, A; Chertkov, M. Optimal capacitor allocation in distribution systems using a genetic algorithm and a fast energy loss computation technique. *IEEE Trans. Power Delivery.* April 2000, Vol. 15, No. 2, 623–628.

[8]   Huang, SJ. An immune-based optimization method to capacitor placement in a radial distribution system. *IEEE Trans. Power Delivery,* April 2000, Vol. 15, No. 2, 744–749.

[9]   Mantovani, JRS; Modesto, SAG; Garcia, AV. VAr planning using genetic algorithm and linear programming. *IEE Proc.-GTD, May* 2001, Vol. 148, No. 3, 257–262.

[10]   Gallego, RA; Monticelli, AJ; Romero, R. Optimal capacitor placement in radial distribution networks. *IEEE Trans. Power Systems,* November 2001, Vol. 16, No. 4, 630–637.

[11]   Li, F; Pilgrim, JD; Dabeedin, C; Chebbo, A; Aggarwal, RK. Genetic algorithms for optimal reactive power compensation on the national grid system. *IEEE Trans. Power Systems,* February 2005, Vol. 20, No. 1, 493–500.

[12]   Maifeld, TT; Sheblé, GB. Genetic-based unit commitment algorithm. *IEEE Trans. Power Systems,* August 1996, Vol. 11, No. 3, 1359–1370.

[13]   Carneiro, A; Leite, P; Carvalho, A. A genetic algorithm approach to optimize the operation planning of hydrothermal system scheduling. 1998 International Conference on Power System Technology, August 1998, Beijing, China, Vol. 1, 547–551.

[14]   Richter, CW; Sheblé, GB. A profit-based unit commitment GA for the competitive environment. *IEEE Trans. Power Systems,* May 2000, Vol. 15, No. 2, 715–721.

[15]   Chen, H; Wang, X. Cooperative coevolutionary algorithm for unit commitment. *IEEE Trans. Power Systems,* February 2002, Vol. 17, No. 1, 128–133.

[16]   Arroyo, JM; Conejo, AJ. A parallel repair genetic algorithm to solve the unit commitment problem. *IEEE Trans. Power Systems,* November 2002, Vol. 17, No. 4, 1216–1224.

[17]   Walters, DC; Sheblé, GB. Genetic algorithm solution of economic dispatch with valve point loading. *IEEE Trans. Power Systems,* August 1993, Vol. 8, No. 3, 1325–1332.

[18]   Sheblé GB; Brittig, K. Refined genetic algorithm - Economic dispatch example. *IEEE Trans. Power Systems,* February 1995, Vol. 10, No. 1, 117–124.

[19]   Yang, HT; Yang, PC; Huang, CL. Evolutionary programming based economic dispatch for units with non-smooth fuel cost functions. *IEEE Trans. Power Systems,* February 1996, Vol. 11, No. 1, 112–118.

[20]   Damousis, IG; Bakirtzis, AG; Dokopoulos, PS. Network-constrained economic dispatch using real-coded genetic algorithm. *IEEE Trans. Power Systems,* February 2003, Vol. 18, No. 1, 198–205.

[21]   Wen, F; David AK. Coordination of bidding strategies in energy and spinning reserve markets for competitive suppliers using a genetic algorithm. *IEEE 2000 PES Summer Meeting,* July 2006, Seattle, USA, Vol. 4, 2174–2179.

[22]   Wong, KP; Chung, CY. Evolutionary computation techniques for power market equilibrium determination. *IEEE 2006 PES General Meeting,* June 2006, Montréal, Canada, 1–4.

[23]   Richter, CW; Sheblé, GB. Genetic algorithm evolution of utility bidding strategies for the competitive marketplace. *IEEE Trans. Power Systems,* February 1998, Vol. 13, No. 1, 256–261.

[24]   Sheblé, GB; Richter, CW; Ashlock, D. Comprehensive bidding strategies with genetic programming/finite state automata. *IEEE Trans. Power Systems,* November 1999, Vol. 14, No. 4, 1207–1212.

[25] Petrov, V; Richter, CW; Sheblé, GB. Predatory gaming strategies for electric power markets. Int. Conf. on *"Electric Utility Deregulation and Restructuring and Power Technologies",* April 2000, London, UK, 488–492.

[26] Gao, F; Gutierrez-Alcaraz, G; Sheblé, GB. Comparison of artificial life techniques for market simulation. *39th Hawaii Int. Conf. on "System Sciences",* January 2006, Hawaii, USA, Vol. 10, paper 243a.

[27] Delfanti, M; Granelli, GP; Marannino, P; Montagna, M. Optimal capacitor placement using deterministic and genetic algorithms. *IEEE Trans. Power Systems,* August 2000, Vol. 15, No. 3, 1041–1046.

[28] Granelli, GP; Montagna, M; Zanellini, F; Bresesti, P; Vailati, R. A genetic algorithm-based procedure to optimize system topology against parallel flows. *IEEE Trans. Power Systems,* February 2006, Vol. 21, No. 1, 333–340.

[29] Gastoni, S; Granelli, GP; Montagna, M. Multiple bad data processing by genetic algorithms. *IEEE 2003 Bologna PowerTech Conf.,* June 2003, Bologna, Italy.

[30] Gastoni, S; Granelli, GP; Montagna, M. Robust state estimation procedure based on the maximum agreement between measurements. *IEEE Trans. Power Systems,* November 2004, Vol. 19, No. 4, 2038–2043.

[31] Krishnakumar, K. Micro-genetic algorithms for stationary and nonstationary function optimization. *Proc. SPIE Conf. on "Intelligent Control and Adaptive Systems",* February 1990, Philadelphia, USA, 289–297.

[32] Chu, PC; Beasley, JE. A genetic algorithm for the generalized assignment problem. *Computers Operations Research,* January 1997, Vol. 24, No. 1, 17–23.

[33] Goldberg, DE. Genetic Algorithms in Search, Optimization and Machine Learning. Reading, USA: Addison-Wesley Publishing, 1989.

[34] Lebow, WM; Rouhani, R; Nadira, R; Sobiewski, DW; Pal, MK; Usoro, PB; Mehra, RK; Bhavaraju, MP. A hierarchical approach to reactive volt ampere (VAR) optimization in system planning. *IEEE Trans. Power Appar. and Systems,* August 1985, Vol. PAS-104, No. 8, 2051–2057.

[35] Mili, L; Phaniraj, V; Rousseeuw, PJ. Least median of squares estimation in power systems. *IEEE Trans. Power Systems,* May 1991, Vol. 6, No. 2, 511–523.

[36] Mili, L; Cheniae, MG; Vichare, NS; Rousseeuw, PJ. Algorithms for least median of squares estimation of power systems. *Proc. 35th Midwest Symp. on "Circuits and Systems",* August 1992, Washington, USA, 1276–1283.

[37] Magnago FH; Abur, A. A unified approach to robust meter placement against loss of measurements and branch outages. *IEEE Trans. Power Systems,* August 2000, Vol. 15, No. 3, 945–949.

[38] Land, AH; Powell, S. Fortran Codes for Mathematical Programming: Linear, Quadratic and Discrete. London, UK: John Wiley, 1970.

[39] Monticelli, A; Wu, F; Yen, M. Multiple bad data identification for state estimation by combinatorial optimization. *IEEE Trans. Power Delivery*, July 1986, Vol. 1, No. 3, 361–369.

[40] Monticelli, A. State Estimation in Electric Power Systems. A Generalized Approach. Boston, USA: Kluver Academic Publishers, 1999.

Chapter 3

PREDICTION OF PERFORMANCE AND POLLUTANT EMISSION FROM PULVERIZED COAL UTILITY BOILERS

N. Spitz[1], R. Saveliev[1], E. Korytni[1], M. Perelman[1], E. Bar-Ziv[1] and *B. Chudnovsky[2]***

[1] Ben-Gurion University, P.O.B 653, Beer-Sheva, Israel 84105
[2] Israel Electric Corporation, P.O.B 10, Haifa, Israel 31000

ABSTRACT

A three-step methodology was developed to provide reliable prediction of a coal's behavior in a utility boiler: (1) Extracting the combustion kinetic model parameters by combining experimental data from a pilot-scale test facility, Computational Fluid Dynamic (CFD) codes and an artificial neural network. While the combustion kinetic parameters used in the model code will not correspond to the combustion rate of a single particle of coal, these parameters do describe the combustion behavior of a "macroscopic" sample of tested coal. (2) Validation of the combustion kinetic model parameters by comparing diverse experimental data with simulation results calculated with the same set of model parameters. (3) The model parameters are then used for simulations of full-scale boilers using the same CFD code. For operational engineering information needed by the utility operator, we apply the predicted results to EXPERT SYSTEM, a boiler supervision system developed by Israel Electric Corporation (IEC). Four different bituminous and sub-bituminous coals with known behavior in IEC 550MW opposite-wall and 575MW tangential-fired boilers were used to show the adequacy of the methodology. The predictions are done with the CFD code, GLACIER, propriety of Reaction Engineering International (REI). Preconfigured GLACIER models of the test and full-scale furnaces were purchased from REI and validated by our group. This book chapter will include a detailed description of the methodology, test furnace facility and an example of the experimental and predictive combustion results from the four coals used to test the methodology. In addition, two previously unknown coals will

---

* E-mail address: barziv@bgu.ac.il
** E-mail address: borisz@iec.co.il

be examined prior to their firing in the utility boilers and prediction of their behavior and operational parameters in the two boilers will be carried out.

1. INTRODUCTION

Pulverized coal is an important fuel for electricity production [1] and will continue to be important for decades. Since coal is a natural resource that depends on many factors and parameters, it has variable properties and composition. Because of this heterogeneity, the combustion behavior and pollutant emissions are different for each coal. The days when utility companies used coal from the same mine for years are gone [2-3] and they are faced now with the challenge of firing very different types of coal in the same boiler. The great variability of the properties and composition of the different coals imposes tremendous operational difficulties and requires innovative approaches to aid in decision making and operational strategies. In addition to this, more stringent environmental regulations are being enforced [4] and the utility company needs to lower its emissions while staying profitable. Different types of coals combined with modifications to the combustion process can be a cost-effective way of improving combustion behavior and yet lowering pollutant emissions from power plants. Because of the variability of combustion behavior of each coal, utility companies go to great lengths to test the coals before purchasing them for use in their utility boiler. These tests, which include preliminary tests at coal site and then full-scale tests at operator's site, are very costly. Although much advanced in recent years, the predictions of Computational Fluid Dynamic (CFD) models for coal combustion are not sufficient alone to select a coal for full-scale use [1,5]. CFD models can give good predictions of coal combustion in utility boilers if the coal combustion kinetic model parameters are known. These kinetic parameters are usually determined from sub-models.

The aim of this study was to develop a low-cost method to predict the combustion behavior and pollutant emission from coals previously unknown to the utility company. To predict the combustion behavior and pollutant emissions of coal in pulverized-coal utility boilers, we developed a three-step methodology to provide reliable prediction of a coal's behavior in a utility boiler:

1. Extracting the combustion kinetic model parameters by combining experimental data from a pilot-scale test facility, CFD codes and an artificial neural network. While the combustion kinetic parameters used in the model code will not correspond to the combustion rate of a single particle of coal, these parameters do describe the combustion behavior of a "macroscopic" sample of tested coal.
2. Validation of the combustion kinetic model parameters by comparing diverse experimental data with simulation results calculated with the same set of model parameters.
3. The model parameters are then used for simulations of full-scale boilers using the same CFD code. For operational engineering information needed by the utility operator, we apply the predicted results to EXPERT SYSTEM, a boiler supervision system.

The predictions and full-scale tests were done on two boiler types: 550MW opposite-wall and 575MW tangential-fired. The coals tested and presented here are: (1) Billiton-BB Prime – a South African bituminous coal (Billiton or SA in text), (2) Glencore-Adaro – an Indonesian sub-bituminous coal (Adaro or Ad), (3) Drummond-La Loma – a Colombian bituminous coal (Drummond or Dr), for tangential-fired boiler only, (4) Glencore-Russian – a Russian bituminous coal (Russian or Gln). For both boilers, we predicted the behavior and emissions from two coals previously unknown to IEC: Guasare-Venezuelan (Venezuelan or Ven) – a Venezuelan bituminous coal and KPC-Melawan – an Indonesian sub-bituminous coal (Mel). For opposite-wall boiler we also simulated the combustion of Glencore-Russian coal. The predictions are done with the CFD code, GLACIER, propriety of Reaction Engineering International (REI). Preconfigured GLACIER models of the test and full-scale furnaces were purchased from REI and validated by our group.

2. METHODOLOGY

The combustion and pollutant formation processes involved in utility boilers depend on (1) the geometry and materials of the combustion chamber that includes the burners, air inlets, coal and air feed factors; all of these are known by the utility operator, and (2) the composition and properties of the coal. Of the latter, the chemical composition and heat values can be analyzed in the chemist's laboratory but the CFD code includes twelve model parameters depicting kinetic and thermal properties of the coal; all of which cannot be analyzed or precisely estimated in the laboratory. If all coal quality parameters (thermodynamic, thermal, and optical properties as well as chemical composition and kinetic parameters) are known, one can use a sophisticated CFD code to simulate the performance of the system. However, only some of the required information is available and hence one must carry out, prior to the simulation, experimental testing of the coal in order to acquire the missing information required for the simulation. One can obtain all parameters required by carrying out testing in a series of laboratory-scale equipment. These measurements, however, may lack credibility in the representation of the coal/blend involved as very small samples (normally milligrams) are used in these laboratory testing. In order to obtain credible parameters, one needs to use a representative quantity of coal/blend, normally in the vicinity of 400-600 kg. This amount can be tested only in a pilot scale facility.

Coal combustion processes in test facilities, such as this one, are simplified because many of the physical and chemical processes that take place in a utility boiler are absent. Therefore, it is easier to validate the CFD combustion model for the test furnace and apply the validated model, with the same kinetic parameters but with different boundary conditions, to a utility boiler. The approach described here is based on a simplified kinetics model, shown in Figure 1. The model includes a two-step devolatilization process with three parameters each (activation energy, pre-exponential factor, and weight fraction for each step), totaling 6 parameters for devolatilization; a char combustion model with two parameters (activation energy and pre-exponential factor); one parameter to represent the nitrogen distribution between char and volatile matter; one parameter to represent the distribution of HCN and NH_3, and one parameter to represent NO_x release from char. While it cannot be said that the kinetic parameters used in the model will correspond to the combustion rate of a single

particle of coal, these parameters do describe the combustion behavior of a "macroscopic" sample of the tested coal. For the goal of predicting combustion behavior of coals in a utility boiler furnace, this simplified model used with the CFD code gave good results.

A flow chart of the three-step methodology we developed is detailed in figure 2. From the flow chart one can see that there is no attempt to predict the combustion behavior of the utility boiler based on the combustion behavior of the test furnace. The three-steps are:

1. Obtaining the combustion kinetic model parameters required for the numerical simulations of utility boilers from a series of experiments in a 50 kW pilot-scale test facility, CFD codes and optimization algorithm. A pilot-scale furnace was constructed at our facility for this purpose. The predictions are done with the CFD code, GLACIER, propriety of Reaction Engineering International (REI). The optimization algorithm developed is based on an artificial neural network (ANN). To simplify interpretation of the combustion experiments the furnace was designed with axial symmetry and with a two-zone configuration: a well-mixed reaction zone followed by a plug flow region. Furthermore, the coal is burned within the test furnace at temperatures and concentrations similar to those prevailing in a pulverized-coal furnace. The model parameters used in the simulation are modified until good agreement is obtained between the results of the numerical simulation and the experimental data from the test furnace. The drawback of the multi-parameter fitting technique is whether the set obtained is unique to the system or there are other sets that would also provide good agreements, due to numerical inter-compensation.

2. Validation of the combustion kinetic model parameters by comparison of different experimental data with simulation results obtained by the set of combustion kinetic parameters. To attain confidence in the fitting technique numerous experiments were carried out at different operating conditions (i.e.: stoichiometric ratios, coal size distributions, and staged burning). The same set of model parameters was used to check the correlation of the experimental and the numerical results. The more data to fit the more confidence there is in the obtained set of parameters. Yet, some uncertainty will always remain, though very difficult to quantify.

3. The extracted combustion kinetic model parameters are then used for simulations of full-scale boilers using the same CFD code. For operational engineering information needed by the utility operator, we apply the predicted results to EXPERT SYSTEM, a boiler supervision system developed by Israel Electric Corporation (IEC).
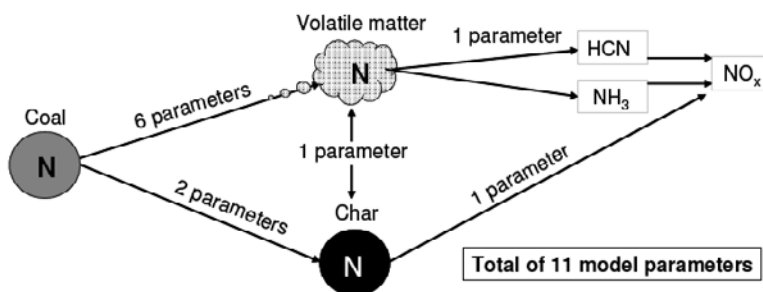
Figure 1. Kinetic model for devolatilization – combustion of coal.
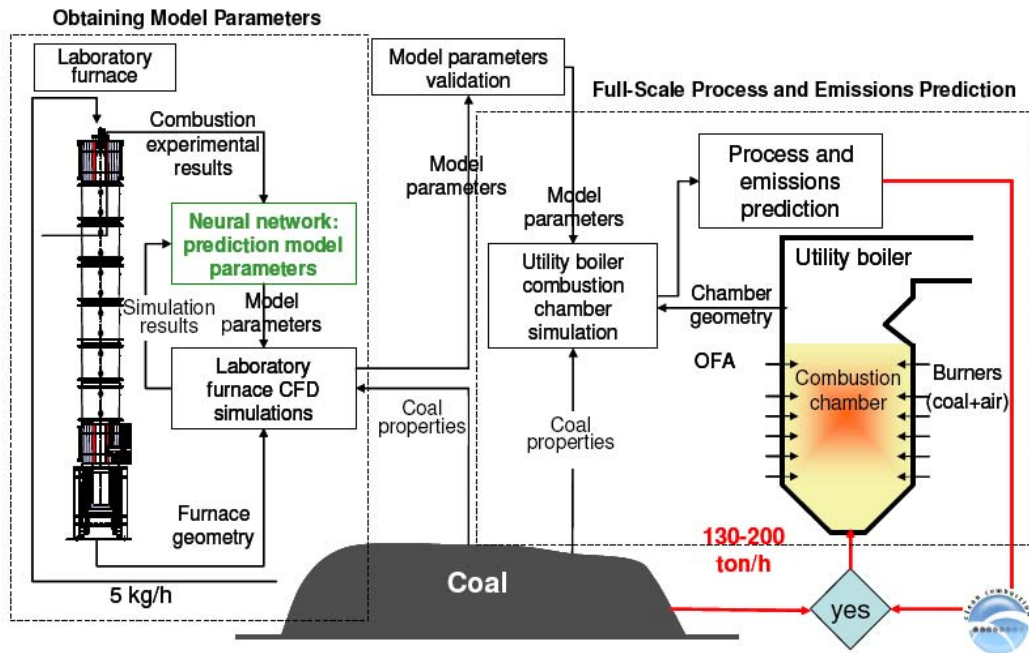
Figure 2. Methodology for predicting coal performance in full-scale utility boilers.
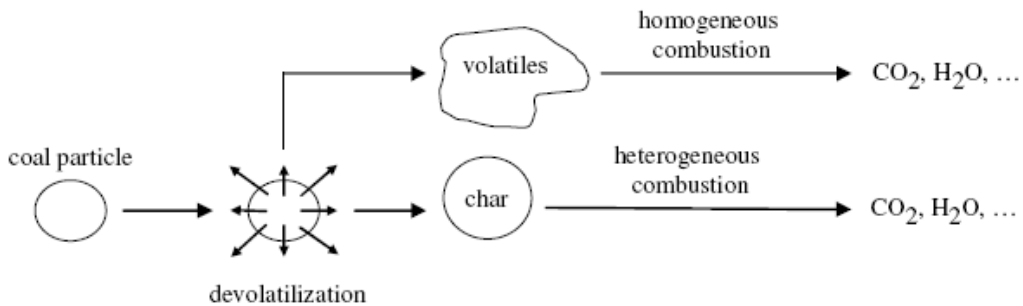
Figure 3. Combustion processes of a coal particle [6].

2.1. Kinetic Model Parameters

This chapter describes the combustion kinetics, mechanisms and parameters, used by GLACIER, the commercial CFD program used for this work.

When coal is burned in a combustion chamber, three main processes occur: (1) coal devolatilization, (2) volatiles combustion and (3) char oxidation (char burnout), where figure 3 shows these processes [6]. During devolatilization water evaporates first and then the light gases and tars. These 'volatiles', which are mainly composed of hydrocarbons, oxidize in the combustor environment. The solid material left is rich in carbon, poor in hydrogen and oxygen, but also contains nitrogen, sulfur and mineral matter. This material is termed 'char'

and the materials in it continue to react during the process of 'char oxidation' or 'char burnout' [7].

The reaction processes of the coal particles in the GLACIER code include devolatilization, char oxidation and gas-particle interchange [8]. Particles are made up of coal, char, ash and moisture. Ash is inert by definition and volatile mineral matter is considered part of the volatile matter of the coal. Particle swelling is empirically accounted for and due to the small Biot number the particles are assumed to be isothermal. Particle reactions are characterized by multiple parallel reaction rates with fixed activation energies. The off-gas from particle reactions is assumed to be of constant elemental composition. Coal devolatilization is modeled using the two-step model proposed by Ubhayakar et al. [9]. Equilibrium codes for volatile matter oxidation are rather common in combustion modeling and various codes are available in the literature [1,5,10,11]. A global Arrhenius model is used to model heterogeneous char oxidation [8].

Volatile Matter Kinetics

When the coal particle is exposed to the high furnace temperatures, instantaneous devolatilization of the volatile matter (VM) occurs. The volatile matter separates from the char portion of the coal, ignites and oxidizes. The time frame for devolatilization is 1-5 ms and for volatile matter oxidation 50-100 ms.

Devolatilization

Ubhayakar et al. [9] modeled the rapid devolatilization process with two competing first-order reactions, depicted in figure 4. Each reaction illustrates a devolatilization process of the DAF (dry ash free) raw coal, CH_x where $0 \leq x \leq 1$. In the devolatilization process volatile matter, V_1 and V_2, is released from the coal leaving a residual char, R_1 and R_2. The rate constants of these two competing processes are k_1 and k_2. Y_1 and Y_2 are the mass stoichiometry coefficients, which are selected according to the (H/C) ratio of the coal; $Y_n \sim x/x_n$, where $x_n = (H/C)$ of V_n, $n=1,2$. Good agreement between predicted and measured values was achieved using Y_1 as the volatile matter fraction and $Y_2 = 2*Y_1$.

$$DAF\ coal,\ CH_x\ (0 \leq x \leq 1) \xrightarrow{k_1} Y_1 V_1 + (1-Y_1)R_1 \quad k_1 = A_1 \exp(-E_1/RT)$$

$$\xrightarrow{k_2} Y_2 V_2 + (1-Y_2)R_2 \quad k_2 = A_2 \exp(-E_2/RT)$$
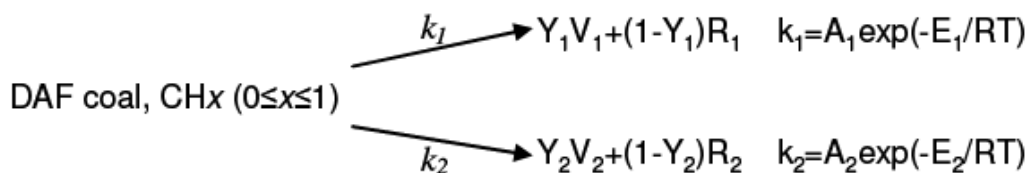
Figure 4. Devolatilization model developed by Ubhayakar et al. [9]: two-step devolatilization model of coal, CH_x ($0 \leq x \leq 1$), DAF stands for dry ash free. k_n are the two rate constants [s⁻¹], Y_n are the mass stoichiometry coefficients, V_n and R_n stand for the volatile matter and residual chars, respectively. For all parameters, $n=1,2$. In the rate constant equations: A_n[s⁻¹] and E_n[kcal/mol] are the preexponential (frequency) factor and activation energy, respectively, R the gas constant and T the temperature of the volatile gases in the vicinity of the coal particle.

The following kinetic parameters for the 2-step devolatilization process are inputted to the GLACIER code: mass stoichiometry coefficients (Y_1 and Y_2), pre-exponential factors (A_1, A_2) and activation energies (E_1, E_2). For a tested blend, the above parameters are entered for

each component coal. The instantaneous mass rates of volatile matter production by the competing reactions in figure 4 at time t are assumed as (Eq. 1):

$$\dot{m}_n \ (r_i, t) = Y_n k_n m_n(r_i, t), n = 1, 2 \tag{1}$$

where $\dot{m}_n$ is the mass of the part of the coal particle with radius r_i that has not yet reacted, the dot represents the time rate of change, $k_n = B_n exp[E_n/RT_c(r_i, t)]$, B_n and E_n are the pre-exponential (frequency) factor and activation energy, respectively, R the gas constant and $T_c(r_i, t)$ the temperature of particles of radius r_i.

The numerical solution to the above model results in a low activation energy reaction significant at lower particle temperatures and a higher activation energy reaction significant at higher temperatures. The high temperature process led to higher yields of lower (H/C) volatiles and char than the low activation process. In other words, a high yield reaction is favored at high temperatures, while a low yield reaction is favored at low temperatures. E_1 and E_2 govern the degree of overall volatile formation as a function of the heating rate. This is the advantage of the two-step devolatilization mechanism, as it predicts the volatile fraction or the changes in the yield as a function of heating rate [12,13].

Volatile Matter Oxidation

Volatile matters oxide very fast in comparison to chemical and physical processes in the furnace; hence it was assumed that once volatile matters are produced, instantaneous equilibrium for gas oxidation reactions takes place. The equilibrium codes are in rather common use in combustion and various codes are available in the literature [1,5,10,11]. One of the most common modeling approaches for volatiles oxidation is the Mixture Fraction approach [5]. Because the reaction time scale is much shorter than the mixing time scale, the physical processes are treated in detail but local instantaneous (infinite rate) chemistry is assumed and computed using an equilibrium algorithm. For two individual inlet streams a conserved scalar variable, f, the mixture fraction, is used to define the level of mixing of primary and secondary mass components, Eq. 2:

$$f = \frac{m_p}{m_p + m_s} \tag{2}$$

where f is the mass fraction of the primary stream component, m_p is the mass originating from the primary source and m_s is the mass originating from the primary source. Normally, for combustion applications, fuel is taken for the primary stream and the oxidizer is the secondary stream. The advantage of this mixture fraction variable is that any other conserved scalar, s, which is a function of f, such as the fuel density, can be calculated from the local value of f. For instance, Eq. 3:

$$s = fs_p + (1 - f)s_s \tag{3}$$

Therefore the mixture fraction approach lowers the number of conserved scalars required to describe a combustion system. The validity of this approach requires that the turbulent diffusivity and boundary conditions of all gas phases be the same, which is not unreasonable in most large combustion systems. Statistical probability density functions (pdf's) are used at each point in the flow field to obtain the mean properties of chemical species, temperature and flow dynamics based on the local instantaneous equilibrium calculated with the mixture fraction approach. The oxidation kinetics, with fixed activation energies, are pre-programmed into the GLACIER CFD code and cannot be changed by the user.

Char Kinetics

The char oxidation rate combines the effects of surface reactivity and pore diffusion as described by Hurt and Mitchell [14] and shown Eq. 4:

$$q = k_s P_s^{\ n} \tag{4}$$

where q is the combustion rate normalized by the particle external surface area, k_s is the global rate coefficient [kg-Carbon/m^2-s], given by Eq. 5:

$$k_s = A \exp(-E / RT_p) \tag{5}$$

A [kg-Carbon/m^2-s] represents the global pre-exponential factor, E [kJ/mol] the global activation energy, R the gas constant and T_p [K] the temperature of the char particle. P_s is the partial pressure of oxygen at the particle surface and n is the global reaction order. Three of the above parameters governing char oxidation can be varied in GLACIER: n, A and E.

NO$_x$ Kinetics

Most coals contain 0.5-2.0% nitrogen by weight. Bituminous coals generally have high nitrogen levels and anthracite low nitrogen levels. The large fuel-nitrogen content of coal can result in substantial NO$_x$ emissions. The formation and destruction of NO$_x$ in combustion systems is very complicated. During combustion, nitrogen found in the coal or in the combustion air is transformed by many different chemical mechanisms to nitrogen-containing species such as: nitric oxide (NO), nitrogen dioxide (NO$_2$), nitrous oxide (N$_2$O), ammonia (NH$_3$), hydrogen cyanide (HCN) and amine compounds. Coal combustion systems emit nitrogen oxides mostly in the form of nitric oxide (NO), with a small part appearing as nitrogen dioxide (NO$_2$) [15,16].

Fuel-nitrogen is distributed between the volatiles and the char and it is released during devolatilization and char oxidation. This split is potentially important for NO$_x$ formation. The fraction of nitrogen released with volatiles during devolatilization depends on the fuel type, the temperature, the heating rate and the residence time [17]. NO$_x$ formation increases with the oxygen content in the coal, meaning that a high rank coal will form less NO$_x$. Increasing the temperature and the residence time favors the conversion of coal-N to volatile-N. In regular pulverized coal combustion, the source of 60-80% of the NO$_x$ is volatile-N. Because of its early release, the volatile matter of the coal has a major impact on the amount of NO formation. The heating rate and final temperature of the coal particle most probably influence the devolatilization processes, and, as a result, the amount of nitrogen released from the coal.

Temperatures above 2000 K are required for complete char-N release. The magnitude of char-NO_x formation is dependent on char properties, reactivity and internal surface area, and the combustor environment; temperature and specie fields [18]. Several factors determine these characteristics of the char: fuel-type, air-staging, burner injection and the thermal and physical conditions in the combustor. Basically, in laboratory conditions, one will find that the contribution of NO_x from char combustion is less than that from volatiles combustion. However, the homogeneous NO_x reactions are influenced more by combustion modifications than the heterogeneous NO_x reactions. Therefore, in "real-life" facilities, the control of NO_x formed from heterogeneous reactions is more difficult. Spinti and Pershing [19] concluded that char-N conversion to NO_x, for typical pulverized coal combustor conditions, is 50-60% with low rank coals and 40-50% with bituminous coals. According to Williams et al. [20] char-N amounts to about 80% of the total NO formed. Rostam-Abadi et al. [21] give a lower amount of 20-30% and state that the amount is dependent on the temperature and the amount of devolatilization.

The type of nitrogen species released during the devolatilization process is determined by the structure of nitrogen in the coal. Coal nitrogen is believed to occur mainly in aromatic ring structures. At high temperatures (>1500 K), 70-90% of the coal nitrogen is devolatilized with 0-20% of the coal nitrogen evolving in the early volatiles, mainly as HCN and NH_3. The tars and nitrogen bound in aromatic rings are the main source of HCN formation while the amines in the coal mainly form NH_3. Combustion of bituminous coals show more HCN formation than NH_3 and sub-bituminous and lignite coals form more NH_3. This is probably because the number of aromatic rings decreases with coal rank but the number of cyclic compounds increases. Accordingly, in low-rank coals the conversion of fuel nitrogen to NH_3 increases [16]. In another study by Schafer and Bonn [22] it was established that the concentration of NH_3 is usually higher than that of HCN in the combustion chamber. These larger amounts of NH_3 cannot be explained only by conversion of the coal-amines but most probably are formed from hydrolysis of HCN. Knowledge of the HCN and NH_3 concentrations is very important since these are the main precursors of nitrogen oxides.

It must be stressed again that the temperature, residence time and fuel/oxygen ratio are main factors in determining the species formed. The three main mechanisms for NO_x formation in combustion systems are [23]: (1) Thermal NO_x - the reaction mechanism that occurs at high temperatures between atmospheric nitrogen and oxygen, (2) Prompt NO_x - the attack of a hydrocarbon free radical on molecular nitrogen producing NO precursors, and (3) Fuel NO_x - the oxidation of the fuel-nitrogen released during devolatilization and char oxidation. Fuel NO is the main source of NO_x in coal-fired systems and accounts for 60-80% of the total NO_x formed [24].

Fuel-N is assumed to proceed through HCN and/or NH_3 [16]. When the fuel-N is bound in an aromatic ring, HCN is formed. NH_3 is formed from amines. Figure 5 presents the reaction paths for fuel NO_x. Nitrogen in the volatile matter is released during the devolatilization process. Part of the nitrogen (α) is quickly transformed to HCN and the remaining part (1-α) of the fuel nitrogen is transformed to NH_3. Depending on local conditions, these two species will react to form either NO or N_2. They will be reduced to N_2 in fuel-rich regions, and in fuel-lean regions they are generally oxidized to NO_x. The NO_x formed can also be reduced via heterogeneous reactions with the char particles.
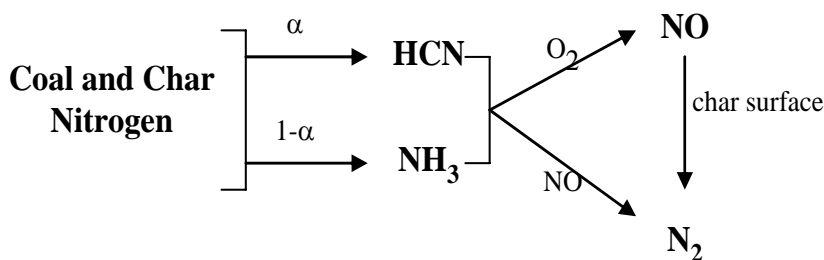
Figure 5. Schematic representation of fuel NOx formation and reduction process [modified from Hill and Smoot [16]].

The GLACIER code used in this work, as most coal combustion CFD models, separates the mechanisms of NO_x formation from the generalized combustion model. The models of nitrogen pollutants are executed after the flame structure has been predicted. The basis for this is that the flame structure is governed by much faster fuel-oxidizer reactions and is not affected by the formation of trace pollutant species. An advantage of this approach is computational efficiency. The time required to solve the system of equations for the fuel combustion can require many hours of computer time while the pollutant sub-models typically make up a small part (~10%) of the time required to converge the combustion case. Thus, NO_x sub-model parameters and pollutant formation mechanisms can be better investigated by solving the NO_x sub-model using a pre-calculated flame structure [16].

GLACIER cannot compute the full chemistry for all intermediate, including nitrogen, species. However, this detail is not required for engineering solutions for utility boiler combustion chambers. As long as all physical mechanisms of first-order importance are included in the investigation, engineering modifications to the furnace can be simulated using a condensed set of chemical kinetic mechanisms [8]. For NO_x formation these are governed by three variables that are used for NO_x postprocessor simulation only. These parameters, described therein, are varied by the user at the beginning of each NO_x simulation and are inputted into the CFD code. The variables used for the NO_x postprocessor are named VMNFR, ZEDA and ZEDAH. The following description was summarized from the GLACIER manual [Reaction Engineering International (REI), "Configured Fireside Simulator IEC 550 MW Wall-Fired Unit" (2004)].

VMNFR is the parameter that defines the division of nitrogen between the volatile matter and the char, and it is experimentally derived as described in Chapter 2.2. This split is potentially important for NO_x formation. The default value for VMNFR is 0.5; equal mass fractions of nitrogen in the volatile matter and the char. The absolute amount of nitrogen can be different in volatile matter and char and is dependent on the volatile yield. For VMNFR = 1.0, all coal-N is in the volatile matter. For VMNFR = 0, all coal-N is in the char. If X symbolizes the volatile matter yield expressed as a fraction of total coal organic mass, for $0.0 \leq VMNFR < 0.5$ there is less coal nitrogen partitioned into volatile matter nitrogen (X*(VMNFR/0.5)) than into char nitrogen (1-(X*(VMNFR/0.5))) relative to equal mass fractions of nitrogen in the volatile matter and char (VMNFR=0.5). For $0.5 < VMNFR \leq 1.0$ there is more coal nitrogen partitioned into volatile matter nitrogen (X+((VMNFR-0.5)/0.5)*(1-X)) than into char nitrogen (1-(X+((VMNFR-0.5/0.5)*(1-X))) relative to equal mass fraction of nitrogen in the volatile matter and char (VMNFR=0.5).

ZEDA is the parameter that partitions the formation of volatile-N between HCN and NH_3. Knowledge of the HCN and NH_3 concentrations is very important since they are the main precursors of nitrogen oxides. ZEDA is defined by:

$$ZEDA = \frac{mol_{HCN}}{mol_{HCN} + mol_{NH_3}} \qquad (6)$$

The default value for ZEDA is 0.8; meaning 80% of the volatile-N forms HCN. ZEDAH is the parameter that defines the fraction of char-N that is converted to NO_x. The CFD default value is 0.1; meaning 10% of the char-N is converted to NO_x. It should be noted that this value is low compared to the literature [19-21].

Since the formation of NO_x is so closely coupled with operation conditions: fuel/air stoichiometry, temperature, heating rate and residence time, the NO_x postprocessor parameters are likely to change with different operating conditions. This differs from the volatiles and char oxidation kinetic parameters which are kept constant for each coal.

Summary

Errors in the model parameters affect combustion characteristics such as coal burnout, temperature fields, flame structure and NO_x formation. Lockwood et al. [25] carried out a comprehensive sensitivity study on their test furnace on the influence of different model parameters on combustion performance. They show that although the volatile yield did not have much of an effect, devolatilization kinetics influence is strongly felt in the near burner zone and volatiles combustion rates strongly influence oxygen levels for their furnace length. Char oxidation kinetics govern the burnout rate. Gera et al. [26] showed that deviating the 1-step devolatilization activation energy by ±12.5% moves the flame root position (defined as the distance from burner where flame temperature reaches 1000 K) in the range of l/D (length/diameter)= 0.69 to 1.04. Jones et al. [13] demonstrate that the devolatilization rate has a marked influence on predicted NO_x emissions. Sheng et al. [10] found that the total coal nitrogen content and volatile yields influence the levels of NO_x formation. Kurose at al. [27] demonstrated that taking into account NO formation from char-N improved their prediction results.

To summarize, kinetic data required for running GLACIER are: nitrogen distribution between volatile matter and coal-char (VMNFR), six parameters for the two-step devolatilization mechanism ($Y_{1,2}$, $A_{1,2}$, $E_{1,2}$), three parameters for the one-step coal oxidation process (n, A_c, E_c), a set of kinetic parameters (plugged-in the code) for a comprehensive gas-phase NO_x mechanism starting from HCN-NH_3 with an initial ratio (ZEDA), one parameter for the conversion of char-nitrogen into NO (ZEDAH); a total of twelve parameters, that none of them could be found in the literature. All of these parameters which describe the coal particles' combustion and NO_x formation rates can be varied in the GLACIER program to obtain the best agreement with the experimental data.

The approach described here determines a simplified kinetics model. While it cannot be said that the kinetic parameters used in the model will correspond to the combustion rate of a single particle of coal, these parameters do describe the combustion behavior of a "macroscopic" sample of the tested coal. For the goal of predicting the combustion behavior

of these coals and blends in a utility boiler furnace, this simplified model used with the GLACIER code gave good results.

From the previous chapter one understands the complexity of modeling coal combustion. The CFD code used in this study, GLACIER, allows the user to define twelve different parameters that influence the different combustion species, combustion rates and temperatures. These parameters are varied in the GLACIER code to obtain the best agreement with the test furnace experimental data for further input in the GLACIER code of full-scale utility boilers.

2.2. Test Facility

The core of the facility is the 50 kW cylindrical down-fired test furnace. As the foremost function of the furnace is to extract kinetic parameters for the large-scale boiler simulations, plug-flow conditions in the furnace are essential, i.e. velocity, concentration and temperature are uniform throughout the radial coordinate. Plug-flow conditions for most of the furnace length were achieved by two means using: (1) a high-swirling burner at the top of furnace to ensure immediate and full mixing of fuel and air and (2) a relatively high length(L)/diameter(D) ratio (L=4.5 m and D=0.2 m; L/D=22.5). To recreate utility boiler kinetic conditions as much as possible, the test furnace works at temperatures up to 1650 °C, corresponding to the temperatures at pulverized-coal facilities. It should be emphasized that the test facility does not simulate the flow dynamics in the utility boiler and that the test facility cannot directly predict the combustion behavior in utility boilers. Coal combustion processes in test facilities, such as this one, are simplified because many of the physical and chemical processes that take place in a utility boiler are absent. Therefore, it is easier to validate CFD combustion models for the test facility and apply the validated model, with the same kinetic parameters but with different boundary conditions, to a utility boiler.

Process Design

Figure 6 describes the test facility's process scheme. The system is designed to burn solid, liquid and gaseous fuels and can burn one or two types of pulverized coals and different ratios of binary blends. Different stoichiometric ratios of air-fuel can be determined; air and fuel staging are done. The hot combustion gases are cooled before being emitted into the atmosphere and the particulate matter is separated using a bag filter. The whole process is semi-automatic and controlled by a fully computerized control system using about fifty strategically placed sensors.

Figure 6. Test facility process scheme. T, P, F, W, L denote measurements (manual and automatic) of temperature (°C), pressure (bar), air and fuel flow (air, F=Nm^3/h), fuel oil (F=L/h), LPG (F=L/min), coal (W=kg/h)) and level of fuel oil in tank respectively.

Coal Preparation and Feeder
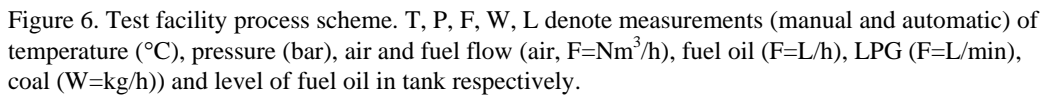
Pulverized coal is prepared on site from raw coal with a size distribution of about 40% smaller than a 5 mm diameter, depending on the coal. The coal is air dried for 24 hours, crushed with Mini Crusher 8F and then pulverized with Laboratory Pulverizer LC67, both from Gilson Inc., USA. Normally the pulverized coal particle size distribution is 60-80% smaller than 75 micron, but this can be changed as needed for the combustion experiments. The coal dosing structure, shown in figure 7, supplies the pulverized coal to the primary air line. It was planned and built specifically to meet the requirements of the test facility. The coal dosing system has the capability of supplying a homogeneous mixture of two types of pulverized coal at a rate of 2-8 kg/h. The coal is gravitationally transferred directly from the drum to the loading hoppers which can each hold a different type of coal. A vibrator periodically shakes the loading hopper to eliminate the formation of bridges or clumps. When the controller relays an order, coal is dropped from the loading hopper to the weighing hopper. An agitator is installed in the weighing hopper to mix the coal and break up lumps. After the coals are weighed, they are transferred to a single outlet and are released to the primary air-line via an ejector.
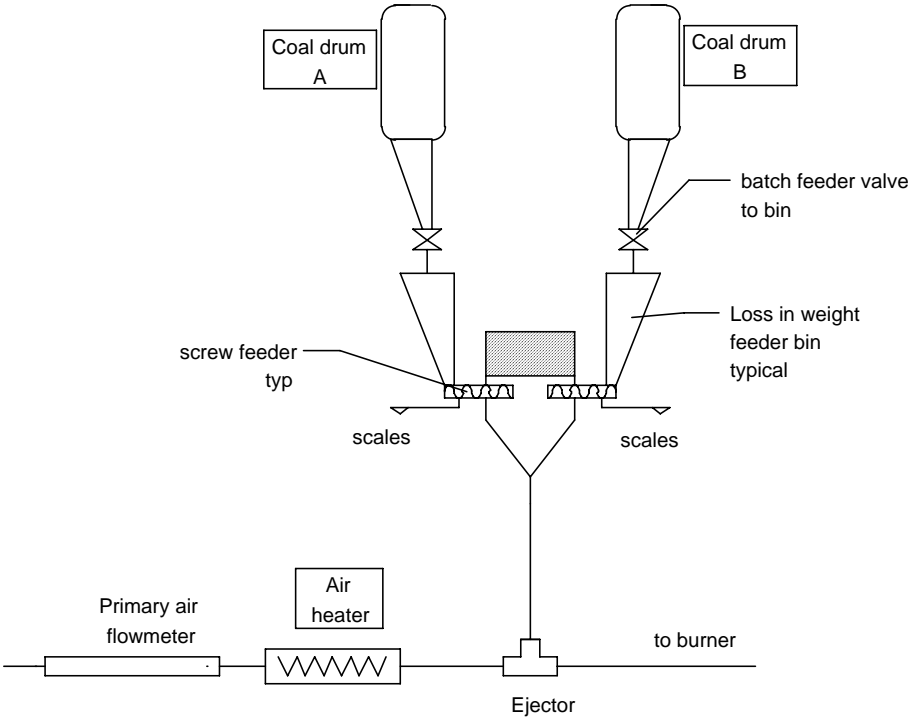
Figure 7. Coal feeder system.

Burner

A highly-swirling burner was designed to create an intense and short mixing region. As shown below, the burner is located at the top of the furnace. Figure 8a shows the cross section of the burner and figure 8b shows the burner face.
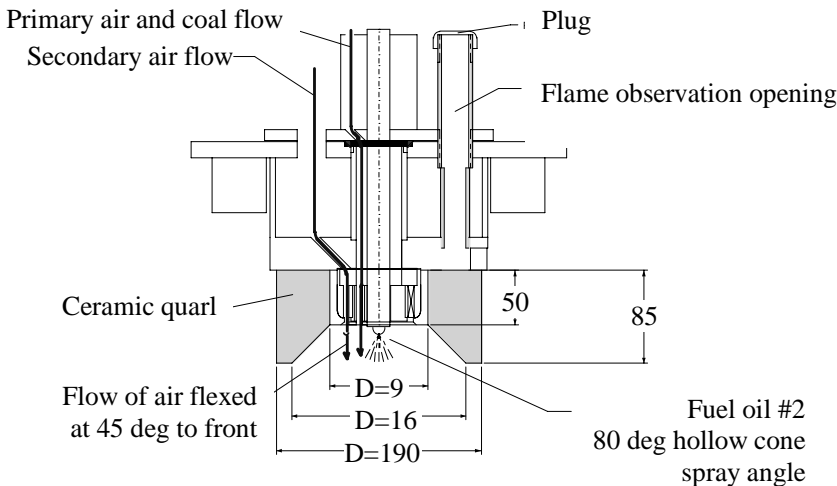
Figure 8a. Cross section of the burner.

Fuel oil #2 nozzle orifice

9 x ID6mm

Primary Air and Coal
Inlet
Area = 254 mm²

Circle
D=35mm

Secondary Air Inlet
OD57, ID47
Area = 376 mm²
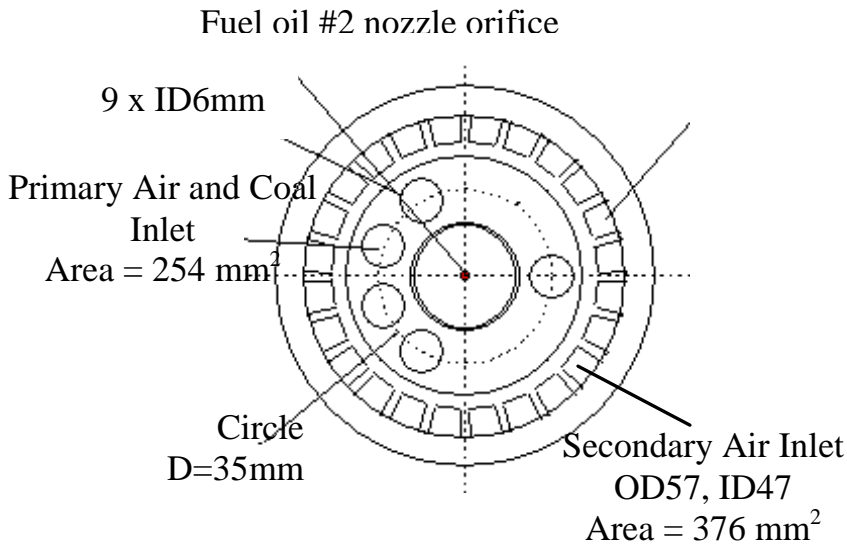
Figure 8b. Burner face.

The burner is made from stainless steel and is insulated from the heat of the furnace by a quarl shaped protective shield. Fuel oil or liquid petroleum gas (lpg) can be injected through an appropriate nozzle located in the center of the burner. Surrounding the nozzle is an open ring (d=0.020/0.022 m) for air flow, whose purpose is to reduce recirculation in the central zone of the burner exit when burning fuel oil. Without this air flow, dendrites of soot form on the burner center. Surrounding this ring are nine holes through which the pulverized coal and primary air mixture enter the furnace. The total area of the coal-air inlet is 2.54E-04 m² while each hole has an ID of 0.006 m. The holes are located on an imaginary circumference of 0.035 m. Secondary air is introduced via twenty-four equal rectangular holes, with the dimensions of 0.0055x0.0035 m, total area 3.76E-04 m². These holes are arranged circularly with an OD (outer diameter) of 0.057 m and ID (inner diameter) of 0.047 m and they are machined at 45° with the furnace centerline, to cause a swirling motion.

Figure 9 shows the axial component of the velocity vector (the velocity in the direction of the furnace length) obtained by the CFD simulations done with the GLACIER code. Because of the grid layout, the vectors are denser in the center of the furnace. As expected from the burner design, the flow is complex with recirculation zones near the burner, but by 0.3 m from the burner face, at l/D (length/diameter) = 1.5, the axial flow pattern is quite uniform for the length of the furnace. In the near-burner zone one can see the recirculation of the flow along the sides of the furnace near the quarl and in the center, back towards the burner face.
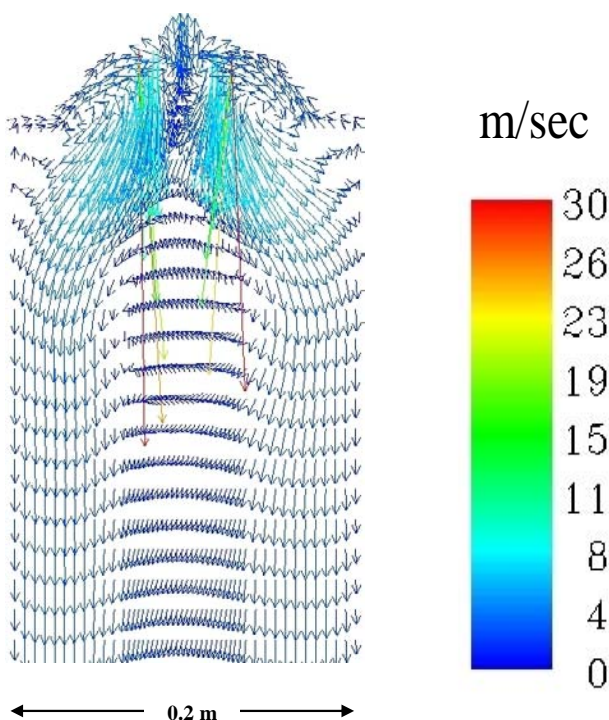
Figure 9. Axial vector presentation of the velocity in the near-burner zone (l/d = 0-2).

Test furnace

Figure 10 shows the test furnace, which dimensions are 0.2 m id (inner diameter) and 4.5 m high. The combination of a high-swirling burner, small furnace diameter and long furnace length creates a short mixing region followed by a plug-flow in the rest of the furnace. This will be explained in greater detail after the furnace description. The furnace can work up to temperatures of 1650 °C, and has a residence time of 1-2 sec. These correspond to the temperatures at pulverized-coal facilities and the time needed for complete coal combustion. The furnace is made up of 8 modular sections, each around 0.6 m high. The cast ceramic center is surrounded by four insulation boards to minimize heat losses. The outer shell temperature decreases monotonically from 85 °C (top furnace) to 45 °C (bottom), depending on the furnace load. For safety and stable flow considerations the furnace is maintained at a slightly sub-atmospheric pressure (about 0.5 mbar below 1 atm). The primary air is heated to about 110 °C with a flow rate around 12 Nm^3/hr, which is the minimum rate required to prevent the coal from settling in the pipe. The coal and primary air mixture enters the burner at a temperature of 60-70 °C. The secondary air is heated to 250 °C, with flow rate depending on the feed rate of the fuel and stoichiometric ratio. Twenty, 0.04 m id, sampling openings were drilled on both sides of the walls of the furnace for gas and temperature sampling, and other input and output applications. Overfire air (OFA) is added via 0.008 m id openings at l/D=13.83. LPG injection is through 0.0006 m id openings at l/D=4.83, 7.83 or 10.83.
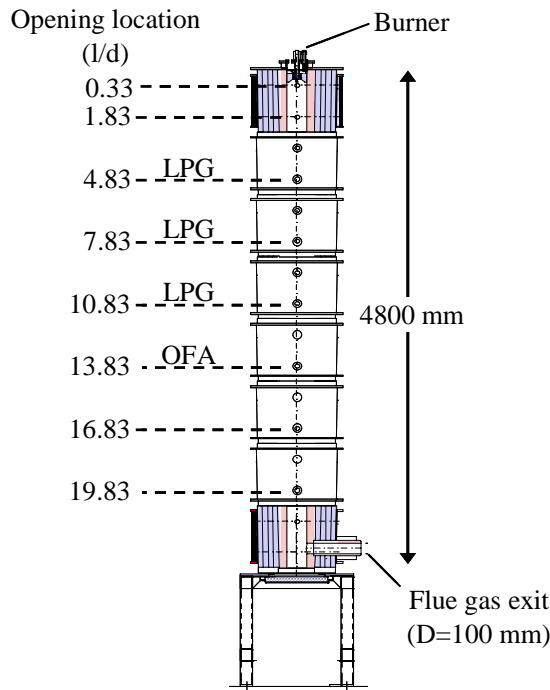
Figure 10. Furnace – general view and sampling, LPG and OFA openings locations.

The specific geometry (2-D with small ID) of the furnace serves two purposes: (1) simplify the CFD simulations, (2) create a short mixing region followed by a plug-flow for most of the furnace length. The high-swirling burner causes a short and yet very intense mixing region. In this mixing region, under certain conditions, both volatile matter and char are burned. For the rest of the furnace chemical reactions are minimal unless reburning fuel or OFA were added. This two-zone configuration was obtained both from simulation and experiment, as shown in figures 11-15. The lengths of these two zones were varied by changing different combustion process parameters, such as feed particle size and stoichiometric ratio.

Figure 11 shows two-dimensional presentation of the predicted temperature within the test furnace (125 ^{o}C between each isotherm-line), distributed in five furnace segments (0.6m for each segment, or 3 l/D). The temperature is not uniform near the burner, the intense mixing zone, and subsequently, by l/D=4, or about 1 m from the burner face, the temperature is homogeneous for the rest of the furnace length. Predicted oxygen levels, depicted in figure 12, show similar trends. Figure 13 portrays the axial velocity component (the velocity in the direction of the flow) distribution in the furnace. The flow pattern is varied near the burner but by the middle of the first segment, l/D=1.5, the flow pattern is uniform and stays that way for the length of the furnace. The radial profile of the velocity shows higher speed in the furnace center, and almost no movement near the furnace walls. Since the burner causes high-swirling, there is radial velocity near the furnace walls, which is not depicted by this figure.
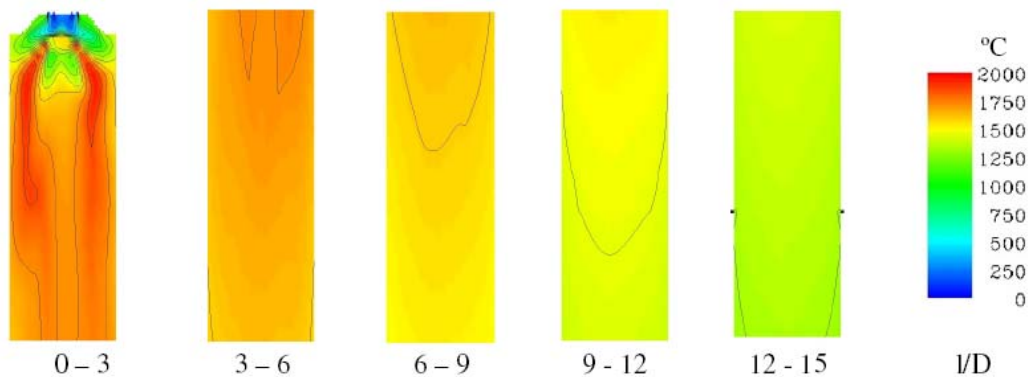
Figure 11. Two-dimensional presentation of the temperature within the test furnace, with isotherms (125 °C between each line), distributed in 5 segments for clarity, each segment represents 3 l/D.
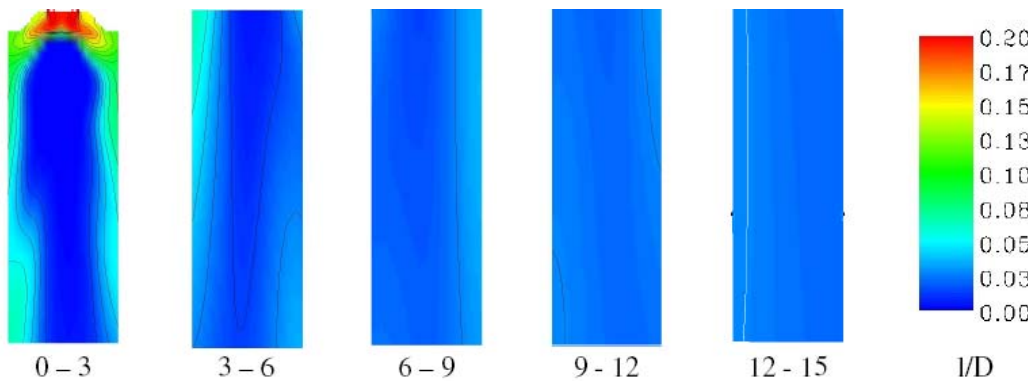
Figure 12. Two-dimensional presentation of oxygen levels within the test furnace distributed in 5 segments for clarity, each segment represents 3 l/D.
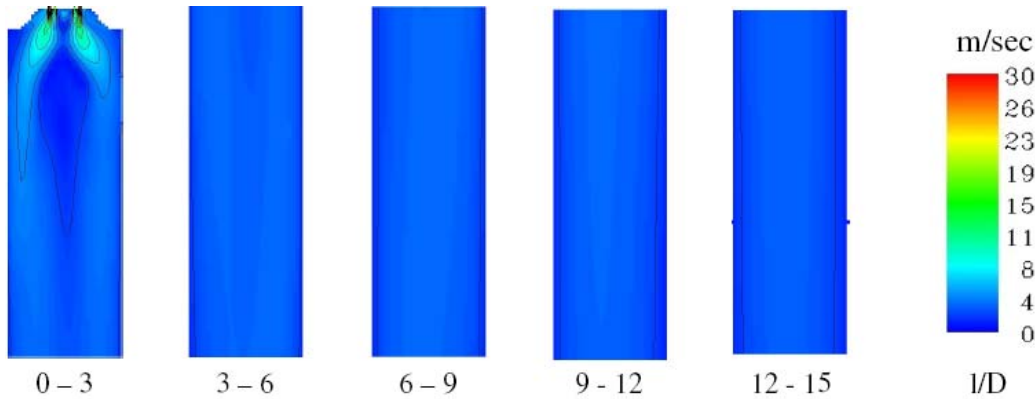
Figure 13. Two-dimensional presentation of the axial velocity within the test furnace distributed in 5 segments for clarity, each segment represents 3 l/D.

The two-zones, high-mixing and plug flow, were also experimentally observed. Figure 14 shows radial profiles of temperature and gas concentrations at l/D=0.33 and 1.83 for three different coals: KPC (bituminous), Drummond (bituminous) and Adaro (sub-bituminous). The temperatures and gas concentrations are not uniform at these locations, which is the area of the high-mixing zone. On the other hand, figure 15 shows the temperature and gases in the plug-flow region by depicting temperature and CO_2 at l/D=4.83, 10.83 and 13.83. These are quite uniform. It should be noted again that both simulations and measurements show the two-zone configuration of the furnace that is essential for the extraction of kinetic parameters.
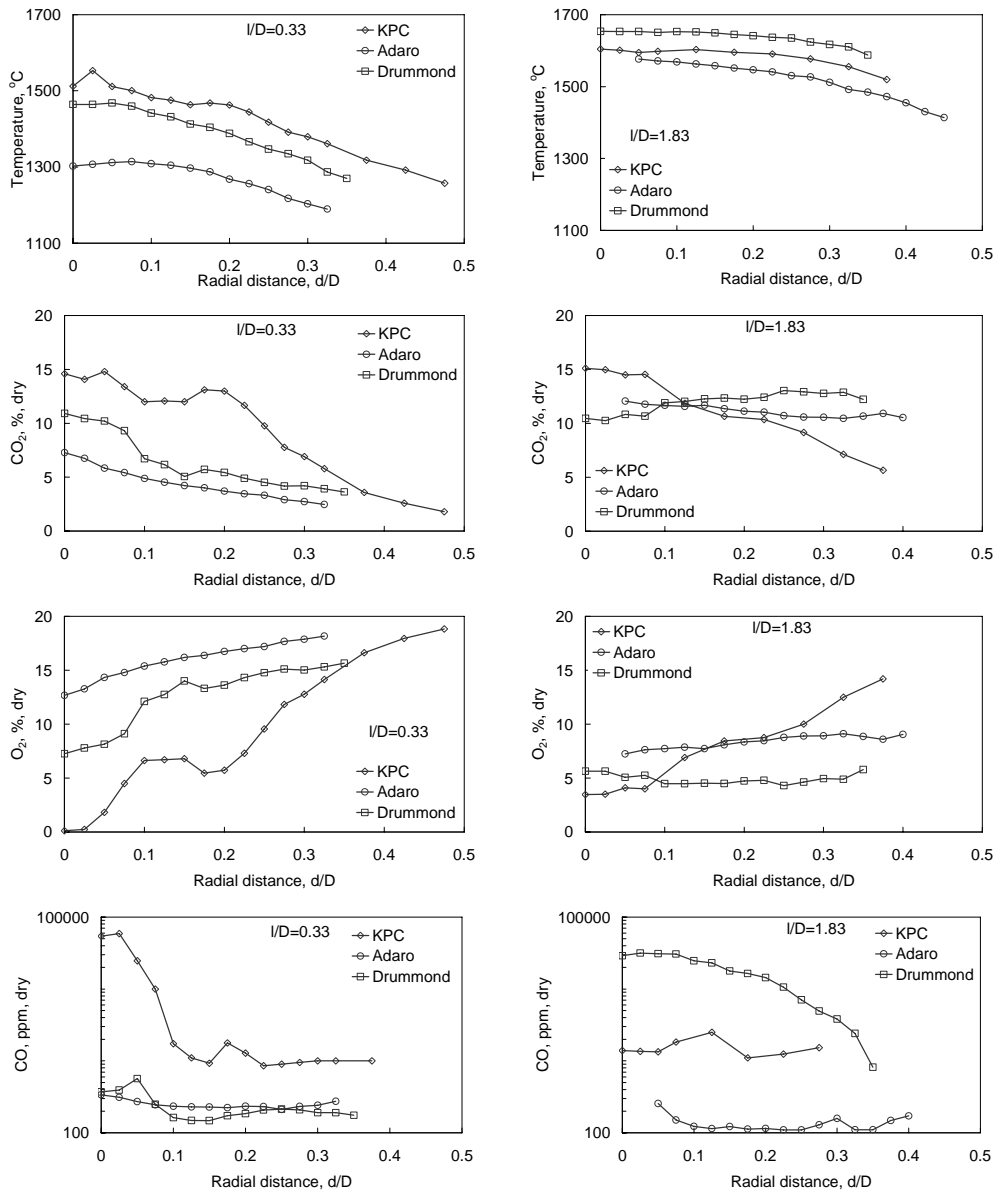
Figure 14 continued on next page.
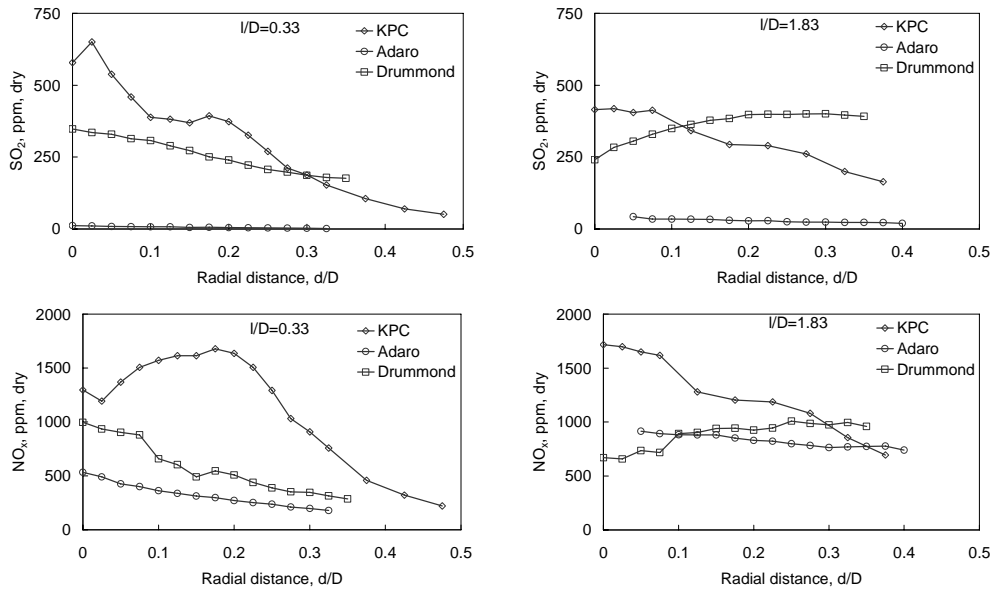
Figure 14. Measured radial profiles of temperature and gas concentrations in the high-mixing zone at l/D=0.33 and 1.83 for KPC, Adaro and Drummond coals.
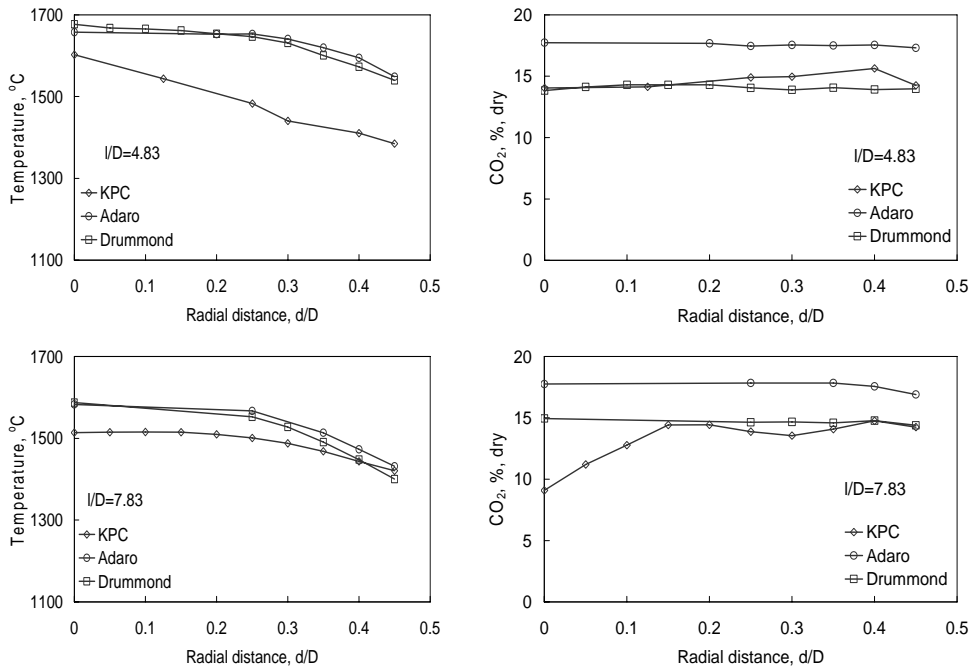
Figure 15 continued on next page.
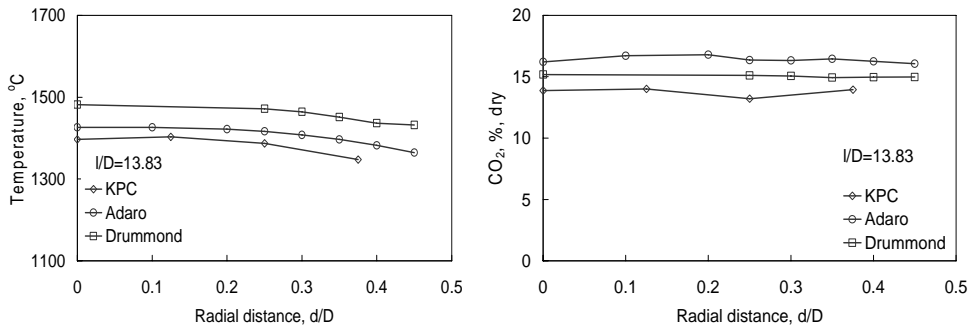
Figure 15. Measured radial profiles of temperature and CO_2 concentrations in the plug-flow zone at l/D=4.83, 7.83 and 13.83 for KPC, Adaro and Drummond coals. CO_2 represents the behavior of all other gases measured in the furnace.

Sampling and Analysis

The temperature and gas concentration monitoring in the test furnace is done semi-automatically. The probes are moved manually but the data is collected and analyzed electronically and stored by the control and data acquisition system. Care was taken to ensure that both temperature and gas concentrations results were reproducible by periodically checking the stability of the test facility and the measuring equipment.

Temperature

Temperature measurement in combustion systems is a very complex issue, especially for coal combustion, where particles are present. Particle and gas temperatures are measured separately. Particle temperatures were not measured in this work. Gas temperature was monitored with thermocouple probes. These include bare fine wire thermocouples and aspirated thermocouples, like the suction pyrometer. The main advantages of bare fine wire thermocouples are that the measurements have the high precision characteristic of electrical measurements; the dimensions are small so that flame disturbance is minimal, and the method is simple and inexpensive.

However, there are several sources of error when using thermocouples in coal combustion, mostly radiation effects. The thermocouple does not filter out the radiative heat transfer from the combustor walls and conduction heat transfer along the leads. Radiative heat transfer between the particles and the bead, and coating of the bead by particles, are sources of error when particles are present. Therefore, the thermocouple will not obtain the same temperature as the hot gas. In suction pyrometers, refractory shields minimize the effects of radiant heat by shielding the thermocouple from its surroundings. The suction pyrometer's drawbacks are its' bulky size and its' requirement of high suction speed (200 m s^{-1}), which can disturb the flow, but mainly that it clogs up very easily because of its small diameter. To solve this problem, the probe was cleaned every 10-15 min or, as needed, with a pulse of pressurized air. This prevented obstruction of the probe but caused cracking of the ceramic (Alsint >99% silica) shield. It was decided to calibrate the thermocouple connected to the gas probe with the suction pyrometer and to use this temperature measurement on a regular basis. As expected, a strong linear correlation between the two probes was found (see figure 16). A

literature correlation from Spinti [28] was also used for this calibration. TypeB (0-1700 °C), 0.50 mm diameter, thermocouples were used.
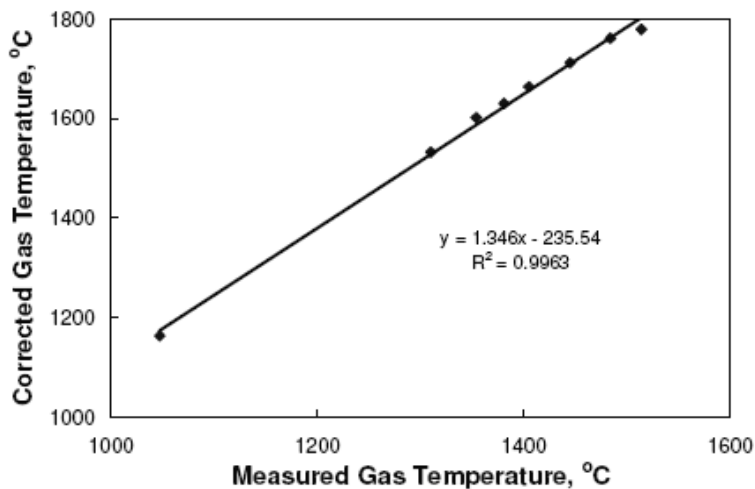
Figure 16. Correlation between corrected temperature (using suction pyrometer and literature correlation) vs. measured gas temperature with gas analyzer probe.

Combustion Gases

The main components of the system for sampling and analyzing gases are (1) gas sampling probes, (2) heated gas sampling tube, particulate filter unit and water condenser, (3) NO_2 to NO converter, and (5) Gas analyzers (CO, CO_2, NO_x, SO_2 and O_2). The ideal gas probe provides a sample that is representative of the gas species' concentrations at the point in the combustor where the probe tip is located. The designs of most sampling probes are based on convective cooling, water or gas quenching or aerodynamic quenching techniques. The quenching of gas samples extracted from combustion systems is essential for the prevention of chemical reactions occurring in the probe or intermediate lines that may significantly change species concentrations. Figure 17 shows the design of the water-cooled probe constructed for this project.
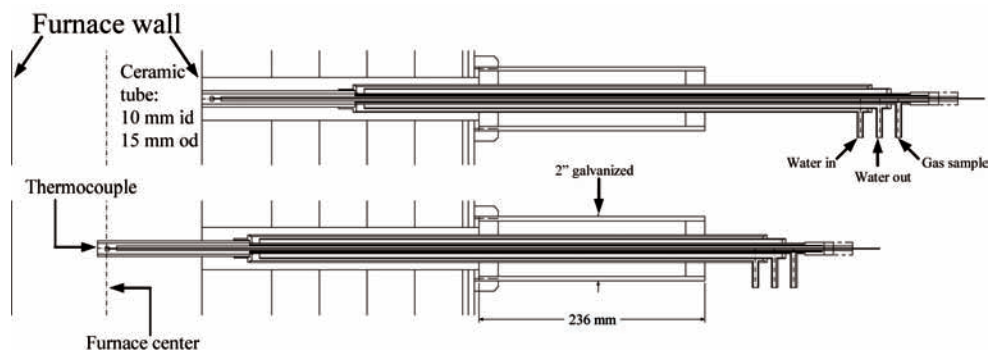
Figure 17. Probe (water-cooled) for temperature measurement and combustion gas extraction.

The sampling tube is Teflon made 7-meter long, 5mm ID, with flow range of 30-300 L h[1]. The tube is electrically heated to 130 °C to prevent water condensation in the pipeline; this is to prevent gas absorption and chemical reactions between the gases and water, which will subsequently lower the gas concentrations. Once the gas reaches the analytical unit, it is cooled and the condensate is separated and pumped out of the unit. A ceramic filter, installed between the gas probe and the analyzer system cabinet, traps the particulates. The capacity of the filter is 0.010 kg m^{-3} and its' retention rate is 99.99% for particles >5μm. The filter is checked and cleaned periodically with compressed air. This cleaning is important, not only to prevent clogging, but also to lower, as much as possible, the absorption of gas on the particulate matter.

The combustion gases analyzers system is installed in a closed cabinet and works in the temperature range (–20)-50 °C. Analyzer calibration is done on a daily basis with built-in permeation tubes. Once a year, calibration is done by an external laboratory.

Monitoring is done with three gas analyzers that monitor the following components in the furnace: (a) CO, CO_2, NO and SO_2 with Infrared Analyzer Module Uras 14 (two analyzers). (b) O_2 Analyzer Module Magnos 106. The Uras 14 uses the fact that different types of heteroatomic molecules absorb infrared radiation at specific wavelengths and utilizes this principle of non-dispersive infrared radiation (2.5-8 μm wavelength range). Each instrument measures two components (CO/CO_2 and NO/SO_2). Table 1 summarizes the measurement parameters for each gas. Since the analyzer cannot recognize NO_2, in order to get an analysis of NO_x ($NO+NO_2$) the gas sample needs to be diverted to a converter, which reduces the NO_2 in the gas sample to NO. The conversion is performed in a reaction tube filled with a catalyst (molybdenum) and heated by the tubular furnace. If the gas sample is not put through the converter then the analyzer will measure only the NO concentration in the sample. The Magnos 106 measures oxygen concentration by utilizing oxygen's paramagnetic behavior. Table 2 summarizes the oxygen measurement parameters.

Table 1. Measurement parameters of CO, CO_2, NO and SO_2 (two analyzers)

Component Parameter	CO	CO_2	NO	SO_2
Range	0-1%, 0-10%	0-30%	0-2,500 ppm	0-10,000 ppm
Linearity deviation	$\leq$ 1% of span	$\leq$ 1% of span	$\leq$ 1% of span	$\leq$ 1% of span
Repeatability	$\leq$ 0.5% of span	$\leq$ 0.5% of span	$\leq$ 0.5% of span	$\leq$ 0.5% of span
Zero drift	$\leq$ 3% of span/week	$\leq$ 3% of span/week	$\leq$ 3% of span/week	$\leq$ 3% of span/week
Sensitivity drift	$\leq$ 1% of measured value per week	$\leq$ 1% of measured value per week	$\leq$ 1% of measured value per week	$\leq$ 1% of measured value per week
Output signal variations	5 sec	5 sec	5 sec	5 sec
Detection limit	$\leq$ 0.5% of span	$\leq$ 0.5% of span	$\leq$ 0.5% of span	$\leq$ 0.5% of span

Table 2. Measurement parameters of O$_2$ (one analyzer)

Range	0-25 Vol.-% O$_2$
Linearity deviation	$\leq 0.5\%$ of span
Repeatability	$\leq 0.5\%$ of span
Zero drift	$\leq 0.03\%$ of span per week
Sensitivity drift	$\leq 1\%$ of measured value per week
Output signal variations	3 sec
Detection limit	$\leq 0.5\%$ of span

Heat Flux

Heat flux is created by a temperature gradient and is the rate of energy transfer through a given surface. This quantity can be measured using a heat flux sensor. Generally, a heat flux sensor is made up from a plate with a differential temperature sensor connected to the top and bottom, at different temperatures. The thermocouple generates an output voltage that is proportional to the temperature difference between the hot- and cold sides. Assuming that the thermal conductivity of the plate is constant and the flow is static, the heat flux is proportional to the measured temperature difference. Figure 18 shows the heat flux probe. The potential difference between the hot and cold leads of the thermocouple (TC) is measured. This potential is translated to heat flux by the equation given in figure 19. The probe was calibrated with a calibrated IEC probe to obtain the graph shown in figure 19. The central thermocouple temperature measurement is important for proper heat flux measurement, because the heat flux is temperature dependent.
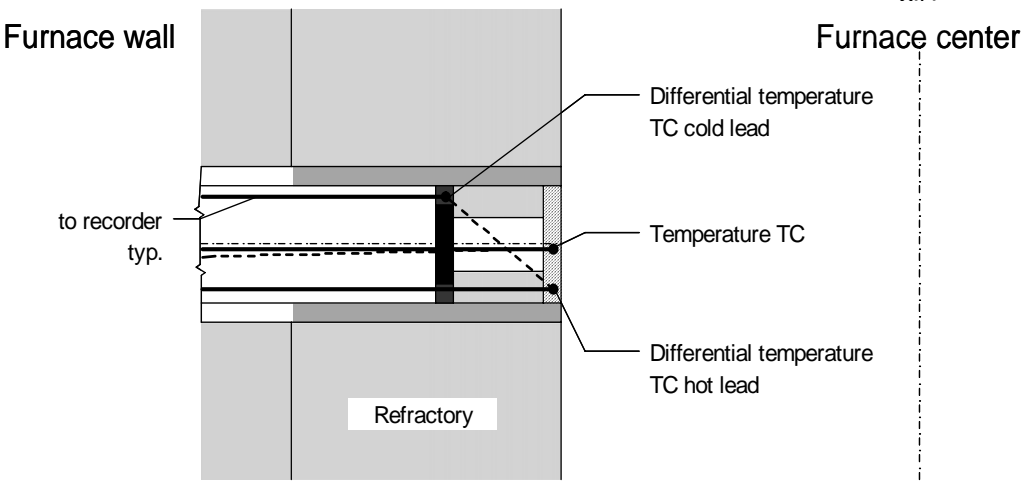
Figure 18. Heat flux probe layout.
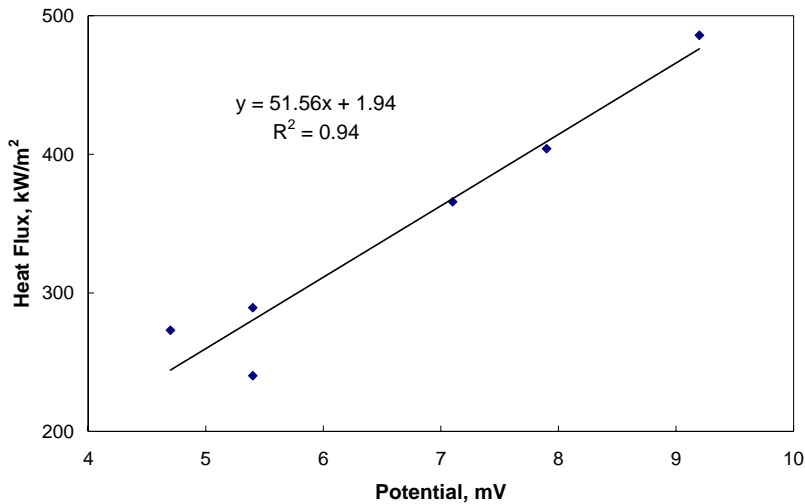
Figure 19. Heat flux probe calibration and equation for potential/heat flux transformation.

LOI (Loss-on-Ignition)

Loss-on-ignition, LOI, is the amount of carbon in the combustion particulate residue (mostly ash). Obviously, low levels of LOI are the goal of the utility boiler operator. LOI levels of 2-5% in the combustion chamber are considered normal and vary according to the coal type. The carbon content and particle size distribution of the unburned coal particles are analyzed by ASTM (American Standard Test Method) D5142-90(1994): *Standard Test Methods for Proximate Analysis of the Analysis Sample of Coal and Coke by Instrumental Procedures*. The burnt coal particles are sampled by US-EPA Method 5: *Determination of Particulate Matter (PM) emissions from Stationary Sources*, using the sampling configuration shown in figure 20.
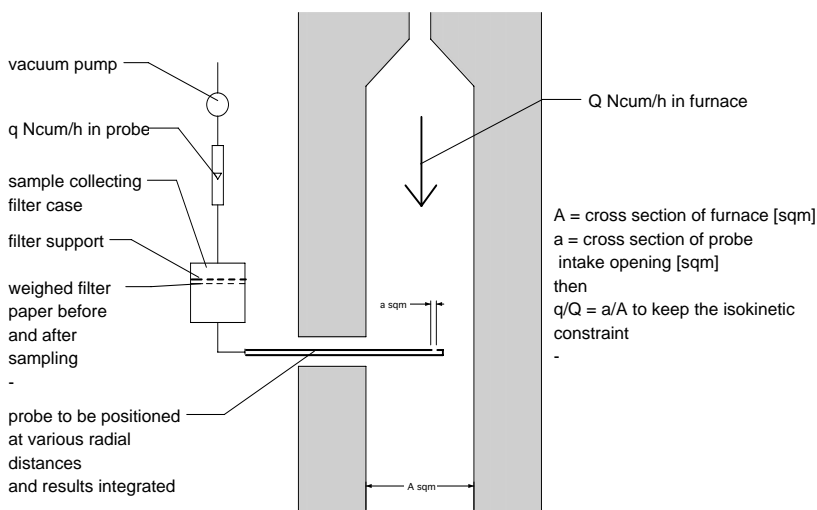
Figure 20. Isokinetic LOI sampling probe.

Slagging

Slagging is the term for the phenomenon of the accumulation of ash on the walls of the combustion chamber. The tendency of a coal to cause ash deposition is affected by the components of the coal ash, as well as the temperature of the ash particles as they come in contact with a surface. Slagging will change the heat transfer properties of the walls of the combustion chamber. These changes are monitored by the slagging probe for the different coals tested in the test furnace. The slagging probe has the same dimensions as the temperature/gas probe shown in figure 17 except that the part placed in the furnace is closed. A thermocouple was welded on the top of the probe to measure the changes in temperature. The temperature differential between the outer and inner parts of the probe is translated to heat flux. Measurements are made at l/D=7.83 for 6-8 hours. A plot depicting the time vs. the decay of heat transfer as deposition occurs is drawn. The results for different coals are compared qualitatively only. It should be noted that GLACIER does not have the capability to predict slagging behavior.

Coal/Char Nitrogen

For the GLACIER prediction of NO_x formation, the distribution of elemental nitrogen (N) between the volatile matter and the char needs to be known. Total nitrogen concentration in the coal and the nitrogen concentration in the coal char were analyzed using ASTM D3179-89 (1993): *Standard Test Methods for Nitrogen in the Analysis Sample of Coal and Coke*. Coal char samples were prepared by heating a 5 g sample of pulverized coal to 950 $^{\circ}$C at a rate of 3 $^{\circ}$C min^{-1}, kept at peak temperature for 6 min and then cooled at a rate of 3 $^{\circ}$C min^{-1}.

2.3. Artificial Neural Network

The whole fitting process required many simulation runs, 25-30, that were carried out with discrete values of the model parameters. Since each simulation took 1-2 days until convergence, we developed an artificial neural network (ANN) [29-31] to reduce computational time and improve efficiency.

For the past 20 years, the potential of ANN as a universal modeling tool has been widely recognized. ANNs can manage complex and non-linear problems, processing information rapidly and thus reducing the engineering effort required in model development. Since ANNs approximate random non-linear functions and they do not need previous knowledge of the process phenomena, ANNs offer an alternative approach to model process behavior. They learn by extracting patterns from data that portray the relationship between inputs and outputs in any given process phenomenon. Neural networks use a trial and error method of learning. Training a neural network involves using a database of examples, which inputs and outputs are the values for the input and output of the network. An input is any information used to arrive at a solution, prediction or decision. An output is the solution, prediction or decision the ANN is being trained to generate. The neural network establishes the relationship between the inputs and the outputs by calculating their relative importance (weights). It calculates and compares its results to the actual example output. The network learns by adjusting the weights to minimize the output errors. In its basic form, a neural network is composed of several layers of neurons, an input layer, one or more hidden layers and an output layer. Each layer of

neurons receives its input from the previous layer or from the network input. The output of each neuron feeds the next layer or the output of the network.

For this work, a feed-forward neural network has been trained to predict the kinetic model parameters of the coal combustion. The ANN was developed using MATLAB$^{\circledR}$ and Simulink$^{\circledR}$ (developed and marketed by The MathWorks, Inc.). The network was trained on the basis of modeling results for three coals: KPC (bituminous), Adaro (sub-bituminous) and KPC-Melawan (sub-bituminous). The input consists of 180 data sets, 60 for each coal. Each data set is comprised of 18 elements: 6 parameters (X – downstream distance from burner, T – gas temperature, NO_x concentration, O_2 concentration, coal index, and N – coal nitrogen content) for 3 different downstream (X) locations. The output data set also contains 18 elements. The neural network was trained using Levenberg–Marquardt optimization algorithm together with a cross-validation based on 'early stopping' mechanism to prevent over-fitting. During network training, the training algorithm continuously checks the network error on the test data set. The training was terminated at the point where the network error on the test data was a minimum. Early stopping is an implicit way to implement regularization, which can improve model robustness [29]. The training strategy implemented has the advantages of speed and not over-fitting the noise in the data.

2.4. GLACIER CFD Code

For the last two decades, much effort was made to develop CFD models to simulate coal combustion in large-scale combustion chambers, such as utility boilers. Combustion system models aim to achieve two main objectives. One is operational problem solving, such as locating areas of corrosion in a combustion chamber. The second objective is pollutant emission reduction through operational strategies. Most of the models are capable of predicting good trend answers for the problems above. All of the models need to be compared to some measured values. However, once good agreement between the predicted and measured values is achieved, the model is used to give the full map of temperature and concentrations and other aspects in the furnace, which is not possible with any measurement system.

These models take into account fluid dynamics, local air/fuel mixing process, heat transfer and chemical kinetics. Coal combustion models are usually made up of four stages: (1) heating up, (2) devolatilization, (3) volatile combustion and (4) char combustion. Sub-models predict the pollutant formation, slagging and the physical aspects (heat transfer and flows). Much information about the combustion system is needed, such as combustion chamber geometry, feed, fuel and kinetic parameters. It is very important to use the correct coal volatilization parameters; otherwise there will be a large error in the model's results. Volatile combustion is treated with chemical reaction sub-models. Char combustion is more complicated and physical phenomena, such as boundary layer diffusion, needs to be taken into consideration [1].

Models were developed for single coal particle combustion as well as combustion systems and boilers. The single particle models contribute to the knowledge of the kinetic parameters and devolatilization and char oxidation rates, but they are unable to predict the behavior of large combustion systems, where the temperature and concentrations fields are

not uniform. Therefore models of large scale systems were developed. These models concentrate on simulating the combustion and NO_x emissions, using CFD codes which are based on numerical finite volume solutions.

There are at least 15 different comprehensive combustion models reported in the literature [5]. Some of these are commercial; some are developed by the authors and most are a combination of a commercial code with sub-models developed by the authors. These models include various sub-models of the physical processes occurring in combustion systems, including gaseous fluid dynamics, radiative and convective heat transfer, homogeneous gas phase reactions, devolatilization, heterogeneous reactions, and particle motion. Most of these comprehensive combustion codes share common features, such as: (1) capability to model three-dimensional geometries; (2) k-ε two-equation turbulence model; (3) Lagrangian model for the entrained coal particles, with single- or two-step volatilization; (4) finite-volume discretization; and (5) SIMPLE or closely related scheme for the fluid dynamics solution. All the models share a common iterative solution scheme to achieve a converged solution.

As mentioned above, there are several commercial CFD codes on the market. For instance, the FLUENT code is popular for simulating combustion systems. As it is deployed in nearly every manufacturing industry, FLUENT is not specific enough for this research (simulating both test furnace and different types of utility boilers). Thus, it was decided to work with the GLACIER code, proprietary of Reaction Engineering International (REI) as it is specific for combustion systems. REI has extensive experience with combustion in utility boilers and has modeled over 150 different utility boilers firing a range of fuels including coal, oil, gas, biomass and blends of these fuels. Types of systems modeled include tangential-fired and wall-fired units, which are the type of furnaces modeled in this work. REI has expertise in pollutant formation, furnace performance and operational impacts.

Comparing the main modeling features of two commercial CFD codes, GLACIER and FLUENT, it was found that the computational features are similar except for three major differences. GLACIER uses only a rectangular elemental mesh while FLUENT enables both linear and curvilinear surfaces. This difference can become important when boundary layer effects along this surface are important to the development of the flow field. GLACIER uses a discrete-ordinate radiation model while FLUENT has a discrete transfer model for gaseous radiation problems. This difference can affect the outcome of chemical species, such as CO_2 and O_2. Lastly, although the NO_x chemistry is similar, the input parameters for the NO_x postprocessor sub-models are different.

The features of the GLACIER code that assure its predictive power are that the GLACIER code includes mass, momentum and energy coupling between the gas and particles and coupling between turbulent fluid flow, chemical reactions, radiative and convective heat transfer, and finite-rate NO formation. It assumes that the flow field is a continuum field that can be described locally by general conservation equations for mass, momentum, energy, and species. The computational approach involves numerical discretization of the partial differential equation set. Typically, 10^5-10^6 discrete computational nodes are used to resolve the most relevant features of a three-dimensional combustion process and approximately forty variables (including gas velocities, thermodynamic properties, and concentrations of various chemical species) are tracked at each node. Coal reactions are characterized by liquid vaporization, coal devolatilization, and char oxidation. Kinetic data required for running GLACIER are: nitrogen distribution between volatile matter

and coal-char, six parameters for the two-step devolatilization mechanism ($Y_{1,2}$, $A_{1,2}$, $E_{1,2}$), three parameters for the one-step coal oxidation process (n, A_c, E_c), a set of parameters (plugged-in the code) for a comprehensive gas-phase NO_x mechanism starting from an inputted initial HCN-NH_3 ratio (ZEDA), one parameter for the conversion of char-nitrogen into NO; a total of twelve parameters, that none of them could be found in the literature. All of these parameters which describe the coal particles' combustion and NO_x formation rates can be varied in the GLACIER program to obtain the best agreement with the experimental data. After validating the GLACIER model for the test furnace, the same model parameters for the coal were used to model the utility boilers. The model validation using a test furnace before attempting prediction in a full-scale boiler strongly supports the full-scale predictions.

The approach described here determines a simplified kinetics model. While it cannot be said that the kinetic parameters used in the model will correspond to the combustion rate of a single particle of coal, these parameters do describe the combustion behavior of a "macroscopic" sample of the tested coal. For the goal of predicting the combustion behavior of these coals and blends in a utility boiler furnace, this simplified model used with the GLACIER code gave good results.

This chapter describes the GLACIER code and the different Configured Fireside Simulators (CFS), which were developed and purchased for the modeling tool.

GLACIER – Code Description

GLACIER, REI's proprietary reacting CFD software package was developed over the last three decades by researchers at the University of Utah, Brigham Young University, and REI. GLACIER is widely used to model the physical and chemical processes occurring in utility boilers. Simulating coal combustion and pollutant formation is stressed in the model code [8], which accounts for radiant and convective heat transfer, turbulent two-phase mixing, devolatilization and heterogeneous coal particle reactions (char oxidation), equilibrium (CO_2, O_2, H_2O, SO_x, CO) and finite rate (NO_x) gas-phase chemical reactions [32].

The GLACIER code includes mass, momentum and energy coupling between the gas and particles and coupling between turbulent fluid flow, chemical reactions, radiative and convective heat transfer, and finite-rate NO formation [33]. It assumes that the flow field is a continuum field that can be described locally by general conservation equations for mass, momentum, energy, and species. The computational approach involves numerical discretization of the partial differential equation set. Typically, 10^5-10^6 discrete computational nodes are used to resolve the most relevant features of a three-dimensional combustion process and approximately forty variables (including gas velocities, thermodynamic properties, and concentrations of various chemical species) are tracked at each node.

The Reynolds averaging of the conservation equations results in partial differential equations for the mean flow field variables. The equations are solved for the time averaged flow field variables. The comprehensive model uses an Eulerian-Lagrangian framework for modeling the gas and particle phases. The code couples the turbulent fluid mechanics and the chemical reaction process, using progress variables to track the turbulent mixing process and thermodynamic equilibrium to describe the chemical reactions associated with the main heat release chemistry.

GLACIER uses a Lagrangian model for particle conservation equations to predict the transport of the entrained particle phase. Particle properties such as burnout, velocities, temperatures, and particle component compositions are obtained by integrating the governing equations along the trajectories. Both gas and particle conservation equations include source terms for addition and loss of mass, momentum and energy due to interactions between the two phases. A coal particle is defined as a combination of coal, char, ash and moisture. Coal reactions are characterized by liquid vaporization, coal devolatilization, and char oxidation. The off-gas from particle reactions is assumed to be of constant elemental composition. Turbulent fluctuations and complete, local, complex chemical equilibrium are included in the particle reactions. Heat, mass and momentum transport effects are included for each particle.

The rate at which the primary combustion reactions occur is assumed to be limited by the rate of mixing between the fuel and the oxidizer, which is a reasonable assumption for the chemical reactions governing heat release. The thermodynamic state at each spatial location is a function of the enthalpy and the degree of mixing of two mixture fractions, one of which corresponds to the coal off-gas. The effect of turbulence and mixing on the mean chemical composition is incorporated by assuming that the mixture fractions are defined by a "clipped Gaussian" probability density function (pdf) having a spatially varying mean and variance. The mean and variance are computed numerically at each grid point and mean chemical species concentrations are obtained by convolution over the pdf. Specie concentrations are calculated as properties based on the local stream mixture and enthalpy. This is much more computationally efficient than tracking individual species. The radiative intensity field is solved based on properties of the surfaces and participating media, and the resulting local flux divergence (net radiant energy) appears as a source term in the gas-phase energy equation.

The numerical method is a based finite volume and the minimum running time for each case is a few days. The post processing is tabular (exit and LOI values) but mostly graphical using FIELDVIEW, by Intelligent Light, USA.

Most coal combustion modeling done today separates the models of nitrogen pollutants from the generalized combustion model. The models of nitrogen pollutants are executed after the flame structure has been predicted. The basis for this is that the formation of trace pollutant species does not affect the flame structure, which is governed by much faster fuel-oxidizer reactions. Another advantage of the approach is computational efficiency. The time required to solve the system of equations for the fuel combustion can require many hours of computer time while the pollutant sub-models typically converge in a small part (~10%) of the time required to converge the combustion case. Thus, NO_x sub-model parameters and pollutant formation mechanisms can be more easily investigated by solving the NO_x sub-model using a pre-calculated flame structure [16]. GLACIER cannot compute the full chemistry for all intermediate species, including nitrogen species. However, this detail is not required for engineering solutions for utility boiler combustion chambers. As long as all the physical mechanisms of first-order importance are also included in the investigation, engineering modifications to the boiler can be simulated using a condensed set of chemical kinetic mechanisms [8].

The Configured Fireside Simulator, CFS, is a Graphical User Interface, GUI, developed by REI for the purpose of running a preconfigured GLACIER model of a coal-fired furnace. CFS's for the test furnace and the opposite-wall and tangential-fired boilers were developed by REI for this project. The CFS's were validated by us by comparing different experimental and simulation results for each furnace type. To reach the stage where the methodology was

fully developed over 100 simulations were run. Detailed descriptions of the different CFS's are described by Spitz Beigelman [34].

2.5. Experimental and Predictive Results

Coal Characteristics and Combustion Conditions

The elemental and proximate analyses of the investigated coals are presented in table 3. Table 4 details the combustion conditions for all coals and furnace types tested.

Table 3. Analyses of the tested coals

	Coal					
Parameter, %	SA	Ad	Dr	Gln	Ven	Mel
C	73.41	72.57	76.38	73.76	75.14	71.38
H	3.93	5.10	5.15	4.99	5.30	5.12
O	5.60	19.54	11.05	7.75	7.42	17.62
N	1.69	0.89	1.47	2.09	1.41	1.36
S	0.51	0.12	0.59	0.37	0.97	0.22
Ash	14.90	1.80	5.35	11.04	9.73	4.3
Total Moisture	7.60	25.70	13.56	9.32	7.50	22.3
Residual Moisture	3.10	14.50	6.01	3.35	1.91	15.5
Volatiles	27.90	50.10	41.54	33.91	36.76	47.1
Fixed C	57.20	48.10	53.11	55.05	53.51	48.6
Gross C.V., AR (MJ kg^{-1})	26.2	21.7	26.7	26.8	29.4	22.4
Net C.V., (MJ kg^{-1})	25.3	20.2	25.4	N/A	27.2	21.5

Test Furnace Results with KPC-Melawan Coal

Results for the test furnace fired with Drummond, Adaro and Venezuelan coals are presented elsewhere [37,38]. Figure 21 shows results for the KPC-Melawan coal fired in the test furnace. The results shown are the experimental and simulation results for the test furnace achieved with the set of model parameters that provided the best agreement between the experimental data and the simulations. In general, better agreement is found for temperature than for the gases. The reasons for this are the physical mechanisms of temperature dispersion and gas formation. The change in temperature is more gradual compared to changes in gas concentrations resulting from the many different chemical reactions occurring in the furnace.
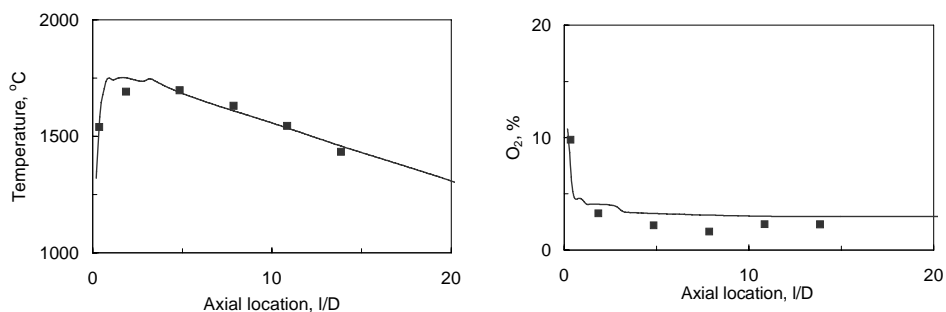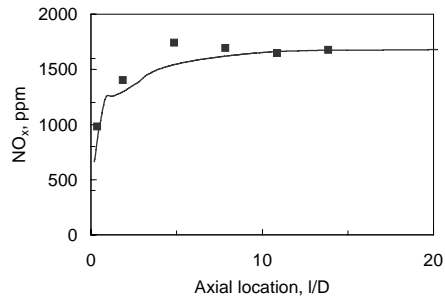
Figure 21. Experimental data (squares) and best agreement of numerical simulation (line) for temperature (top left), O_2 (top right) and NO_x (bottom left) for KPC-Melawan coal fired in test furnace.

Validation of the Model Parameters

The validity of the combustion kinetic model parameters was verified by comparison of different experimental data with simulation results obtained using the same set of combustion kinetic parameters. Good agreement between experimental data and simulation results for CO_2 and heat flux is shown in figure 22.
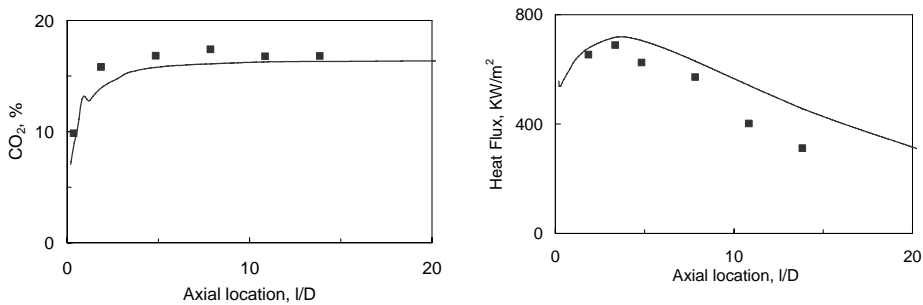
Figure 22. Experimental data (squares) and numerical simulation (line) results for CO_2 (left) and heat flux (right) for KPC-Melawan coal fired in the test furnace.

In addition, the agreement between simulation and measurement data is checked for different radial profiles in the test furnace. Figure 23 shows radial profiles at axial locations: l/D=1.83, representative of the near-burner zone, and l/D=13.83, representative of the plug-flow zone. The radial results are shown from furnace center to wall (from d/D=0 to 0.5, where d is the radial coordinate). In some cases the results at the wall are somewhat deviated due to dilution with external air and these points are not included here. The agreement between experimental data and numerical results is very good in the plug flow zone (l/D=13.83) for both temperature and NO_x. However, for NO_x the agreement to the radial profiles at l/D=0.33, which is very close to the burner (0.066 m), was not as good as the general trend. As in other works [1,25], attempts made to simulate very near-burner zone results for coal combustion did not succeed very much. Various reasons might be for this discrepancy; for example, the mesh was not fine enough, turbulent models are not suitable to these highly turbulent zones, or limitations of the sampling probe in this highly variable zone. However, good agreement

Table 4. Combustion conditions for the tested coals

Parameter	Coal					
	SA	Ad	Dr	Gln	Ven	Mel
Test furnace						
Primary air flow, $Nm^3\ h^{-1}$	9.4	12.0	11.8	11.7	9.3	9.5
Primary air temperature, ^oC	67	60	60	60	59	64
Secondary air flow, $Nm^3\ h^{-1}$	23.5	23.5	25.5	24.7	24.0	26.0
Secondary air temperature, ^oC	250	250	250	250	250	300
Total air flow, $Nm^3\ h^{-1}$	32.9	35.5	37.3	36.4	33.3	35.5
Total coal flow, $kg\ h^{-1}$	4.1	5.2	4.4	4.58	3.8	5.0
Total heat rate, $MJ\ h^{-1}$	113	127	127	119	120	135
550MW Opposite-Wall Unit						
Primary air flow, $Nm^3\ h^{-1}$	307000	371000	290000	287000	282000	358000
Primary air temperature, ^oC	78	60	78	60	78	60
Secondary air flow, $Nm^3\ h^{-1}$	831000	759000	901000	837000	789000	971000
Secondary air temperature, ^oC	322	327	311	327	322	330
Total air flow, $Nm^3\ h^{-1}$	1675000	1458000	1536000	1474000	1469000	1538000
Total coal flow, $kg\ h^{-1}$	200000	237000	181000	181000	169000	187000
Total heat rate, $MJ\ h^{-1}$	5271000	6356000	4864000	4854000	4532000	5044000
575MW Tangential-Fired Unit						
Primary air flow, $Nm^3\ h^{-1}$	313000	381000	310000	375000	302000	302000
Primary air temperature, ^oC	70	60	70	60	70	60
Secondary air flow, $Nm^3\ h^{-1}$	976000	928000	970000	926000	1054000	1054000
Secondary air temperature, ^oC	315	318	315	307	315	330
Total air flow, $Nm^3\ h^{-1}$	1560000	1549000	1551000	1541000	1550000	1527000
Total coal flow, $kg\ h^{-1}$	196000	247000	193000	192000	177000	177000
Total heat rate, $MJ\ h^{-1}$	5266000	6624000	5188000	5149000	4747000	3705000

between simulation results and experimental data was attained further downstream in the test furnace.

One must re-emphasize that the fitting procedure did not use radial profiles, but centerline profiles only. Thus, the good agreements for the radial profiles were an outcome of the whole process. Finally experiments were done at different operating conditions using the same model parameters to simulate the case with the new conditions. Example results for centerline NO_x profiles are given in figure 24 for KPC-Melawan coal fired in the test furnace with and without overfire air (OFA). For the OFA experiment a near-burner stoichiometric ratio of 0.85 was tested. OFA addition at l/D=13.83, illustrated in the dashed curve by the dip in NO_x concentration (caused by dilution), completed the total stoichiometric ratio to 1.15. Coal rate flow was the same as for the previous experiments. Good agreement between simulation results and experimental data was obtained. The model clearly predicts the influence of the near-burner lower stoichiometric ratio on the ~85% reduction of NO_x concentration at the test furnace exit.

Temperature, l/D=1.83

Temperature, l/D=13.83

NO_x, l/D=1.83

NO_x, l/D=13.83
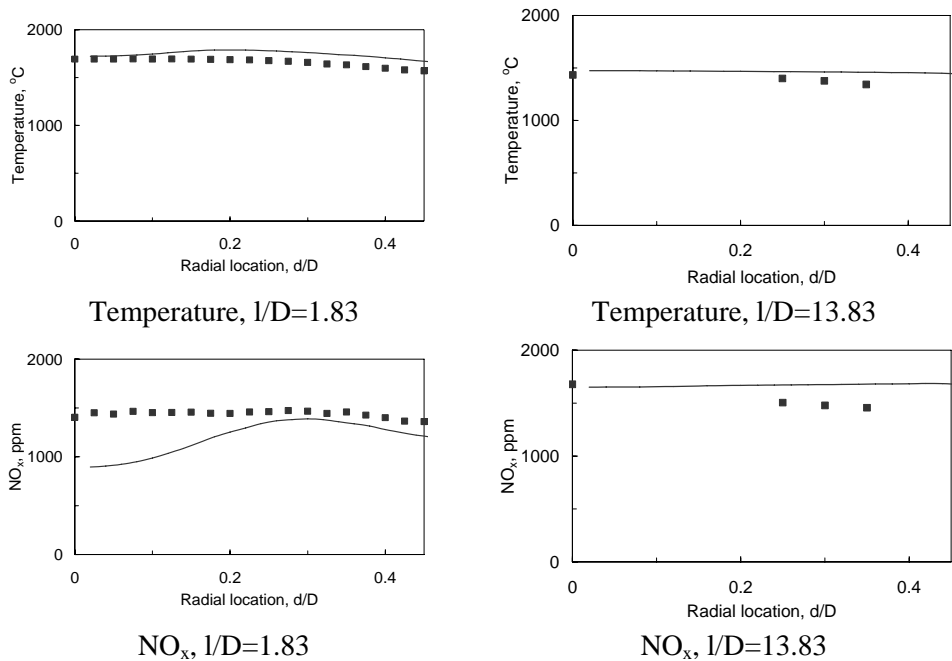
Figure 23. Experimental data (squares) and numerical simulation (line) of radial profiles at axial locations: l/D = 1.83 (left) and l/D = 13.83 (right), for temperature (top) and NO_x (bottom) when firing KPC-Melawan coal in test furnace.
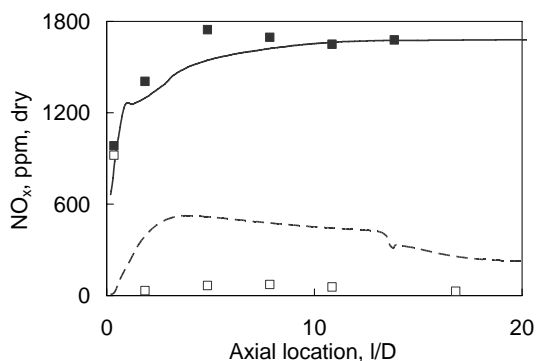
Figure 24. Experimental data (squares) and numerical simulation (line) results for NO_x from KPC-Melawan coal fired in test furnace at different firing conditions. Solid squares and line represent fired stoichiometric ratio of 1.15 (no OFA). Open squares and dashed line represent near-burner ratio of 0.85 and total stoichiometric ratio of 1.15 after OFA addition.

Table 5. Model parameters that provided best agreement between numerical results and experimental data for KPC-Melawan coal at all experimental conditions. Explanation on each parameter in text after table

Model Parameter	KPC-Melawan coal
Devolatilization	
Y_1	0.4
Y_2	0.8
A_1, sec^{-1}	375000
A_2, sec^{-1}	1.46E+13
E_1, kJ mol^{-1}	96.3
E_2, kJ mol^{-1}	251.0
Char oxidation	
n	1
A_c, kgC s^{-1} m^{-2} PaO$_2$$^{-1}$	6.55E-07
E_c, kJ mol^{-1}	87.9
NO_x formation	
VMNFR	0.3
ZEDA	0.9
ZEDAH	0.1

The devolatilization reactions are defined by the following six parameters: Y_1 and Y_2 (the mass stoichiometry coefficients), A_1 and A_2 (devolatilization rate preexponential factors), E_1 and E_2 (devolatilization activation energies). The char oxidation rate and combustion behavior are defined by three global kinetic parameters: A_c (preexponential factor), E_c (activation energy) and n (reaction order). The parameters, VMNFR, ZEDAH and ZEDA, are used for the NO_x postprocessor simulation. VMNFR defines the nitrogen division between volatile matter and char. ZEDA is the parameter that partitions the formation of volatile matter nitrogen between HCN and NH_3. ZEDAH specifies the fraction of char nitrogen that is converted to NO_x.

After verifying that the set of kinetic combustion model parameters used in the CFD model of the test furnace show good agreement between numerical results and experimental data in the test furnace plug flow zone for both axial and radial profiles and for different combustion conditions, we use the same set of model parameters for the same CFD model configured for full-scale boilers. Table 5 lists the values of the model parameters for KPC-Melawan coal, as determined from the procedure detailed in the Methodology section.

Full-Scale Predictions

The extracted combustion kinetic parameters are used for simulations of full-scale boilers using the same CFD code. We verified the full-scale model predictions with results from a series of full-scale firings done with different coals fired by IEC: three coals for the opposite-wall boiler and four coals for the tangential-fired boiler. The 550 MW drum type radiant opposite-wall fired unit was designed by Babcok and Wilcox (B and W) and the 575 MW tangential-fired unit was designed by Combustion Engineering (CE). The opposite-wall boiler is equipped with low-NO_x burners. Detailed descriptions of the boilers and furnace performance are described elsewhere [35,36] and a short description is provided here.

Boilers design

The evaluations were performed for a 550 MW drum type radiant opposite fired unit designed by Babcok and Wilcox (figure 25) and for a 575 MW unit equipped with tangentially fired boilers designed by Combustion Engineering (figure 26).

The 550 MW drum type radiant opposite wall unit (figure 25) comprises of two (2) balanced draft pulverized coal fired B and W drum type radiant boilers. These have a parallel backend arrangement for reheat steam temperature control, conservatively designed to supply steam to a single reheat steam turbine generator. Furnace walls utilize a gas-tight welded membrane construction. Internally, ribbed tubes are used in the furnace wall areas where the heat flux, fluid velocity and quality dictate their use to maintain nucleate boiling and minimum metal temperature. The furnace wall tubes are bent to accommodate the burners, observation ports, access doors and wall blowers. Integral windboxes are attached to the furnace walls of the unit in the burner zone for air distribution to the burners. Each opposite wall boiler is equipped with thirty (30) B and W DRB-XCL combination oil and coal burners. The burners are arranged in five rows high of six burners each elevation (3 burners are located at front wall and 3 burners at rear wall). The aerodynamic features of the DRB-XCL burner reduce oxygen availability during the early stages of combustion and redirect a portion of the secondary air further into the furnace to complete char burnout. The burner design also promotes a high temperature region just downstream of the nozzle exit to promote the conversion of fuel nitrogen to volatile molecular nitrogen and to thereby minimize the amount of fuel nitrogen retained in the char. Each opposite wall boiler is also supplied with six (6) B and W dual air zone over-fired air ports for reducing NO_x formation. The dual air zone NO_x port is designed to provide optimum mixing of air and flue gas in the second stage of combustion. NO_x port size selection is based upon full load operation with all the burners in service and a burner zone stoichiometry of 0.85. To improve and homogenize temperature distribution, an "aerodynamic nose" is installed in the furnace exit. The furnace bottom design is the open hopper type. The boiler consists of superheater and reheater elements. Steam capacity of the boiler is 1650 t/h, main steam/reheat steam pressures are 176/43 atm. Main steam/reheat steam temperatures are 540/540 °C.

Figure 25. Scheme of B and W drum type radiant 550 MW opposite wall furnace.

The CE 575 MW tangentially fired boiler (figure 26) is equipped with twenty straight flow burners, located in five levels, tangentially fired within the furnace. Five pulverizers, one for each level, are used. A fraction of the secondary air is fed through the closed-coupled overfire air ports located above the burners. Steam capacity of the boiler is 1700 t/h, main steam/reheat steam pressures are 181/43 atm. Main steam/reheat steam temperatures are 540/540 °C.

Figure 26. Scheme of CE 575 MW tangential fired furnace.

Figure 27. Mass-weighted averaged values before convective pass plane: full-scale tests (filled columns) and predictions (open columns) of temperature, O_2, NO_x and LOI for IEC 550MW opposite-wall boiler (left) and 575MW tangential-fired boiler (right).

Prediction Results

Figure 27 displays the results for both boiler types and compares experimental (filled columns) with simulation (open columns) results [37,38]. The results shown are the mass-

weighted averages measured or calculated at the furnace exit, before the convective pass, for temperature, O_2, NO_x and LOI. Good agreement between experiment and simulation was obtained for the temperature. It should be mentioned that the temperature results obtained from IEC are not directly measured but are calculated by the heat balance code developed at IEC [39]. For both boilers, the results are about 10-25°C higher than the predictions. This gives an error of about 1-2% for the temperature, which is in the operational limits. Oxygen test results are very close to our predictions. In most of the cases the concentration differences are between 0-0.1%, which is an error of about 3%. An exception is the oxygen concentration for Drummond coal in the opposite-wall boiler, where the measured level is 25% higher than the predicted level. We do not have an explanation for the large difference compared to the other coals and boiler. When comparing the general NO_x levels for both boilers, the opposite-wall NO_x concentrations are about 45% less than those for the tangential-fired furnace fired with the same coals. It is apparent that the opposite-wall boiler is fitted with low-NO_x burners. The simulation results clearly show this reduction trend, with an average error of about 7-8%. There is an average error of about 32% in LOI results for both boilers. It is known that the models for char burnout need to be further developed [1]. However, because the general trend of measured LOI for the different coals is reproduced by the numerical simulations; extreme values determined by simulations should be researched further before actually firing the coal.

3. PREDICTION RESULTS FOR COALS NOT TESTED BY IEC

3.1. Combustion Behavior and Pollutant Emission

As described in the previous chapter, the full-scale model predictions were verified with results from a series of full-scale tests done with different coals fired by IEC [37,38]. We verified our methodology for different coals well known to IEC: Billiton-BB Prime – a bituminous coal from South Africa, Glencore-Adaro – a sub-bituminous coal from Indonesia, Drummond-La Loma – a bituminous coal from Colombia and, for tangential-fired boiler only, Glencore-Russian – a bituminous coal from Russia. For both boilers, see figure 27, we predicted the behaviour and emissions from two coals previously unknown to IEC: Guasare-Venezuelan – a bituminous coal from Venezuela and KPC-Melawan – a sub-bituminous coal from Indonesia. For opposite-wall boiler we also simulated the combustion of Glencore-Russian coal. The combustion model parameters used in these predictions were obtained and verified according to the methodology described. The predictions were done using the operation parameters in table 4 and model parameters in table 5.

Predictions for the Russian coal fired in the opposite-wall boiler gave similar results to the Drummond coal. Further support to the validity of the simulation is shown in figure 27; the tested Russian coal is similar to Drummond also in the tangential-fired boiler.

The predicted mass-weighted averages at the furnace exit, before the convective pass, for temperature, O_2, NO_x and LOI are shown in figure 27 for the Venezuelan and Melawan coals for both boiler types. The Venezuelan coal shows a higher exit temperature for the opposite-wall boiler and a lower exit temperature for the tangential-fired boiler while KPC-Melawan coal has similar temperature values compared to the other coals. Still, the predicted

temperatures meet the design temperatures for both boilers. The oxygen levels for both coals are similar to the others with the exception of KPC-Melawan fired in the tangential-fired boiler which shows a lower than average value. NO_x levels are similar to the average for both coals in both boilers. The KPC-Melawan coal shows similar behavior to Adaro, the other sub-bituminous Indonesian coal tested, in the lower than average NO_x levels in the opposite-wall boiler. It is interesting to note for both boilers: the extremely high LOI levels predicted for Guasare-Venezuelan coal and the very low LOI levels predicted for KPC-Melawan coal, similar to Adaro. Figure 28 shows a comprehensive picture of the temperature, O_2, NO_x and velocity at the exit plane (before the convective pass) of each boiler type, which is the area of most significance for the utility operator. The different parameters need to meet the design values for good operation or environmental requirements.

Figure 28. KPC-Melawan coal: predicted gas temperature, O_2 and NO_x concentrations and velocity at the plane before the convective pass of the opposite-wall boiler (top) and tangential-fired boiler (bottom).

Table 6 presents predicted mass-weighted exit values for temperature, O_2, CO, NO_x and LOI for both boilers for three different firing configurations for KPC-Melawan coal. The results of different firing configurations were received by varying the staged combustion

stoichiometric ratios. This gives IEC a better representation of the influence of the unknown coal on combustion behavior and emissions. All parameters are within normal operational limits. The LOI results are low for full-scale boiler operation but since these results are comparable with the tested Adaro coal, we feel that the general trend of low LOI values is valid.

Table 6. Predicted mass-weighted averaged results for different firing conditions of IEC boilers fired with KPC-Melawan coal

	SR = Total stoichiometric ratio and (near-burner stoichiometric ratio)		
	550MW Opposite-wall furnace		
Parameter	*1.16(0.82)*	*1.16(0.86)*	*1.16(0.88)*
Temp., °C	1292	1295	1301
O_2, %	3.22	3.26	3.26
CO, ppm	20	27	30
NO_x, mg/dNm^3@6%O_2	174	190	193
LOI, %	0.31	0.83	0.23
	575MW Tangential-fired furnace		
Parameter	*1.2(1.0)*	*1.187(1.0)*	*1.187(0.98)*
Temp., °C	1314	1346	1361
O_2, %	2.56	2.84	3.13
CO, ppm	750	726	557
NO_x, mg/dNm^3@6%O_2	595	612	600
LOI, %	0	0.5	0

3.2. Operational Parameters Using Expert System

Expert System - Description

IEC's boilers are equipped with an on-line supervision system called EXPERT SYSTEM. The purpose of the supervision system is to quantify the performance of the combustion and heat transfer processes in real time, reporting continuously on the controlled parameter deviations from their reference values. For prediction purposes the supervision system is used in off-line mode as "what-if-then" mode. EXPERT SYSTEM is programmed to give the boiler performance based on the coal's characteristics and boiler data and now it uses data provided by BGU. The EXPERT SYSTEM analyzes the BGU data and outputs information in a manner known to IEC. Figures 29a-32c describe the process of obtaining the coal's kinetic parameters by BGU and the input of this data to EXPERT SYSTEM by IEC.

The on-line supervision system used by IEC, EXPERT SYSTEM, is described in detail by Chudnovsky et al. [35]. The basic functional aim of the supervision system is to quantify the performance of the process (of all units and its elements) in real time, reporting continuously on the controlled parameter deviations from their reference values. For prediction purposes the developed supervision system is used in off-line mode as "what-if-then" mode. The system consists of a data acquisition module, data validation model, on-line interface system, calculation modules and a data storage module. The three most important independent calculation modules are:

1. Turbine and unit heat rate calculation.
2. Boiler performance and efficiency calculation.
3. Furnace performance calculation.

The first module is intended for turbine cycle performance calculation. The second module is based on an algorithm that enables the provision of on-line boiler efficiency, heat duty and cleanliness factors for each monitoring stage. The third module is based on the FURNACE Code, which can operate in the on-line and off-line modes. The FURNACE Code uses 3D-zonal calculation model of heat transfer, which is described by Karasina et al. [40]. The code calculates the distribution of the flue gases temperature, as well as absorbed and incident heat fluxes at the furnace walls. The furnace design, burners design and arrangement, radiant heat transfer properties of the flue gas and all the operating conditions are taken into account. Validation of the calculation results is done by comparison with full-scale furnace test data [41]. Besides data of flue gas temperature and flue heat flux distribution, the FURNACE code calculates the water-wall cleanliness and coal burn-out characteristics in the furnace and estimates the temperature of the water-walls, superheater and reheater tube metal. The slagging and fouling indices of the coal are calculated separately and then used for the FURNACE code module.

Figure 29a. Extraction of combustion kinetic parameters by BGU.

INPUT DATA   **CFD OUTPUT DATA**

Combustion kinetic parameters

Utility Boiler Simulation

- Furnace exit gas temperature
- Furnace heat fluxes distribution through furnace width and height
- Furnace temperatures distribution through furnace height and width
- NO_x, CO, O_2, CO_2 emissions
- LOI.

BOILER FURNACE INPUT DATA

Design and performance data:
Furnace and burners design
Coal and air flows
Emissivity & fouling
Coal characteristics:
Proximate and ultimate analysis
Coal fineness

Figure 29b. Prediction of coal combustion behavior in boilers by BGU using boiler input information from IEC.

CFD OUTPUT DATA

IEC Boiler Calculation Module

RH&SH spray

Flue gas temperature distribution through the boiler

Steam and water temperature distribution

Metal temperature distribution

Boiler heat losses and efficiency

Power Plant Calculation Module

Unit Heat Rate

Specific Coal Consumption

Figure 29c. Prediction and on-line supervision of coal combustion behavior in boilers by IEC using coal kinetic data given by BGU.

Prediction of Boilers' Performance Using Expert System

Chudnovsky and Talanker (2004) showed that fouling factor of the furnace is the function of the basic content of ash. Generalization of the obtained data for burning of certain coals [41] show that furnace cleanliness (fouling), under the same operation conditions, depends on ash characteristics. As was shown by Chudnovsky and Talanker (2004) the

fouling factor of the furnace is a function of the basic content of ash, at NCR load for furnace clean condition after soot blowing. The basic content is equal to the ratio in Eq. 7:

$$Basic = \frac{B}{A+B} * 100\%$$

(7)

As follows, a basic content value increase leads to increase in fouling.

In order to clarify the influence of water wall absorptivity factor on radiation heat transfer process in the furnace, we provided a numerical analysis. For this purpose we ran a furnace calculation for different absorptivity factors, using commonly accepted water-wall fouling. Generalization of the obtained results enabled us to receive water-wall absorptivity as a function of acid ratio [41]. Approximation of the obtained results may be expressed by Eqs. 7 and 8:

For bituminous type ash:

$$a_w = -6.61 \left(\frac{A}{A+B}\right)^2 + 8.05 \left(\frac{A}{A+B}\right) - 1.19$$

(8)

$$where \ 0.82 < \frac{A}{A+B} < 0.95$$

For lignite type ash:

$$a_w = 1.81 \left(\frac{A}{A+B}\right)^2 + 1.92 \left(\frac{A}{A+B}\right) + 0.49$$

(9)

$$where \ 0.80 < \frac{A}{A+B} < 0.87$$

Due to low acid index of the ash the Melawan coal belongs to high emissivity ash and

it will increase heat absorption in the furnace. From the other hand base index of the coal is very high and it may lead to increasing of fouling of the waterwall. However because of low ash content both factors influence should be less that in the case of typically burned, bituminous, coals.

Using the above data and the coal characteristics, we ran simulations of the opposite-wall and tangential-fired boilers for firing KPC-Melawan coal. The results are summarized in table 7.

Table 7. Predicted KPC-Melawan coal firing performance at NCR unit load

Parameter	Opposite wall boiler, 550 MW	Tangential fired boiler, 575 MW
Coal consumption, t/h	212	240.0
PA+FD Fan Air Flow, t/h	2040	2200.0
ID Fan Flue Gas Flow, t/h	2280.0	2460
SH spray, t/h	110-120	<5
RH spray, t/h	0.0	<3

Table 8. Predicted KPC-Melawan coal firing performance at NCR unit load

Parameter	Opposite wall boiler, 550 MW	Tangential fired boiler, 575 MW
Maximum mill capacity, t/h	52	52.5
Boiler heat losses, %	5.80	5.4
Unburned carbon heat losses, %	0.2	0.28
Efficiency, %	93.8	94.1
FEGT (preliminary)	1300	1370

**Table 9. Evaluation summary: IEC opposite-wall boiler
behavior when firing KPC-Melawan coal**

Topic	Evaluation
Coal type	Sub-bituminous coal
Ash type	KPC-Melawan coal belongs to lignitic ash category and corresponds to high slagging indices. Fe_2O_3/CaO indices are equal to 1.16 and may increase its fouling potential.
Ash softening temperature	The hemispheric ash temperature of KPC-Melawan coal is 1190° C. For lignitic ash category slagging index depends on AFT. For the current ash composition the coal has high slagging and low fouling potential. However, low ash content may prevent furnace slagging.
Mills capability	Due to low KPC-Melawan coal heating value, NCR load is provided by five pulverizers in operation
Mills outlet temperature	High moisture content in the coal requires raising primary air temperature before pulverizer in order to provide acceptable pulverizer exit temperature. Maximum possible pulverizers exit temperature 60 ℃ was achieved with primary inlet temperature about 240-260 ℃.
PA, FD and ID Fans capability	Primary air flow through the pulverizers is about 95 t/h. PA+FD fan air flow is about 2040 t/h. ID Fan capacity is 2280 t/h. (This causes a unit load limitation, especially for the summer period.)
SO_2 and NO_x emission tendency	Low sulfur content in KPC-Melawan coal leads to reduced SO_2 concentrations. Fixed carbon to volatile matter ratio reduction also leads to lower NO_x formation.
Boiler performance	The boiler efficiency at NCR load is equal to 93.8% (low heating value base). SO_2 emission is equal to approximately 400 mg/dNm^3@6%O_2. NO_x emission is equal to 210 mg/dNm^3@6%O_2 at NCR load. LOI is equal to 4.0% at NCR load. Furnace exit temperature 1300 ℃.
Ignition point	The selected pulverizer and firing system operation conditions provide reliable coal nozzle operation with acceptable distance of ignition point from the nozzle.
Coal burnout	Coal burnout rate is slightly less in comparison with typical coal firing.
Furnace absorptivity	Furnace absorptivity is equal to 0.9-0.95 and fouling slightly less than for typical coal with the same ash composition.
Furnace performance	KPC-Melawan coal firing provides stable ignition and combustion process. Due low ash content incident heat flux less than for typical coal firing. Furnace exit temperature does not exceed allowable limit.
Metal temperature and steam temperature control	Superheater spray flow at full load was approximately 110 t/h and reheater spray flow was 0 t/h. Superheater spray provides steam temperature control in acceptable range. Water wall, SH and RH tube metal temperature less than allowable limit. Burner tilt and superheater spray provide steam temperature control in acceptable range. Flame position is the furnace is not symmetric and it is recommended to perform additional combustion tuning to correct this.

Opposite-Wall Boiler

The total air flow is almost similar for KPC-Melawan coal in comparison with typical bituminous coals burning in IEC. At the same time, high water content in Melawan coal leads to a slightly increased flue gas flow (about 3%) and as a result a pressure drop increase about 7% is expected in flue gas path and therefore the ID Fan capacity is increased about 10%. This may cause a unit load limitation, especially for the summer period. Due to high water content in the burning coal, pulverizers inlet temperature is increased to approximately 240-260 ºC. Pulverizers outlet temperature control is limited, however we expect to be able to provide pulverizers outlet temperature about 60 ºC which is acceptable for stable ignition and firing. The primary air flow through pulverizes is about 95 t/h. Together with lower primary air temperature the ignition point will be located away from coal nozzles and will provide reliable operation of the nozzles. Superheater spray flow at full load is approximately 110 t/h and reheater spray flow is 0 t/h. Sulfur content in burning coal is 0.22% (DCB). Coal index (ratio of fixed carbon to volatile matter content) is equal to 1.064. Ash content reduction (4.3%) in comparison with typical coals decreases the unburned carbon losses and, as a result, boiler efficiency will increase when KPC-Melawan coal is fired. Evaluation is summarized in table 8.

Tangential-Fired Boiler

The total air flow for KPC-Melawan coal is slightly less in comparison with typical coals burning in IEC. At the same time, high water content in KPC-Melawan coal will lead to slightly increasing flue gas flow (about 3%) and as a result the pressure drop in flue gas path will increase about 7%. Therefore the ID Fan capacity will increase about 10% and will cause a unit load limitation, especially during summer period. This limitation may cause maximum achievable excess oxygen content at full load reduction up to 2.8%. This is less than boiler design data and may lead to increasing CO and unburned carbon in the ash. Due to high water content in the burning coal, pulverizers inlet temperature is increased to approximately 250-260 ºC and in this case pulverizers outlet temperature control is limited. However a pulverizers outlet temperature about 60 ºC can be provided which is acceptable for stable ignition and firing. The primary air flow through the pulverizers is about 95-98 t/h and together with lower primary air temperature the ignition point will move away from the coal nozzles and reliable operation of the nozzles can be achieved. Fuel air damper control is handled according to CE recommendation. The fuel oil and auxiliary air nozzle control compensates to keep the required windbox - furnace pressure drop. Superheater spray flow at full load does not exceed 5-10 t/h and reheater spray flow is equal to 0 t/h. Sulfur content in burning coal is 0.22% (DCB). Coal index (ratio of fixed carbon to volatile matter content) is equal to 1.064. Low ash content of 4.3% in comparison with typical, bituminous, coals will decrease the unburned carbon losses and, as a result, the boiler efficiency increases when KPC-Melawan coal is fired. Evaluation is summarized in table 9.

**Table 10.Evaluation summary: IEC tangential-fired boiler
behavior when firing KPC-Melawan coal**

Topic	Evaluation
Coal type	Sub-bituminous coal
Ash type	KPC-Melawan coal belongs to lignitic ash category and corresponds to high slagging indices. Fe_2O_3/CaO indices are equal to 1.16 and may increase its fouling potential.
Ash softening temperature	The hemispheric ash temperature of KPC-Melawan coal is 1190° C. For lignitic ash category slagging index depends on AFT. For the current ash composition the coal has high slagging and low fouling potential. However, low ash content may prevent furnace slagging.
Mills capability	Due to low KPC-Melawan coal heating value, NCR load is provided by five pulverizers in operation
Mills outlet temperature	High moisture content in the coal requires raising primary air temperature before pulverizer in order to provide acceptable pulverizer exit temperature. Maximum possible pulverizers exit temperature 60 ºC was achieved with primary inlet temperature about 250-260 ºC.
PA, FD and ID Fans capability	Primary air flow through the pulverizers was about 95 t/h. PA+FD fan air flow about 2200 t/h. ID fan flue gas flow about 2460 t/h. Due to high moisture content in the Melawan coal the total flue gas flow is approximately 3% higher than typical coals fired which leads to an increase in the ID Fan capacity of about 9-10%. As result ID Fan achieved its allowable limit at 575 MW load and it causes unit load limitation especially during summer period.
SO_2 and NO_x emission tendency	Low sulfur content in KPC-Melawan coal leads to reduced SO_2 concentrations. Fixed carbon to volatile matter ratio reduction also leads to lower NO_x formation.
Boiler performance	The boiler efficiency at NCR load is equal to 94.0% (low heating value base) with boiler exit temperature less than design value. SO_2 emission is equal to approximately 400 mg/dNm^3@6% O_2. NO_x emission is equal to 550 mg/dNm^3@6% O_2 at NCR load. LOI is equal to 2.5 % at NCR load. Furnace exit temperature 1370ºC.
Ignition point	The selected pulverizer and firing system operation conditions provide reliable coal nozzle operation with acceptable distance of ignition point from the nozzle.
Coal burnout	Coal burnout rate is slightly less in comparison with typical coal firing.
Furnace absorptivity	Furnace absorptivity is equal to 0.9-0.95 and fouling slightly less than for typical coal with the same ash composition.
Furnace performance	KPC-Melawan coal firing provides stable ignition and combustion process. Due to low ash content, incident heat flux less than for typical coal firing. Furnace exit temperature does not exceed allowable limit.
Metal temperature and steam temperature control	Superheater spray flow at full load is less than 10 t/h and reheater spray flow 0 t/h. Burner tilt and superheater spray provide steam temperature control in acceptable range. SH and RH tube metal temperature less than allowable limit. Burner tilt and superheater spray provide steam temperature control in acceptable range.

4. Conclusions

We successfully predicted performance and emissions of bituminous and sub-bituminous coal types from full-scale pulverized coal utility boilers of type: opposite wall and tangential fired. To predict the combustion behavior and pollutant emissions of coal in pulverized-coal utility boilers, we developed a methodology combining measurements in a 50kW pilot-scale test facility with simulations using the same CFD code configured for both test and full-scale furnaces. There is no attempt to predict the combustion behavior of the utility boiler based on the combustion behavior of the test furnace. For the goal of predicting combustion behavior

of coals in a utility boiler furnace, our methodology gives good results. In addition to predicting combustion behavior and emissions, IEC developed an online supervision system called EXPERT system. This system calculates the different operator needed information.

REFERENCES

[1]   A. Williams, R. Backreedy, R. Habib, J. M. Jones and M. Pourkashanian, *Fuel.* 81 (5), 605-618 (2002).

[2]   A. Arenillas, R. I. Backreedy, J. M. Jones, J. J. Pis, M. Pourkashanian, F. Rubiera and A. Williams, *Fuel.* 81 (5), 627-636 (2002).

[3]   A. M. Carpenter, *IEACR/81*. IEA Coal Research, London, 1995.

[4]   J. M. Beer, Progress in Energy and Combustion Science. 26, 301-327 (2000).

[5]   A. M. Eaton, L. D. Smoot, S. C. Hill and C. N. Eatough, *Progress in Energy and Combustion Science*. 25, 387-436 (1999).

[6]   D. W. Pershing and J.O.L. Wendt, *Ind. Eng. Chem. Process Des. Dev.* 18(1), 60-67 (1979).

[7]   H. Lorenz, E. Carrea, M. Tamura and J. Haas, *Fuel*. 79, 1161-1172 (2000).

[8]   K-T Wu, H. T. Lee, C. I. Juch, H. P. Wan, H. S. Shim, B. R. Adams and S. L. Chen, *Fuel*. 83, 1991-2000 (2004).

[9]   S. K. Ubhayakar, D. B. Stickler, C. W. Von Rosenberg and R. E. Gannon, *16th Symposium (International) on Combustion*. The Combustion Institute, 427-436 (1977).

[10]   C. Sheng, B. Moghtaderi, R. Gupta and T. F. Wall, *Fuel*. 83, 1543-1552 (2004).

[11]   R. I. Backreedy, R. Habib, J. M. Jones, M. Pourkashanian and A. Williams, *Fuel*. 78, 1745-1754 (1999).

[12]   T. H. Fletcher and D. R. Hardesty, *SAND92-8209 UC-362*. Sandia Report (1992).

[13]   J. M. Jones, P. M. Patterson, M. Pourkashanian, A. Williams, A. Arenillas, F. Rubiera and J. J. Pis, *Fuel*. 78, 1171-1179 (1999).

[14]   R. H. Hurt and R. E. Mitchell, *24th Symposium (International) on Combustion*. The Combustion Institute, 1243-1250 (1992).

[15]   D. Smoot, editor, "Fundamentals of coal combustion for clean and efficient use", *Elsevier Science Publishers B.V.*, The Netherlands (1993).

[16]   S. C. Hill and L. D. Smoot, *Progress in Energy and Combustion Science*. 26, 417-458 (2000).

[17]   P. Glarborg, A. D. Jensen and J. E. Johnsson, *Progress in Energy and Combustion Science*. 29, 89-113 (2003).

[18]   A. Molina, A. G. Eddings, D. W. Pershing and A. F. Sarofim, *Progress in Energy and Combustion Science*. 26, 507-531, (2000).

[19]   J. P. Spinti and D. W. Pershing, *Combustion and Flame*. 135, 299-313 (2003).

[20]   A. Williams, M. Pourkashanian and J. M. Jones, *Progress in Energy and Combustion Science*. 27, 587-610 (2001).

[21]   M. Rostam-Abadi, L. Khan, J. A. DeBarr, L. D. Smoot, G. J. Germane and C. N. Eatough, *American Chemical Society, Division Fuel Chemistry*. 41(3), 1132-1137 (1996).

[22]   S. Schafer and B. Bonn, *Fuel*. 79, 1239-1246 (2000).

[23]   J. A. Miller and C. T. Bowman, *Progress in Energy and Combustion Science*. 15, 287-338 (1989).

[24]   P. J. Smith, S. C. Hill and L. D. Smoot, *19th Symposium (International) on Combustion*. The Combustion Institute, 1263-1270 (1982).

[25]   F. C. Lockwood, T. Mahmud and M. A. Yehia, *Fuel*. 77(12), 1329-1337 (1998).

[26]   D. Gera, M. Mathur and M. Freeman, *Energy and Fuels*. 17, 794-795 (2003).

[27]   R. Kurose, H. Makino and A. Suzuki, *Fuel*. 83, 693-703 (2004).

[28]   J. C. P. Spinti, "An experimental study of the fate of char nitrogen in pulverized coal flames", PhD dissertation, The University of Utah (1997).

[29]   Q. Zhu, J. M. Jones, A. Williams and K. M. Thomas, *Fuel*. 78, 1755-1762 (1999).

[30]   J. Z. Chu, S. S. Shieh, S. S. Jang, C. I. Chien, H. P. Wan and H. H. Ko, *Fuel*. 82, 693-703 (2003).

[31]   H. Zhou, K. Cen and J. Fan, *International Journal of Energy Research*. 29, 499-510 (2005).

[32]   J. Valentine, M. Cremer, K. Davis, J. J. Letcavits and S. Vierstra, *2003 Power-Gen Conference*, December 9-11, 2003, Las Vegas, NV USA (2003).

[33]   H. S. Shim, private correspondence (2004).

[34]   N. Spitz Beigelman, Combustion of Coal Blends. PhD dissertation, Ben-Gurion University of the Negev, Israel (2006).

[35]   B. Chudnovsky, L. Levin and A. Talanker, Advanced On-line Diagnostic for Improvement of Boiler Performance and Reduction of NO_x Emission. Proceedings of the PowerGen 2001 Conference, Brussels, Europe (CDROM).

[36]   B. Chudnovsky, E. Karasina, B. Livshits and A. Talanker, Application of Zonal Combustion Model for On-line Furnace Analysis of 575 MW Tangentially Coal Firing Boilers. Proceedings of the PowerGen 1999 Conference, Frankfurt, Europe (CDROM).

[37]   N. Spitz, E. Bar-Ziv, R. Saveliev, M. Perelman., E. Korytni, G. Dyganov and B. Chudnowsky, POWER2006-88065. Proceedings of ASME POWER2006, Atlanta, Georgia (2006).

[38]   R. Saveliev, B. Chudnowsky, B. Kogan, E. Korytni, M. Perelman, Y. Sella, N. Spitz and E. Bar-Ziv, POWER2007-22065. Proceedings of ASME POWER2007, San Antonio, Texas (2007).

[39]   A. Vikhansky, E. Bar-Ziv, B. Chudnovsky, A. Talanker, E. Eddings and A. Sarofim, *International Journal of Energy Research*. 28, 391-401 (2004).

[40]   E. Karasina, Z. Shrago and S. Borevskaya, *Teploenergetika*. 7, 42-47 (1982) (in Russian).

[41]   B. Chudnovsky and A. Talanker, Effect of Bituminous Coal Properties on Heat Transfer Characteristic in the Boiler Furnaces. 2004 ASME International Mechanical Engineering Conference, Anaheim, California, USA (CDROM).

Chapter 4

STUDY OF ELECTROGASDYNAMIC POWER CONVERSION

Armando C. Oliveira[1], Saffa B. Riffat[2], Szabolcs Varga[1] and Yuehong Su[2]*

[1] Faculty of Engineering, University of Porto; New Energy Technologies Unit
Rua Dr Roberto Frias, 4200-465 Porto, Portugal
[2] School of the Built Environment, The University of Nottingham
University Park, Nottingham, NG7 2RD , UK

ABSTRACT

Electrogasdynamic (EGD) power conversion is a process that converts thermal (internal/kinetic) energy into electric energy, without moving parts. Discrete particles are charged using a corona electrode and transported at high velocity against an electric field, to produce useful electric power. An important advantage of such a device is that it doesn't include moving parts and so requires very little maintenance.

The basic equations for EGD power conversion are presented, as well as theoretical results for the process. The efficiency of EGD power conversion is calculated for different fluids. The existence of practical limits for conversion is discussed: electric breakdown strength and charge-to-mass ratio. Theoretical results for different working fluids are presented.

A Computational Fluid Dynamics model was developed as a tool to simulate the major characteristics of fluid flow in such a device and to identify the most important factors in the power conversion process. Model results are presented and discussed. The importance of the particle/electric field interaction is evaluated, taking into account turbulent effects. A parametric study to identify the best collector location is carried out.

Experimental results for an EGD apparatus and different operating fluids were also obtained in a test rig. An EGD nozzle was designed, built and experimentally tested. Tests of electric breakdown were carried out for different working fluids (refrigerants). Results are presented and compared to theoretical values. The use of electrospray is also investigated.

---

* Armando C. Oliveira: corresponding author, acoliv@fe.up.pt

1. INTRODUCTION. PRINCIPLE OF EGD CONVERSION AND EQUATIONS

Electric power conversion in an Electro-Gas-Dynamic (EGD) converter is accomplished by charging discrete particles using a corona electrode and transporting these particles at high velocity against an electric field. As electric potential builds up, useful electric power can be collected. An important advantage of such a device is that it doesn't include moving parts and so requires very little maintenance. Another interesting property is that it uses a high pressure fluid, e.g. coming from a generator (boiler), and therefore could be easily incorporated in a combined heat and power cycle.

Marks [1] started and conducted most of the existing research on EGD power generation, using a charged aerosol without moving mechanical parts. He presented a mathematical analysis based on a thermodynamic approach and Poisson's equation. Several simplifying assumptions were made in his first work with relation to fluid flow (e.g. constant volumetric flow rate along the converter, uniform velocity distribution, negligible frictional losses, etc.). Marks [2], [3] and [4] extended his analysis to compressible flow, assuming a power law equation for the cross section, one dimensional velocity and charge density variation, and solved the Poisson equation analytically. Other methods based on EGD conversion developed by the same author can be found in [5], [6] and [7]. Borok [8] published the simulation results of an EGD generator operating at very high pressure but relatively low gas flow rates. Lawson and Ohain [9] presented an early overview of the characteristics and mathematical equations of power conversion involving electro-fluids. At the time of those studies, EGD power generation was thought for large scale and centralised applications, and research was carried out in this context. However it has been shown that distributed power generation is a more sustainable method of electricity production. EGD conversion could be advantageous for such applications, although a significant amount of research is needed in order to efficiently explore the concept.

The principle of an EGD power converter is shown schematically in figure 1. An EGD converter comprises a corona electrode, an attractor electrode, a collector electrode and a duct of electrically insulating materials. A sufficiently high voltage is imposed between the corona and attractor electrodes to charge the particles carried by a gas stream. When a high pressure vapour (e.g., steam) flows through a converging duct, the expansion causes it to be cooled and part of it to condense. The corona electrode emits electrons which then attach themselves to the condensed droplets to form a charged aerosol and the charged aerosol is swept downstream by high speed vapour. The high speed vapour goes on to push the charged droplets to the collector electrode, building up a higher electric potential than the corona electrode. In this process, the kinetic energy of the vapour is converted to electrical energy. Useful electrical energy is obtained from the converter by connecting a load between the collector and corona electrodes, i.e., direct conversion of heat (that generates the vapour) to power is accomplished by the EGD converter.
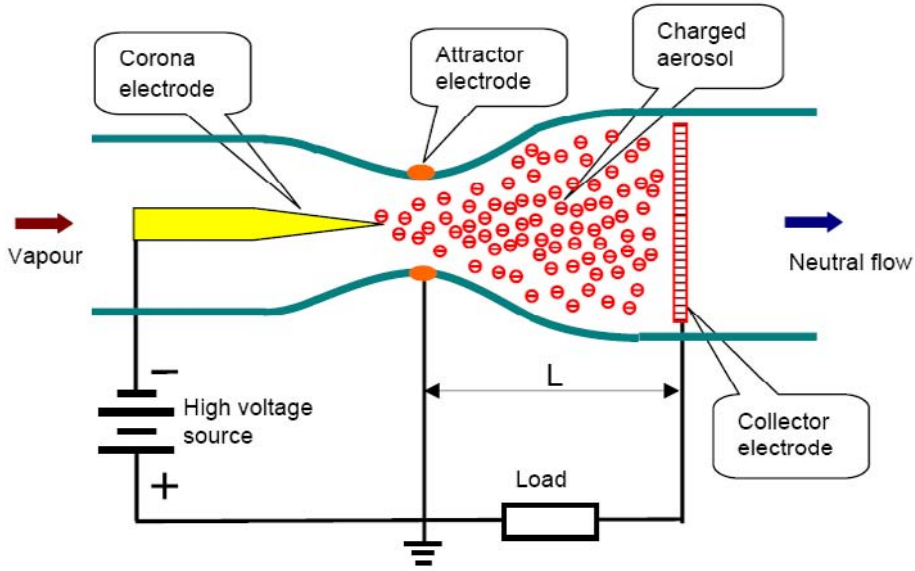
Figure 1. Principle of EGD power converter.

The corona electrode is positioned centrally in the converging section of the nozzle, the annular attractor electrode is placed at the throat and the collector electrode is positioned centrally in the diverging section. Compared with conventional turbine/generator power conversion systems, this system is simple and reliable as there are no high-speed moving parts involved.

Some basic equations for EGD power conversion will be presented next. A simple model includes the following assumptions: (1) one-dimensional flow; (2) constant charge density; (3) negligible charge mobility (i.e., negligible velocity difference between charge and working fluid); (4) the working fluid is already charged.

One may consider 2 fluid evolutions between converter inlet (1) and outlet (2): process 1-0 in the converging section of the nozzle, where the electrically charged aerosol is formed, and process 0-2 in the diverging section where EGD conversion is accomplished. Energy conservation within the converging section of the nozzle is given by:

$$h_1 + \frac{1}{2}u_1^2 = h_0 + \frac{1}{2}u_0^2 \tag{1}$$

Energy conservation within the EGD conversion section may be given by:

$$h_0 + \frac{1}{2}u_0^2 = h_2 + \frac{1}{2}u_2^2 + \frac{IV}{\dot{m}_g} \tag{2}$$

where I is the current and V is the voltage produced, and the mass flow rate is equal to $\dot{m}_g = \rho_0 u_0 A_0$.

An isentropic efficiency may be defined for the nozzle as:

$$\frac{h_1 - h_0}{h_1 - h_{0,s}} = \eta_n = \frac{h_0 - h_2}{h_0 - h_{2,s}}$$

(3)

The efficiency of EGD power conversion is given by the ratio between electricity output and energy spent to generate the vapour. This one corresponds to increase vapour enthalpy to the converter inlet condition (after condensation which follows converter outlet). It may be expressed as:

$$\eta = \frac{P_e}{Q_g} = \frac{IV}{\dot{m}_g \Delta h_g}$$

(4)

The current I is given in the electrically charging process and only the voltage V needs to be determined. This may be given by applying the Poisson's equation which correlates the electric potential and the space charge density:

$$\nabla^2 V = -\frac{\delta}{\varepsilon_0 \varepsilon_r}$$

(5)

where δ is the charge density, ε_0 is the dielectric constant of free space and ε_r is the relative dielectric constant of the working fluid.

The charge density may be given by:

$$\delta = I/(u_p A)$$

(6)

where the velocity of charged particles is given by:

$$u_p = u + \lambda \cdot E$$

(7)

where λ is the mobility of charges and E is the electric field strength.

The distribution of the charge density is needed to solve equation (5). It may be calculated by assuming some kind of cross section variation along the flow direction, as shown in [10]. In order to simplify the analysis, it may be assumed that the electric field is one-dimensional, the charge density is constant and the mobility is negligible. Integrating equation (5) and substituting into equation (4) gives:

$$\eta = \frac{I^2 \rho_0 L^2}{2\dot{m}_g^2 \Delta h_g \varepsilon_0 \varepsilon_r}$$

(8)

where L is the length of the EGD conversion section. Equation (8) shows that EGD conversion efficiency changes proportionally with the square of current.

Unlike other conversion methods, EGD conversion efficiency is limited by the electric breakdown strength of the working fluid. Along the conversion section the electric field is maximum at the entrance and may be expressed as:

$$E_0 = \sqrt{\frac{2\eta\rho_0 \Delta h_g}{\varepsilon_0 \varepsilon_r}} \qquad (9)$$

Taking into account the electric breakdown strength of working fluids, the possible maximum conversion efficiency may be expressed as:

$$\eta_b = \frac{\rho_0 b_0^2 \varepsilon_0 \varepsilon_r}{2\Delta h_g} \qquad (10)$$

where b_0 is the standard electric breakdown strength of working fluids, which may be multiplied by the relative density of the working fluid to give its electric breakdown strength under corresponding condition [2]. It is obvious that the conversion efficiency at electric breakdown increases proportionally with increasing inlet vapour density and squared reference electric breakdown strength.

When calculating EGD conversion, a charge-to-mass ratio limitation should also be considered. The ratio of the collector current to the nozzle mass flow rate should be beyond this limit. Usually, the charge-to-mass ratio is around 10^{-6} C/kg for gas and 10^{-3} C/kg for liquid, respectively. As indicated by equation (8), for a given EGD efficiency, a smaller charge-to-mass ratio is related to larger density and smaller latent heat of evaporation.

2. THEORETICAL RESULTS

The previous equations can be used to assess EGD conversion characteristics and efficiency. Different fluids were considered, such as air, water and some refrigerants. The use of refrigerants is due to the possible application to combined heat and power systems.

Calculation results for compressed air are given in figures 2-4. EGD power output increases with the collector current. Electric breakdown happens at a collector current of 20 μA for a nozzle throat diameter of 3 mm and an inlet pressure of 4 bar, as shown in figure 2. The power output at electric breakdown is about 0.5 W. Figure 3 shows that higher inlet pressure will cause larger electric breakdown current and power due to larger density and flow rate. Larger power output can also be achieved using a bigger nozzle, as shown in figure 4.

Figures 5 gives the change of EGD efficiency with collector current for water as the working fluid. It is assumed that the converter has a nozzle efficiency of $\eta_n = 0.90$. From a thermodynamic point of view, 10.8% is the maximum power conversion efficiency of the cycle under the calculation condition. However, EGD power conversion efficiency is limited by the electric breakdown strength of the working fluid. Electric breakdown happens at a collector current of 5.7 μA for the calculation condition. EGD efficiency at electric

breakdown is 0.0012%. The maximum EGD efficiencies for various inlet pressures are given in Figure 6. As steam has very low density and small electric breakdown strength (b_0=3.22 kV/mm), the performance of EGD conversion for steam (water) is very poor. Steam is therefore not a suitable working fluid.
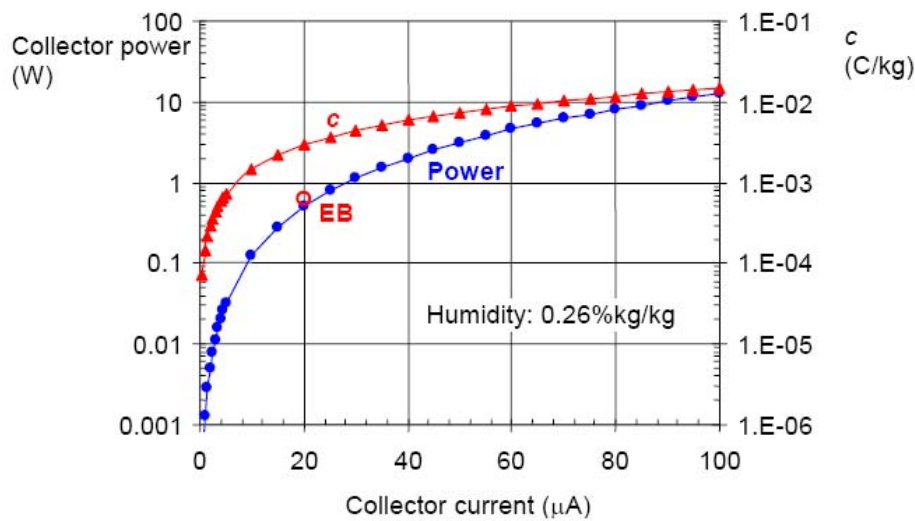
Figure 2. Power and charge-to-mass ratio against collector current (Air, p=4 bar, d_n=3 mm, L=10 mm, EB – electrical breakdown condition).
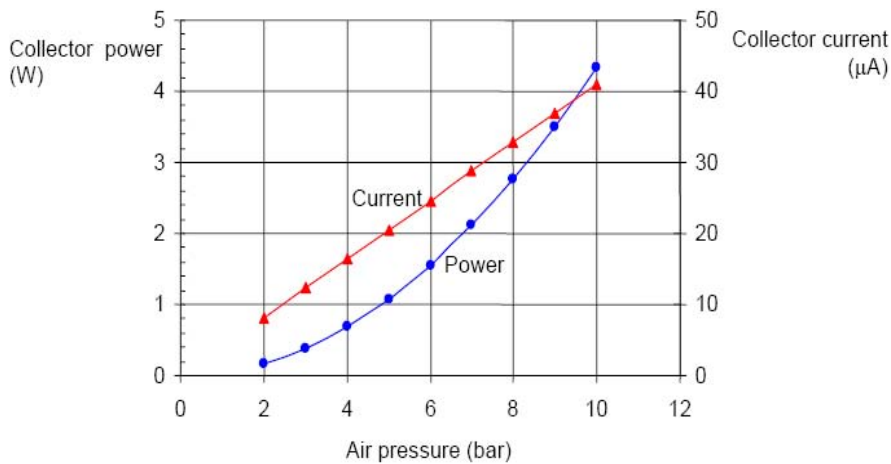
Figure 3. Maximum collector power and current at electrical breakdown (Air, d_n=3 mm, L=10 mm).
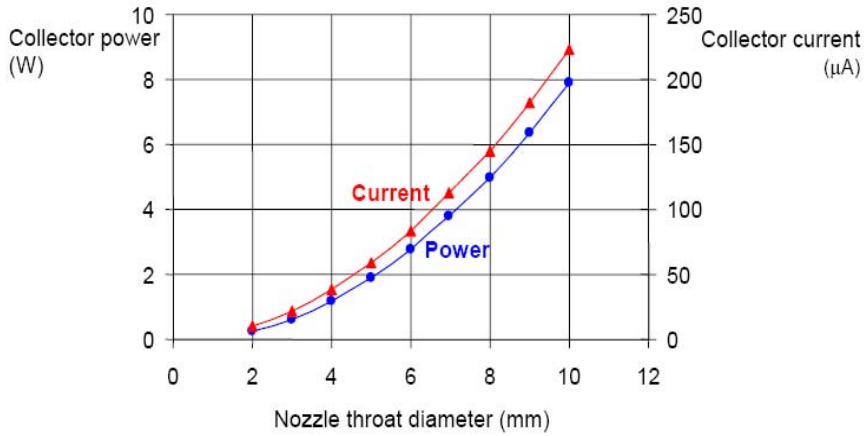
Figure 4. Maximum collector power and current at electrical breakdown (Air, p=4 bar, L=10 mm).
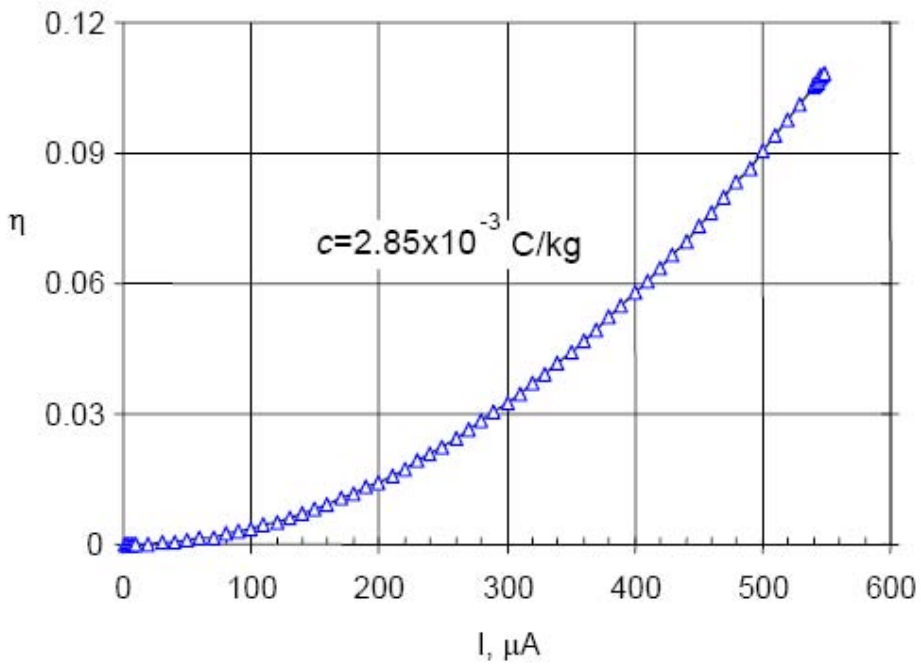
Figure 5. Efficiency against current (Water, T_g=120ºC, T_c=40ºC, d_n=3 mm, L=10 mm).
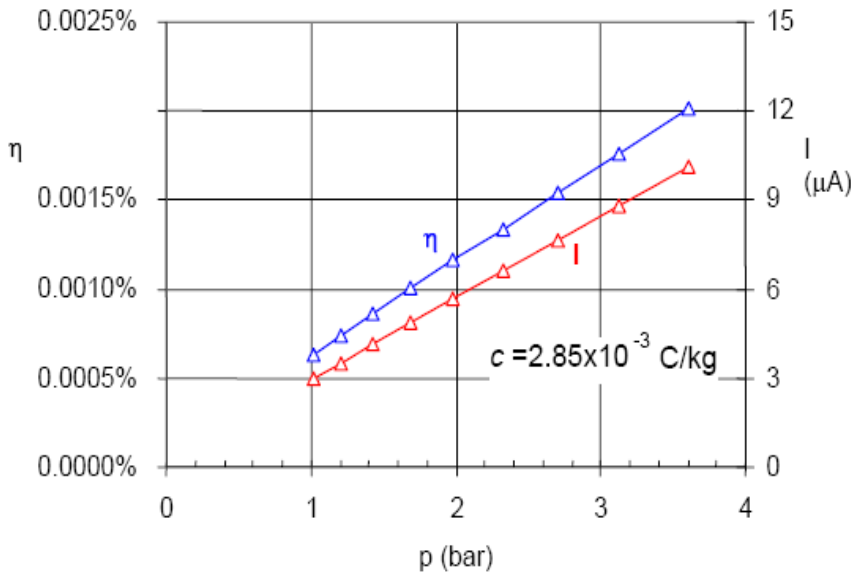
Figure 6. Efficiency and current at electrical breakdown as a function of vapour pressure (Water, T_c=40ºC, d_n=3 mm, L=10 mm).

Calculation results for refrigerant HFE7100 as working fluid are given in figures 7-10. Compared with steam, HFE7100 has larger vapour density and smaller latent heat of evaporation. This will allow a larger EGD efficiency at electric breakdown. Figure 7 shows the change of EGD power output with collector current and figure 8 gives EGD efficiency. If HFE7100 has the same standard dielectric strength as air, i.e., $b_0=b_{0\_air}$, electric breakdown will happen at the collector current of 33 μA for the calculation condition. This will allow an EGD power output of 8 W and an EGD efficiency of 0.33%. If $b_0=3xb_{0\_air}$, the electric breakdown current will be 100 μA and the EGD power and efficiency will be 72 W and 3%. The charge-to-mass ratio at electric breakdown is $2.85x10^{-3}$ C/kg if $b_0=b_{0\_air}$ and $8.5x10^{-3}$ C/kg if $b_0=3xb_{0\_air}$. It is around the charge-to-mass ratio limit for liquid. This would play a main limit on EGD conversion for HFE7100 as working fluid. Figure 9 shows the effect of conversion duct length on charge-to-mass ratio. For a given EGD efficiency, a longer conversion duct allows a smaller charge-to-mass ratio. However, in practice, this would cause a larger charge loss to the wall of the conversion duct. Figure 10 gives the change of EGD efficiency and power with the inlet pressure of nozzle when $b_0=3xb_{0\_air}$.
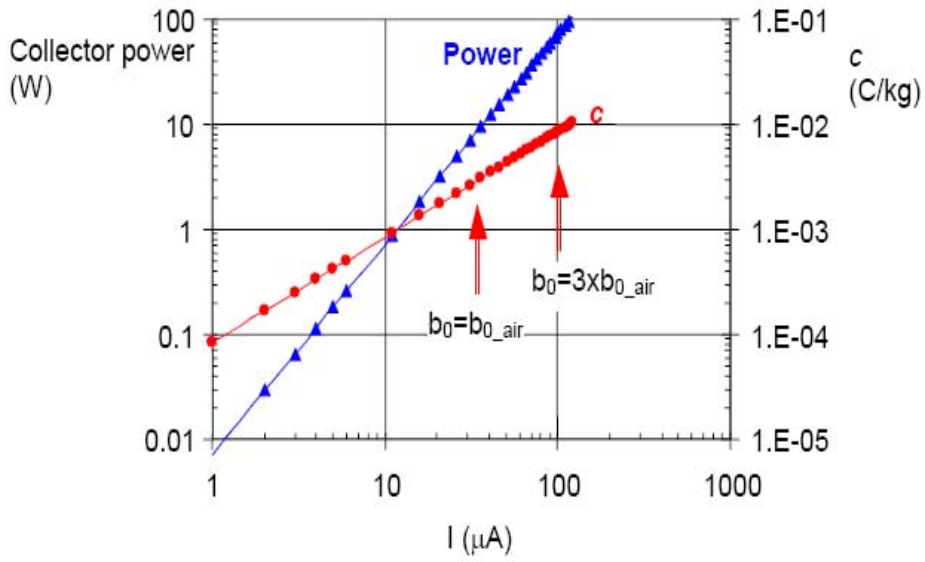
Figure 7. Power and charge-to-mass ratio against current (HFE7100, T_g=100ºC, d_n=3 mm, L=10 mm).
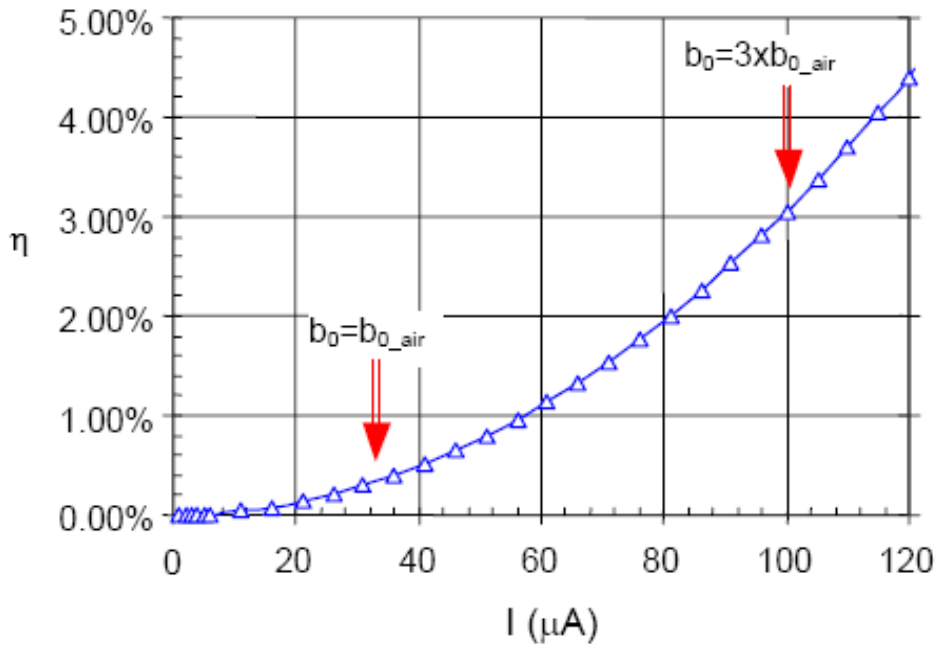
Figure 8. EGD efficiency against current (HFE7100, T_g=100ºC, d_n=3 mm, L=10 mm).
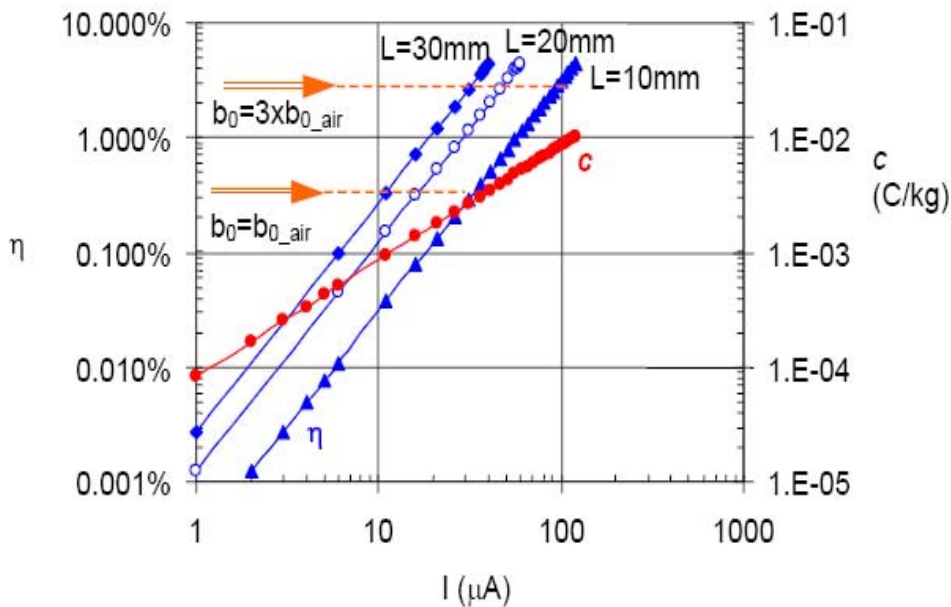
Figure 9. Efficiency and charge-to-mass ratio against current (HFE7100, T_g=100ºC, d_n=3 mm).
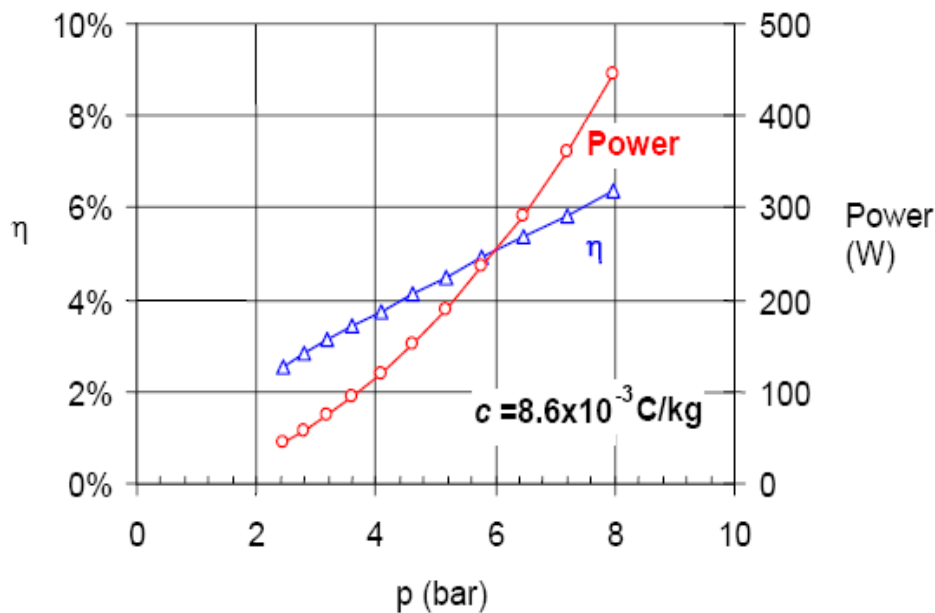
Figure 10. Efficiency and power at electrical breakdown for different pressures (HFE7100, $b_0=3xb_{0\_air}$, d_n=3 mm, L=10 mm).
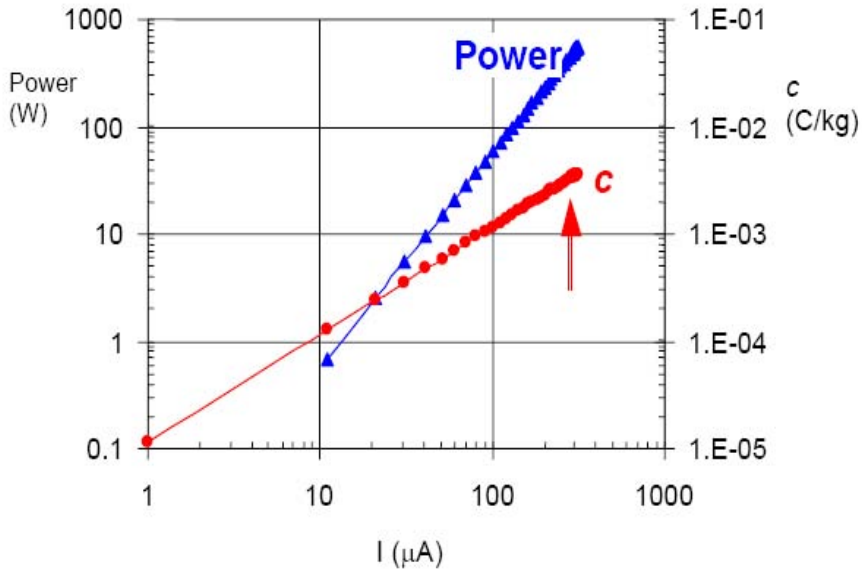
Figure 11. Power and charge-to-mass ratio against current (R134a, T_g=90ºC, d_n=3 mm, L=10 mm).
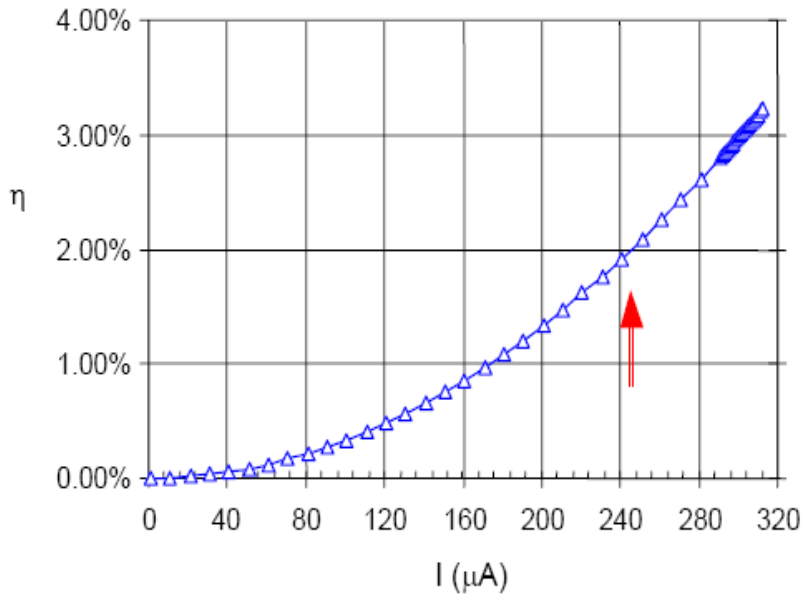
Figure 12. Efficiency against current (R134a, T_g=90ºC, d_n=3 mm, L=10 mm).
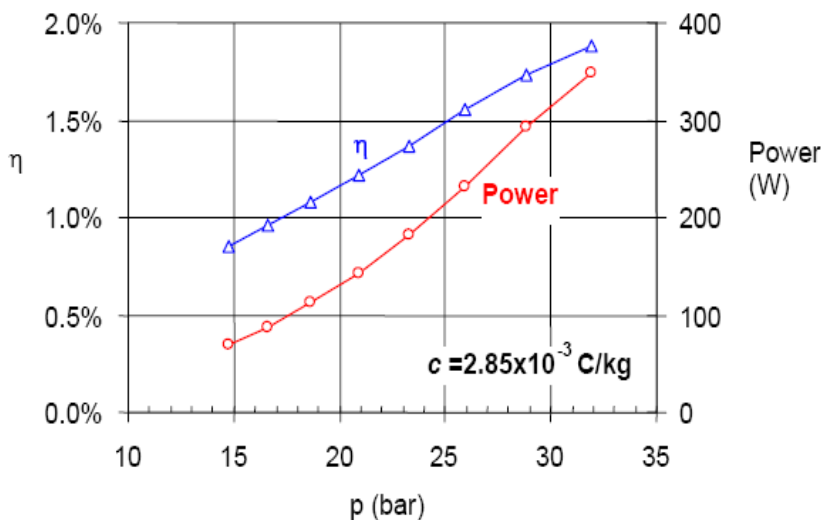
Figure 13. Efficiency and power at electrical breakdown for different pressures (R134a, $b_0=3xb_{0\_air}$, d_n=3 mm, L=10 mm).

The calculation results for refrigerant R134a as working fluid are given in figures 11-13. For a given generating temperature, R134a has a much higher density than HFE7100. When $b_0=b_{0\_air}$, electric breakdown will happen at the collector current of 247 μA for the calculation condition. This will allow an EGD power output of 350 W and an EGD efficiency of 1.9%.

As a conclusion of the results shown, in order to overcome the limits of electrical breakdown and charge-to-mass ratio on EGD conversion, the working fluid with larger vapour density and smaller latent heat of evaporation will be preferred, as indicated by equations (8) and (10). The calculation results show that HFE7100 is a much better EGD working fluid than steam, due to its larger density and smaller latent heat. For the same reason, R134a is better than HFE7100, although it has higher operating pressures. In addition, high collector voltage, large electric field strength and small mobility of charges will be necessary for a certain EGD efficiency required.

3. COMPUTATIONAL MODEL OF CONVERTER AND RESULTS

In the last few years, a number of publications available in the open literature dealt with computer simulation of charged particles in an electro-hydrodynamic flow field. However, these publications are limited to industrial processes (e.g. electrostatic precipitators, powder coating) where fluid velocities are relatively slow compared to EGD converter velocities. An overview on the existing modelling approaches can be found in [11]. The two most widely used methods for the secondary phase modelling - Lagrangian and Eulerian approach - were compared for different turbulent field conditions. Particle charging mechanism was taken into consideration.

One of the objectives of this work was to develop a Computational Fluid Dynamics (CFD) tool for the simulation of the major characteristics of fluid flow in such a device and to identify the most important factors in the power conversion process. As no work was done in the past on simulation of charged particles under high velocity flow conditions, this implies some major differences compared to available models and data, such as the assumption of incompressibility (which is no longer valid), and electric charge transport by convection (which is dominant, and may have an effect on particle charging). In the present work, an effort has been made to address these issues and assess EGD conversion. The importance of the particle/electric field interaction was evaluated. Turbulence effects were also taken into consideration.

A geometrical domain representative of an EGD converter was chosen for the CFD model, as shown in figure 14. The cross section of the converter is similar to a Venturi tube, with a relatively short converging section and a smooth diverging part on the flow downside of the device. The fluid enters under pressure on the inlet side of the converter. The particles receive electric charge from the corona electrode (of 1 mm diameter), positioned along the centreline between the inlet and the nozzle, because of the electric potential between the corona and attractor electrodes. The latter is situated at the nozzle throat. It was considered that the electrode has a negligible effect on gas flow; therefore, it was not explicitly modeled. The charged particles and the carrier gas stream are accelerated in the nozzle throat to sonic velocity. The electric charge can then be collected at several locations, indicated by four collector rings in the diffuser, as particles contact with the EGD wall. The fluid leaves the converter at low pressure through the outlet surface.
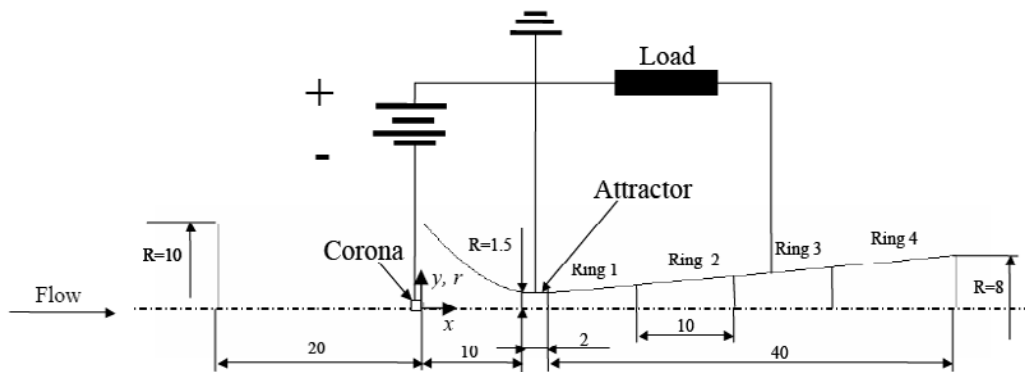
Figure 14. Schematic representation of the simulated EGD converter section.

The CFD model uses 3 different sets of equations: one set for the compressible gas flow, one for the charged particles and one for the electric field.

Gas flow was considered as compressible and turbulent, due to the high velocities in the EGD converter. Axi-symmetry was considered in the geometrical domain of figure 14. Velocity (vector), pressure and temperature are the major unknown variables, which are related through the set of partial differential equations (PDEs) expressing continuity, conservation of momentum and conservation of energy [12]:

$$\frac{D\rho}{Dt} - \rho \cdot div\, u = 0 \tag{11}$$

$$\rho \frac{Du}{Dt} = \rho\, \mathbf{g} - \nabla p + \nabla\, \tau_{ij} \tag{12}$$

$$\rho \frac{Dh}{Dt} = \frac{Dp}{Dt} + div(k\nabla T) + \tau_{ij} \frac{\partial u_i}{\partial x_j} \tag{13}$$

where τ_{ij} depend on viscosity and turbulence [12]. The turbulent behaviour was treated using the Reynolds averaging principle (RANS). In this work the standard $K\text{-}\varepsilon$ *model* was used, which added two more semi-empirical transport equations to the original set of PDEs. Their actual form for the turbulent kinetic energy (K) and the turbulent kinetic energy dissipation rate (ε) equations is well known and can be found in [13].

It was considered that the density variability in the carrier gas stream follows the ideal gas law and the physical properties of air were used. In order to solve the set of equations (11-13), proper boundary conditions must be applied. Symmetry conditions for velocity and temperature were used for r=0 (axis), zero velocity and heat fluxes at the walls, as well as an inlet pressure of 4 bar and an outlet pressure of 0 bar.

The set of PDEs (11-13), together with the two additional turbulence equations, can only be solved numerically. A commercial package, FLUENT [14], was used to simulate fluid flow in the EGD converter. In FLUENT, the space domain is subdivided into a number of small control volumes called finite volumes. For each finite volume each PDE is transformed into a set of algebraic equations and then solved using numerical techniques. Global iterations are performed until convergence is obtained for each unknown. For details of the finite volume method the reader is referred to [15]. Several mesh densities were examined from the finer to the coarser. The final structured mesh consisted of 18400 quadrangular cells.

Particle motion (transport), as a secondary phase, is usually modelled using the Eulerian or Lagrangian reference frames. In the first case, the particles are considered as a second continuum; therefore, another set of continuity, momentum, and energy equations must be solved. Coupling between the phases is provided, satisfying local mass balances and pressure. It is generally computationally less demanding. In the Lagrangian reference frame, each particle (stream) is individually considered. The trajectory is calculated using Newton's second law, integrating all existing forces over the particle for small time steps. Schmidt and Vogel [11] compared these two approaches for modelling particle flow in an electric field. It was concluded that Lagrangian particle tracking is superior for scientific investigation, while the Eulerian approach is, in most cases, suitable for practical purposes. In this work the Lagrangian model was applied, using FLUENT's discrete phase model. Several assumptions were made. The interaction between continuous and discrete phases was neglected, because of the relatively small volume fraction of the particles (0.3%). Particle/particle interaction includes collision and break up. Because of the specific operating conditions in the EGD converter (small particle diameter, short residence time), droplet collision and break up was not modelled in this work. Mass transfer between particles and continuous phase was also neglected.

Under these circumstances, there are only two relevant forces acting on the particle: drag and electrical forces. The equation of motion in the Lagrangian reference frame can be written as:

$$\frac{du_p}{dt} = f_D(u - u_p) + \frac{\mathbf{g}(\rho_p - \rho)}{\rho_p} + F_E \Big/ m_p \qquad (14)$$

where on the right hand side, the first and third terms represent the drag and Coulomb vector forces per unit mass, respectively. The factor f_D is related to the drag coefficient (C_D), particle diameter (d_p) and particle Reynolds number (Re_p) through the following expression:

$$f_D = \frac{18\mu}{\rho_p d_p^2} \frac{C_D \, Re_p}{24} \qquad (15)$$

C_D is also a function of the particle Reynolds number and an empirical formula for its calculation can be found in [14]. The term F_E in equation (14) can be calculated from the electric field vector (E) and the electric charge (C) carried by the particle as:

$$F_E = C \cdot E \qquad (16)$$

The effect of turbulent eddies on the transport of particles was taken into consideration. Equation (14) was integrated for small time steps using estimates for the instantaneous fluid velocity vector (u). It is common to model velocity fluctuations as a random process with a normal distribution, which has two parameters: mean and variance. The mean of the fluctuating velocity component is considered to be zero, which is consistent with the RANS turbulence model. The variance is scaled by the mean square of the average fluctuating velocity, estimated from local K and ε values [14]. In order to calculate particle trajectory, it is also necessary to estimate the time that a particle spends in a turbulent eddy (Lagrangian integral time scale, t^*). There are different methods to estimate it. In FLUENT, its value is estimated by:

$$t^* = C_L \frac{K}{\varepsilon} \qquad (17)$$

In equation (17) it is assumed that particles are small, and thus essentially travel with zero drift. C_L is a constant and has a significant influence on turbulent dispersion [16]. Unfortunately, its value is not well known: it should be determined experimentally. The published values vary in a significant range (0.1 to 0.6) [11]. In the present paper C_L=0.15 was used, which is near the value that most authors considered. From the Lagrangian time scale, the eddy lifetime can be directly estimated. For more details the reader is referred to [17]. It was also necessary to define boundary conditions for the discrete phase. A symmetry condition was applied at the centre line with zero normal particle velocity. For all the other boundaries, it was assumed that once an aerosol droplet reached the surface, it vanished. It is not completely true for the EGD converter wall, because there would be particles re-entering

the main stream or dragged towards the outlet. However, these particles would loose their electric charge, either to be collected either to be "lost" to the attractor electrode at the first contact. Therefore, their further trajectory is not interesting from the electrical conversion point of view.

The final set of equations concerns the electric field. The electric field in the presence of discrete charged particles is governed by the Poisson equation – equation (5). The particle charge density, δ, can be calculated from the particle mass flow rate ($\dot{m}_i$) and the particle charge-to-mass ratio (c):

$$\delta = \frac{c}{V_{cell}} \sum_{i=1}^{n} \dot{m}_i \left(t_{out} - t_{in} \right) \qquad (18)$$

where V_{cell} is the volume of the computational cell and n the number of particle streams. t_{in} and t_{out} refer to the times a particle enters or exits a cell, respectively.

The particle charge-to-mass ratio depends on the charging kinetics with unipolar electric field. The charging mechanism is usually divided in two mechanisms – field and diffusion charging. In this work, the charging mechanism was not modelled, and the charge-to-mass ratio was considered to be a constant value.

The relationship between the electric field and electric potential is given by:

$$E = \nabla V \qquad (19)$$

Equations (5) and (14) were iteratively solved. Constant electric potential was applied as boundary condition on the collector and corona electrodes. Zero potential was considered for the attractor, while zero flux ($E=0$) was used on the other boundaries. The electric current can be calculated from the current density (J). The current density is described by the following equation:

$$J = \sigma E + \delta \overline{u}_p \cong \delta \overline{u}_p \qquad (20)$$

where σ is the electric conductivity and $\overline{u}_p$ is the mass average particle velocity. In equation (20) the first and second term on the right hand side represent the current density by conduction and convection, respectively. It was assumed that conductive current was negligible because of the ion drift velocity (50-100 m/s [18]) being considerably lower than the gas velocity [9]. The electric current on a boundary surface was obtained by integrating the current density vector with respect to the boundary area vector. For more details see [17].

CFD model results will be presented next. Velocity contours of the gas flow inside the converter are shown in figure 15. The gas enters the converter at about 8 m/s and accelerates in the nozzle to Mach=1. The maximum velocity is then reached at about 1.5 cm downstream of the throat (Mach $\approx$ 2.7), where wall separation appears. From this point on, the fluid velocity decreases. The total mass flow rate was 6.5×10^{-3} kg/s, which agrees well with experimentally measured values under similar conditions (6.8×10^{-3} kg/s) [19]. Sample particle trajectories, considering turbulent dispersion, are shown in figure 16. The particles were

injected at 0 m/s, and were quickly accelerated to the fluid velocity, because of the small particle inertia. A significant portion of the charged particles gets trapped on the collector wall. Maximum residence times in the converter were around one microsecond.

Several simulations were performed in order to evaluate the effect of the imposed electric field on particle trajectory, for various operating conditions. In each case the pressure difference between inlet and outlet was kept at 4 bar. The corona potential in all cases was –5 kV, located at the inlet of the converging section. The particles were injected at nozzle inlet, and turbulent dispersion was taken into account. Because of the random nature of turbulent dispersion, a large number of particle streams were tracked (49000). The collector was located at ring 2 (see figure 14) and the electric potential varied from relatively small (–10 kV) to relatively high (–50 kV) values. Two particle diameters were tested: 1 μm and 100 μm. The results for the proportion of the particles collected (escaped) at the diverging part and outlet of the converter are summarised in table 1. It can be seen that particle diameter influenced particle fate considerably. Increasing the particle diameter increases by a factor of two the proportion of particles "collected" on the collector rings ($\approx$70 % vs $\approx$35%). Charge to particle mass ratio had no effect on the proportion of the particles collected within the considered range, probably because drag force was dominating. Therefore, the collected current can be considered directly proportional to the charge to mass ratio. It is worth mentioning that there is an upper limit for c before electric breakdown occurs.
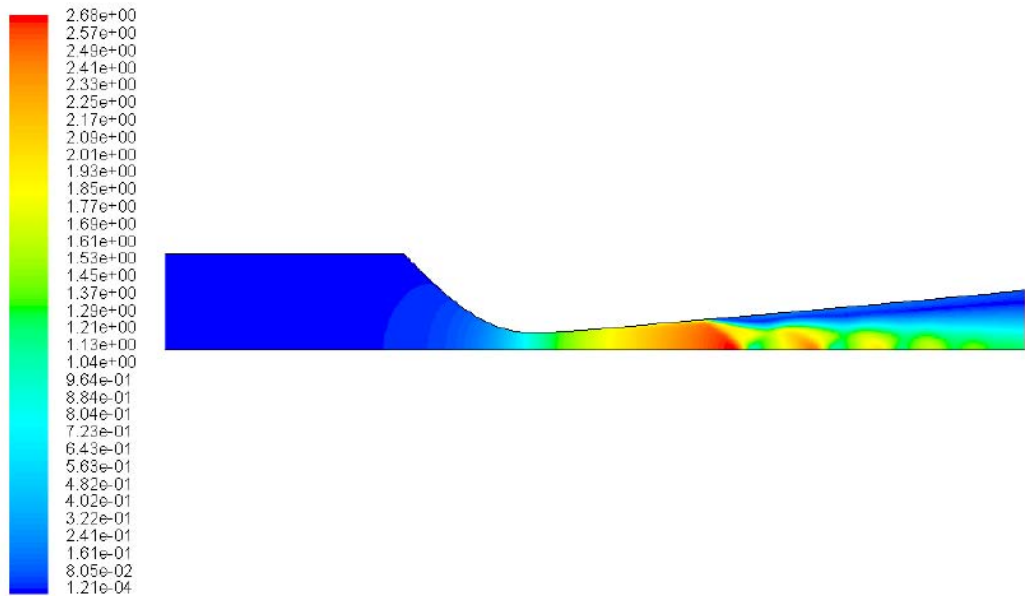
Figure 15. Mach number contours in the EGD converter.
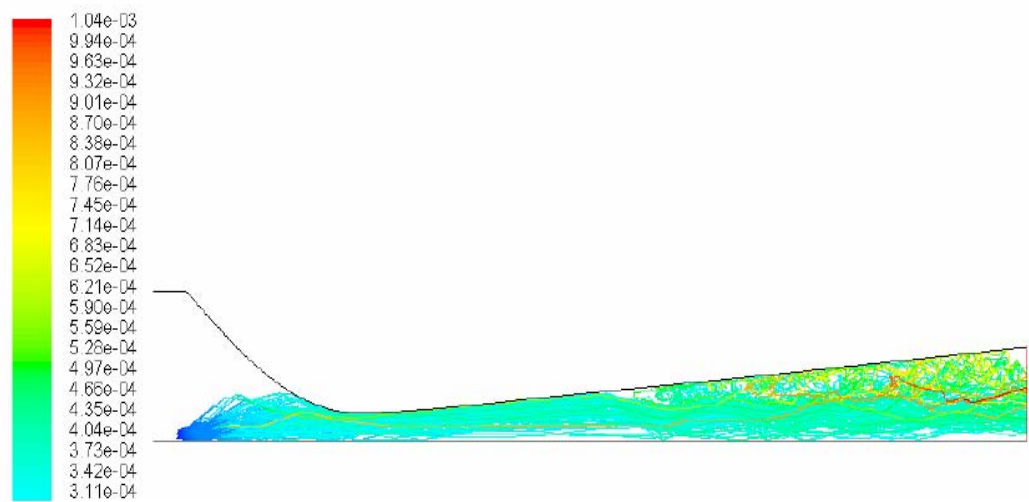
Figure 16. Particle trajectory calculation with turbulent dispersion, scaled by residence time (in seconds).
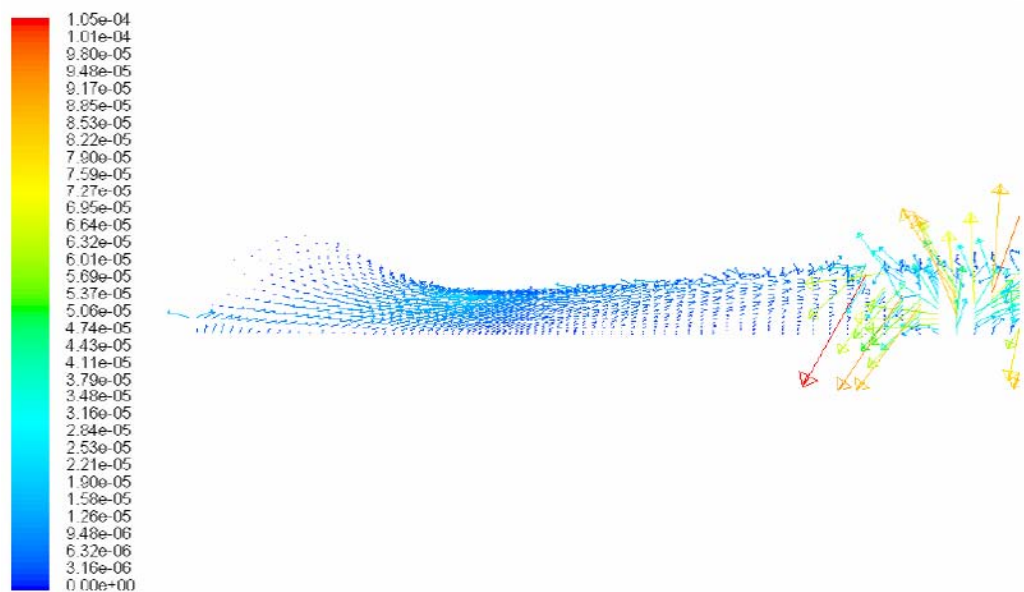
Figure 17. Drag force vectors on droplets (in N).

Table 1. Proportion of injected/trapped (escaped) particles at different locations of the converter.

	ref_1	ref_2	V_1,d_1,c_1	V_1,d_2,c_1	V_2,d_1,c_1	V_2,d_2,c_1	V_3,d_1,c_1	V_3,d_2,c_1	V_2,d_1,c_2	V_2,d_2,c_2
Ring 1	0.100	0.078	0.036	0.080	0.037	0.078	0.037	0.078	0.035	0.081
Ring 2 (collector)	0.063	0.100	0.063	0.099	0.063	0.101	0.062	0.100	0.062	0.099
Ring 3	0.111	0.172	0.109	0.175	0.112	0.172	0.108	0.171	0.110	0.176
Ring 4	0.099	0.135	0.103	0.133	0.099	0.135	0.099	0.135	0.100	0.134
Outlet	0.643	0.293	0.645	0.294	0.642	0.294	0.648	0.291	0.653	0.288

- ref: reference conditions, particle trajectory calculated with no electric field (1- $d_p=1$ μm; 2- $d_p=100$ μm)

- V: collector potential (1- V= -10 kV; 2- V= -20 kV; 3- V= -50 kV)

- d: particle diameter (1- $d_p=1$ μm; 2- $d_p=100$ μm)

- c: charge/mass ratio (1- $c=10^{-4}$ C/kg; 2- $c=10^{-2}$ C/kg)

It is clear from table 1 that the applied collector potential (electric field) had no influence on particle trajectories, due to the high carrier gas velocities. The dominating force acting on the particles was drag, with the Coulomb force playing a secondary role. Figure 17 shows drag forces acting on the droplets. Especially high values of drag were observed in locations where the fluid shows high rates of acceleration and deceleration (e.g. injection point, near wall). Some similar results were found in the literature, although the carrier gas velocities were much lower. An analysis made indicated that the electric field altered particle trajectories for an average inlet velocity as low as 0.2 m/s, that is, 40 times smaller than in the present simulation study.

The presence of charged particles altered the electric field inside the conversion section (space charging effect). Figure 18 shows the iso-potential lines with charged particles. Simulation conditions were 0.0007 kg/s particle flow rate, 0.01 C/kg charge-to-mass ratio, which correspond to a total current of 7 μA. Collector and corona potential were -20 kV and -5 kV, respectively. It can be seen that iso-potential lines are denser near the corona and attractor indicating a larger electric field (divergence of the potential) for the charged particle.
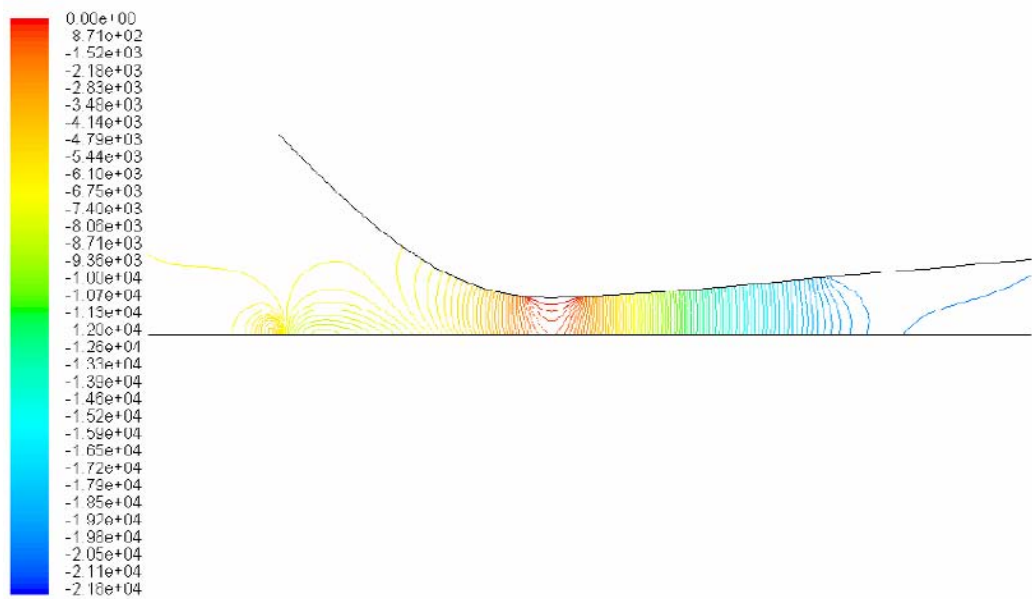
Figure 18. Iso-electric potential lines in the EGD converter with charged particles (in V).
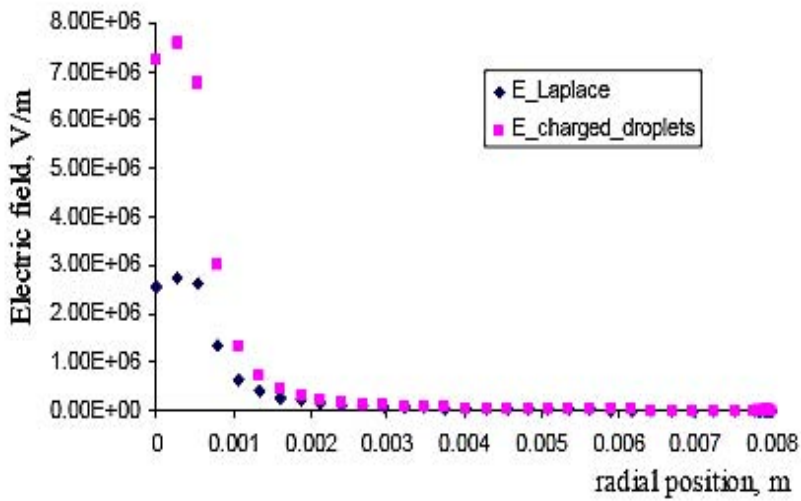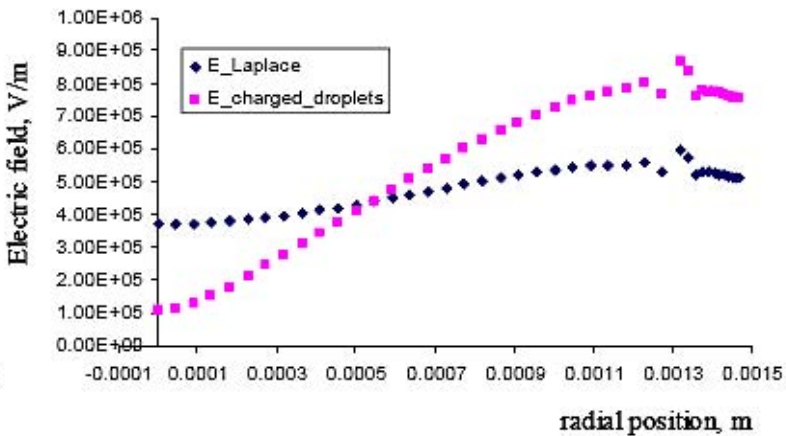
Figure 19. Electric field with and without charged droplets along the radial direction at the corona electrode (A) and at the throat (B).

The effect of the charged particles on the electric field (E) can be analysed through figure 19, where E is plotted at two different axial positions, with and without charged droplets. As can be seen, the electric field strength is larger by about a factor of two at both the corona and attractor electrodes, in the presence of discrete charges. The knowledge of local electric field strength is very important. Above a certain value, which depends on fluid properties and pressure, as seen before electric breakdown occurs, which represents a limit for EGD conversion.

**Table 2. Electric current as a function of charge-to-mass ratio (*c*) at
different locations in the EGD converter**

Charge/mass ratio (*c*)	Electric current (μA)				
	ring 4	ring 3	ring 2	attractor	outlet
10^{-3} C/kg	0.05	0.06	0.03	0.01	0.09
10^{-2} C/kg	0.50	0.57	0.23	0.10	0.87
10^{-1} C/kg	4.91	5.78	2.41	0.96	8.66

The electric current at different locations, for different particle charge-to-mass ratios, is shown in table 2. It can be seen that increasing *c* by a factor of ten, the collector current increases proportionally. The best collector locations seem to be at rings 3 and 4, and the worst at ring 2. Experimentally determined currents, under similar conditions [19], were in the range of 10 μA, that is higher than simulated values. The charge-to-mass ratio was not experimentally determined; its actual value might be higher than the cases considered in table 2.

Electric conversion efficiency of an EGD converter is generally low for air, as seen in section 2. Calculated (CFD model) EGD conversion efficiencies can be considered in agreement with the range published values.

The main conclusions of the CFD model are: (a) gas flow is choked in the throat of the EGD converter; (b) after the throat the velocity gradually increases to a value as high as Mach=2.7 at about 1.5 cm in the downstream direction, where boundary layer separation occurrs; (c) because of the high velocities, electric field has a negligible effect on particle trajectories; (d) the most important factors in collector efficiency are particle diameter and parameters in the turbulence model; (e) increasing particle diameter might improve collection efficiency, however decreasing charging efficiency, with optimal values still to be found. It could also be concluded that the electric field is highly altered by the presence of charged particles, compared to a Laplace field. This information is important to determine maximum particle mass load before electric breakdown occurs between corona and attractor electrodes, which represents an upper limit for EGD conversion. The collector current was directly calculated from the current density at different locations in the EGD converter, for three different charge-to-mass ratios; the results suggest that the best collector location is about 2 cm from the nozzle, and the worst just downstream of the nozzle. The low process efficiency could be improved by optimisation of flow conditions, by applying fluids with better electrical properties, and by multi-staging of the conversion process. The model presented can be considered as a good tool for optimisation purposes.

4. EXPERIMENTAL RESULTS FOR AN EGD CONVERTER

In order to investigate the EGD conversion process, an EGD nozzle was designed, built and experimentally tested – figure 20 shows a view of the experimental facility. The nozzle comprised a nylon housing and a sliced duct set, as shown in figure 21. The diverging section of the nozzle included 5 slices of 10 mm thick metal or ring. A metal ring was used so the

collector electrode and the collector could be put at 5 different positions. A long needle could also be used as the collector to be compared with the ring. The metal throat served as the attractor electrode. The corona needle was hold by a tube and positioned in the centre of the converging section. A liquid could be injected through the needle to the working gas. This design allowed an easy change of the nozzle configuration and dimensions. In the experiments, 3 mm straight throat and tapered throat, 4 mm straight, various throat and collector positions were tested. The EGD nozzle assembly was connected with a generator, a condenser and a circulation pump to form a test rig. Both the generator and condenser were plate heat exchangers.
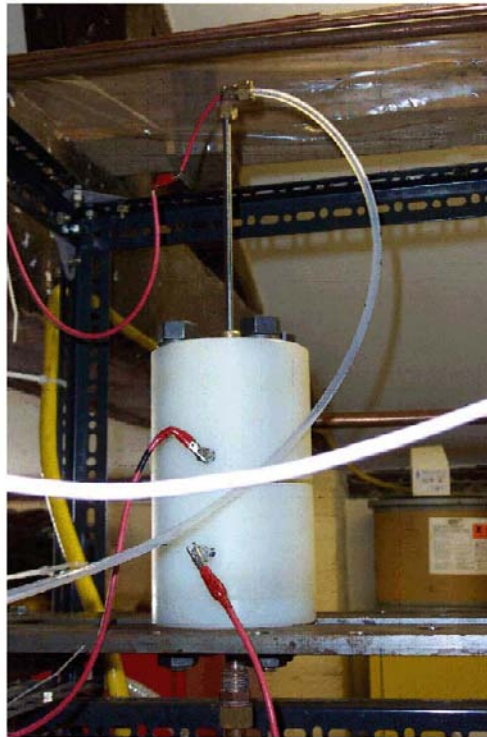
Figure 20. View of EGD converter experimental apparatus.
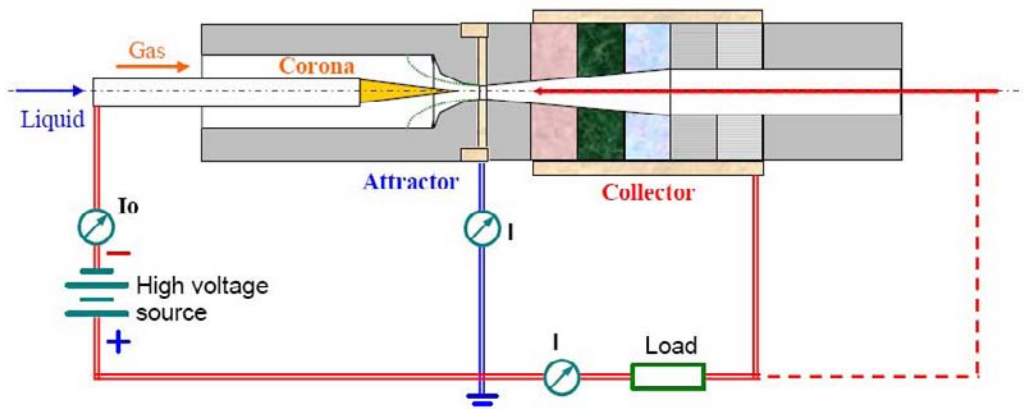
Figure 21. Schematic representation of sliced EGD nozzle.

Compressed air was used to test the performance of the EGD nozzle and to determine its proper configuration. Results showed that there is a slight effect of changing the ring collector position on the collector current. However, changing the needle collector position has a larger effect on the collector current, as shown in figure 22. The collector current decreases with increasing conversion duct length when L>10 mm. A larger charge was lost to the duct wall when the conversion duct was longer, this resulted in a larger attractor current. A conversion duct length of 5 to 10 mm would be suitable for a needle as the collector. However, the ring collector had a more stable charge collection.
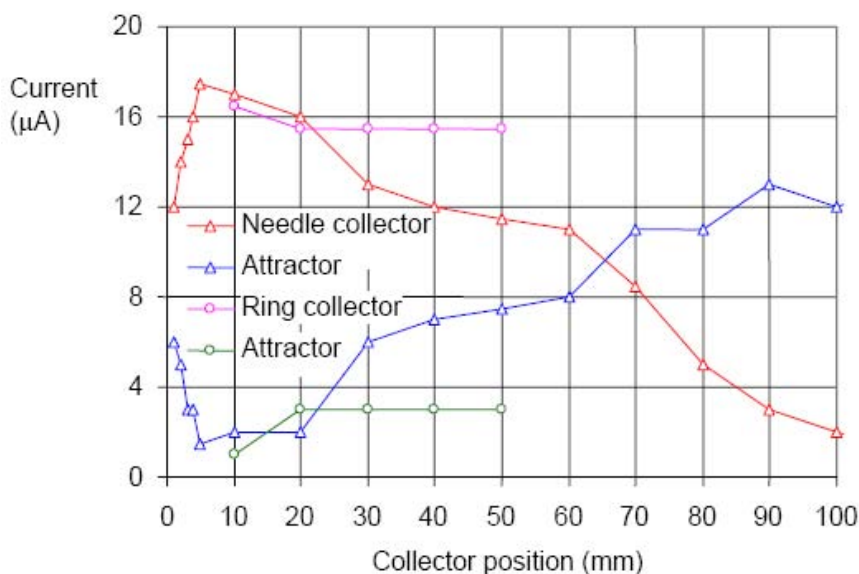
Figure 22. Collector current against collector position (Air, 4 bar, d_n=3mm, Load=1 Gohm, I_o=20 µA).

In the following tests, the ring collector at L=10mm was used. Increasing the input current did not cause a significant rise in collector current. Electric breakdown was observed for a larger input current. A power output of 0.3 W was achieved, at around 1 Gohm external load.

Tests of electric breakdown were carried out for three different reasons: (a) the data of dielectric strength is very limited for refrigerants; (b) EGD electrode configuration (ring or needle) is different from those (e.g., plate, cylinder and sphere) in the existing theory of dielectric strength; (c) to confirm happening of electric breakdown in the EGD converter. The device was comprised of two ring electrodes spaced by a 2.54 mm ring Plexiglas. High voltage was applied between the two electrodes. Breakdown voltage was taken when the current started to jump. It was found that the breakdown voltage for the existing EGD configuration was smaller than the values in the references [20], as shown in figure 23. When the experimental dielectric strength was used to calculate EGD conversion, a good agreement between experiment and calculations (section 2) was obtained. Figure 24 gives the breakdown voltage for 3 refrigerants. Their dielectric strength is less than 3 times that of air.

When these experimental dielectric strength values were used to calculate EGD conversion, the results showed smaller EGD power than those in section 2. Figure 25 shows new calculations for 4 different refrigerants.
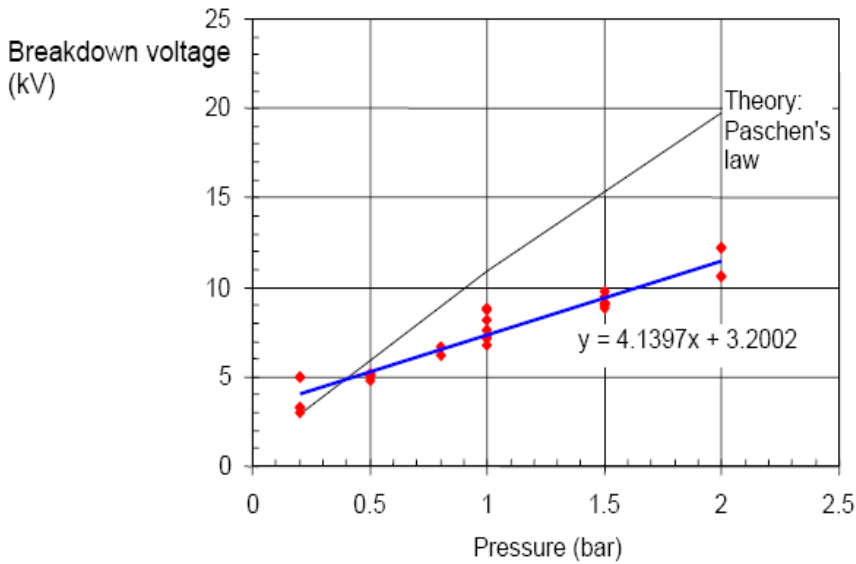
Figure 23. Electric breakdown voltage for air (theory [20] and experiment).
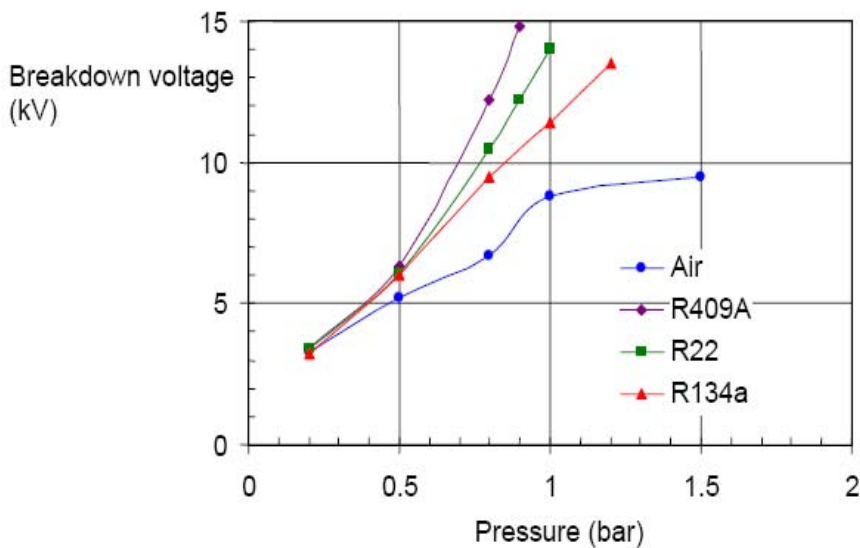
Figure 24. Experimental electric breakdown voltage for different refrigerants and air.
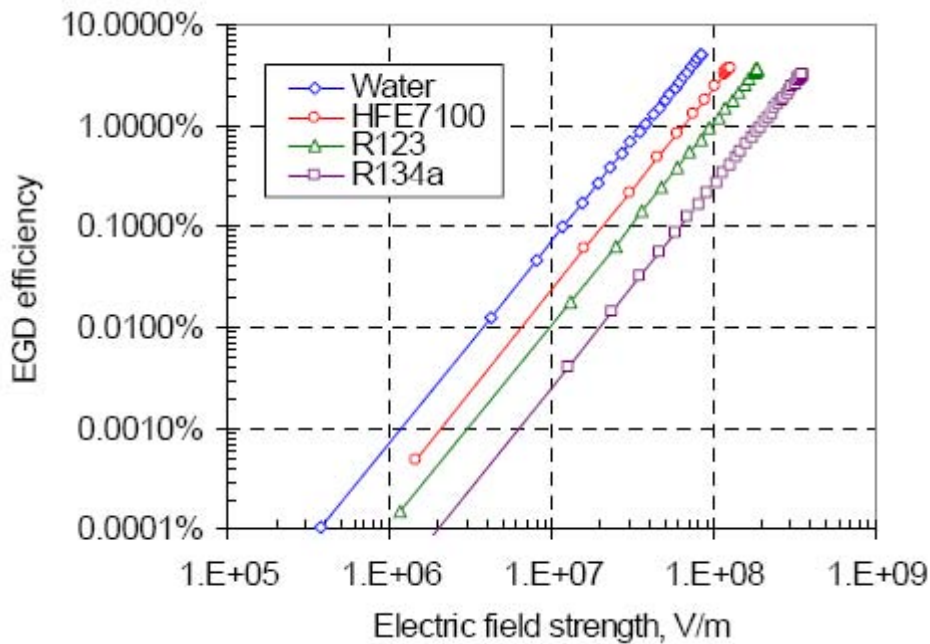
Figure 25. Calculated EGD efficiency versus electric field strength for different refrigerants (T_g=90ºC, T_c=30ºC, d_n=3 mm, L=10 mm).

Due to the requirement in mobility of charge, liquid droplets were considered as an efficient charge carrier in EGD conversion. However, the tests using liquid spray in the EGD nozzle showed a very limited effect on the working current. A test was therefore arranged to determine how much current could be given by a single electrospray. The principle of electrospray consists in a liquid from a capillary being split into fine droplets by an applied electrical force; charges are attached to the surface of droplets. Testing was carried out with a distance of 10 mm between the needle and plate, with a voltage applied between them. The water was sprayed through the needle. Effective electrospray current was 1~3 μA for a flow rate of 2 ml/minute. It became clear that the current by electrospray was very limited. For example, 100 ml/min only gave 3μA. Also, the mobility by electrospray is too large, which doesn't recommend it as an option for EGD operation.

5. CONCLUSION

Electric power conversion in an Electro-Gas-Dynamic (EGD) converter is accomplished by charging discrete particles using a corona electrode and transporting these particles at high velocity against an electric field. As electric potential builds up, useful electric power can be collected. An important advantage of such a device is that it doesn't include moving parts and so requires very little maintenance.

As seen, EGD conversion is subjected to two limits: electric breakdown and charge-to-mass ratio. The limit of electric breakdown may be overcome by using low boiling point

working fluids. For example, R134a has a much higher EGD efficiency than the other working fluids analysed, which could reach 2% at electric breakdown. For a given charge-to-mass ratio, R134a also has a higher EGD efficiency than others. The charge-to-mass ratio limit can be mitigated by using high pressure working fluids. It was found in the calculations that if a longer EGD conversion duct is used, the charge-to-mass ratio will be smaller for a required EGD efficiency.

A Computational Fluid Dynamics (CFD) model allowed to simulate the major characteristics of fluid flow in such a device and to identify the most important factors in the power conversion process. The importance of the particle/electric field interaction was evaluated, taking into account turbulent effects. It was concluded that the electric field has a negligible effect on particle trajectories and that the most important factors in collector efficiency are particle diameter and parameters in the turbulence model. It could also be concluded that the electric field is highly altered by the presence of charged particles, compared to a Laplace field. The results suggest that the best collector location is about 2 cm from the nozzle, and the worst just downstream of the nozzle. The low process efficiency could be improved by optimisation of flow conditions, by applying fluids with better electrical properties, and by multi-staging of the conversion process. The model presented can be considered as a good tool for optimisation purposes.

In order to investigate the EGD conversion process, an EGD nozzle was designed, built and experimentally tested. Tests of electric breakdown were carried out for different working fluids (refrigerants). It was found that the breakdown voltage for the existing EGD configuration was smaller than the values in the references. When these experimental dielectric strength values were used to calculate EGD conversion, the results showed smaller EGD power than previously calculated. The tests using liquid spray in the EGD nozzle showed a very limited effect on the working current. It was also found that the use of electrospray is not suitable for EGD operation.

The ability of this technology to produce electricity needs improvements. In spite of its low efficiency and limited power output, it can nevertheless be interesting when taking into account that it may be driven by waste heat or renewable energy sources (to obtain the incoming vapour). Generalisation of the concept of distributed power generation, as well as combined heat and power systems, which allow a significant reduction of energy consumption and harmful emissions to the environment, will put this technology in a different level. For those applications, no large power outputs are needed, and combination with available heat output may be economically justifiable. The capital cost of the technology associated with this equipment is expected to fall below the cost of conventional alternatives.

ACKNOWLEDGMENTS

REFERENCES

[1]   Marks, A.M. (1953). Heat Electrical Power Conversion Through the Medium of a Charged Aerosol. USA Patent 2,638,555.

[2]   Marks, A.M. (1983). Electrothermodynamic (ETD) Power Converter. USA Patent 4,395,648.

[3]   Marks, A.M. (1987). Electrothermodynamic Power Converter With Converging Flows. USA Patent 4,677,326.

[4]   Marks, A.M. (1984). Electrohydrodynamic equations of a charged aerosol generator. *Proceedings of the 19th Intersociety Energy Conversion Engineering Conference*. San Francisco, CA, USA, pp.43-51.

[5]   Marks, A.M. (1980). Charged Aerosol Generator With Uni-Electrode Source. USA Patent 4,206,396.

[6]   Marks, A.M. (1969). *Gas Flywheel Power Converter.* USA Patent 3,456,135.

[7]   Marks, A.M. (1968). Heat-Electrical Power Interconversion Devices. USA Patent 3,417,267.

[8]   Borok, A.M. (1970). Electrogasdynamic generator with neutralization of the space charge. *High-Temp*, 8 (1), pp.166-170.

[9]   Lawson, M. and Ohain, H.V. (1971). Electrofluid dynamic energy conversion present status and research areas. *Trans. ASME A, Journal of Engineering for Power*, 93, pp.201-221.

[10]   Riffat, S.B. and Su, Y.H. (2002). Integration of An Electrogasdynamic Power Converter Into Ejection Refrigeration Systems. *Proceedings of 1st Sustainable Energy Technologies Conference*. paper RHP18. FEUP, Porto, Portugal.

[11]   Schmidt, H.-J. and Vogel, L. (2003). On the modelling of the particle dynamics in electro-hydrodynamic flow fields: I. comparison of Eulerian and Lagrangian modelling approach. *Powder Technology*, 135-136, pp.118-135.

[12]   White, F.M. (1991). *Viscous Fluid Flow*. 2nd ed., McGraw Hill.

[13]   Launder, B.E. and Spalding, D.B. (1974). Numerical computation of turbulent flows. *Computer Methods in Applied Mechanics and Engineering*, 3 (2), pp.269-289.

[14]   Fluent Inc. (2001). *Fluent 6.0 User's Guide*. New Hampshire (USA).

[15]   Versteeg, H.K. and Malalasekera, W. (1995). *An Introduction to Computational Fluid Dynamics: The Finite Volume Method*. Prentice Hall.

[16]   Varga, S. and Oliveira, A.C. (2004). CFD Simulation of Fluid Flow in An Electrogasdynamic Power Converter. *Proceedings of 3rd Combined Heat and Powered Cycles Conference*. Larnaca, Cyprus, October 11-13.

[17]   Varga, S. and Oliveira, A.C. (2006). Simulation Study of An Electrogasdynamic Power Converter Using CFD. *International Journal of Low Carbon Technologies*, 1 (3). Manchester University Press, pp.245-261.

[18]   Schmidt, H.-J. (2003). On the modelling of the particle dynamics in electro-hydrodynamic flow fields: II. Influences of inhomogeneities on electrostatic precipitation. *Powder Technology*, 135-136, pp.136-149.

[19]   Riffat, S.B. and Su, Y.H. (2004). Design And Testing Of An Electrogasdynamic Power Generator For EGD Ejection Systems. *Proceedings of 3rd Combined Heat and Powered Cycles Conference*. Larnaca, Cyprus, October 11-13.

[20] Paschen, F. (1889). Über die zum Funkenübergang in Luft, Wasserstoff and Kohlensäure bei verschiedenen Drücken erforderliche Potentialdifferenz. *Weid. Annalen der Physick*, 37: pp. 69-75.

NOMENCLATURE

A   EGD converter section area (m^2)
b_0 standard electric breakdown strength ($V.m^{-1}$)
C   electric charge (C)
C_D drag coefficient (-)
C_L constant of turbulent dispersion (-)
c   charge-to-mass ratio (C/kg)
d   diameter (m)
E   electric field strength ($V.m^{-1}$)
F_E electric force (N)
f_D factor related to drag coefficient (s^{-1})
g   gravity acceleration ($m.s^{-2}$)
h   specific enthalpy ($J.kg^{-1}$)
I   current (A)
J   current density ($A.m^{-2}$)
K   specific turbulent kinetic energy ($m^2.s^{-2}$)
k   fluid thermal conductivity ($W. m^{-1}.K^{-1}$)
L   Length of EGD conversion section (m)
m   mass (kg)
$\dot{m}$ mass flow rate ($kg.s^{-1}$)
p   pressure ($N.m^{-2}$, bar)
Q   heat transfer rate (W)
Re Reynolds number (-)
r   radius (m)
T   temperature (ºC)
t   time (s)
u   velocity ($m.s^{-1}$)
V   electric potential (V)
V_{cell}volume of computational cell (m^3)
x   space coordinate (m)

Greek Symbols

Δ difference
δ charge density ($C.m^{-3}$)
ε turbulent kinetic energy dissipation rate ($m^2.s^{-3}$)
ε_0 dielectric constant of free space ($=8.85 \times 10^{-12} F.m^{-1}$)
ε_r relative dielectric constant

η efficiency of EGD conversion (-)

λ mobility ($m^2.s^{-1}.V^{-1}$)

μ viscosity ($N.s.m^{-2}$)

ρ density ($kg.m^{-3}$)

σ electric conductivity ($A.m^{-1}.V^{-1}$)

τ viscous stress ($N.m^{-2}$)

Subscripts

0   state point at the throat of the nozzle

1, *2* state points at EGD converter inlet and outlet

b   breakdown

c   vapour condenser (after converter)

g   vapour generator (boiler)

i, *j* generic space coordinates (2D)

in   inlet

n   nozzle

out outlet

p   charged particle

s   isentropic process

Acronyms

CFD   computational fluid dynamics

EGD   electrogasdynamic

PDE   partial differential equation

Chapter 5

NEW TRENDS ON LOAD MODELLING AND FORECASTING: MODELS AND TECHNIQUES

Lia Toledo Moreira Mota[*1] *and Alexandre de Assis Mota*[2]

[1] State University of Campinas, Av. Albert Einstein, 400 – CP 6101,
DSEE/FEEC/UNICAMP – Campinas – SP – Brazil
[2] Pontifical Catholic University of Campinas, Rodovia D. Pedro I,
km 136 - Parque das, Universidades CEP: 13086-900 – Campinas – SP – Brazil

ABSTRACT

Load modeling has a significant impact on power systems operation, simulation and analysis. However, little attention has been paid to develop adequate load models and forecasting methods when compared to the effort spent with other power systems related problems. In this context, this chapter presents a review on load models and load forecasting techniques, and also discusses the new trends on these issues. These late tendencies include bottom-up and top-down approaches, gray box identification techniques and the use of fuzzy logic, among others. The discussion emphasizes the main concepts of each method. Especially in distribution networks, load modeling and forecasting may cause some inconvenient because, in order to monitor the residential consumption (in terms of total active power and the type of appliance), many measuring equipment must be installed, leading to undesired costs. This is also troublesome to the residents of the household. In order to minimize these inconvenient, non-intrusive load modeling and forecasting techniques must be applied. This new methodology is also presented and examined in this chapter, in a combination of a bottom-up approach with object-oriented programming techniques.

1. INTRODUCTION

Load modeling plays a critical role on power systems operation and studies, since the results of important analysis functions are strongly dependent on the models assumed. For

---

[*] email: amota@puc-campinas.edu.br

example, in voltage stability assessment, it influences the determination of the stability margin; in power flow studies, the resulting operation point depends on the adopted load model; the state estimation function is strongly affected by load parameters; the usage of inadequate models to represent the load behavior in post-blackout conditions can lead the system to other outages and finally, for expansion planning studies, the forecasting of the load is of great importance, because small errors can represent the loss of millions of dollars. As the decisions related to the system reinforcement and/or expansion are based on the results of these studies, accurate models must be assumed for all the electrical network components, including the system loads.

Despite its significance, few attention has been paid to the load modeling and forecasting area [1]. This fact can be possibly explained by the difficulties associated to the development of accurate load models, as the lack of information about the load composition, the large number of different load components, the problem of accessing load devices located inside residences, among others. Consequently, the loads are usually poorly modeled as constant power injections. However, this approach may be not acceptable in many situations, especially when the system dynamics must be considered. Besides, the growing of the power system demand, which is usually not followed by the corresponding power generation growth, reduces the system operation margins and consequently, more precise load models are required. These aspects show the relevance of developing more accurate load models and forecasting techniques, since they can provide a more realistic view of the power system under analysis, helping to avoid, for example, stability problems and voltage collapses.

In this context, first, this chapter presents a brief review on the load models and forecasting techniques commonly used in power system analysis. In the sequence, new trends on these issues are introduced, followed by a discussion about the tendencies on the load model and forecasting area. Finally, some conclusions on this subject are also presented.

2. REVIEW ON LOAD MODELLING AND FORECASTING METHODS

Basically, there are two main approaches in load modeling: the bottom-up and the top-down approaches. The bottom-up approach builds up the load model based on information associated to the individual components of the load of a particular bus. The top-down approach forms the load model from information related to the aggregated behavior of all components of the load bus. These two different approaches can be developed using models to represent the dynamic behavior of the components or system measurements with identification techniques. In the following, the models normally used in power system analysis are presented. It is important to note that they can be used to build a model for the load bus, using the top-down or the bottom-up approaches. This means that a specific model can be used to develop the model of individual electrical devices and other load components (bottom-up approach) or to build the model to represent the aggregated behavior of a load bus (top-down approach).

2.1. Mathematical Load Models

A mathematical load model can be defined as the mathematical representation of the relation between the power (active or reactive) and the voltage in a load bus. The models commonly found in the literature are described in the sequence [2].

2.1.1. Static Models

These models can be used to represent static loads, as for example, resistive loads, or as an approximation of dynamic loads as induction motors. In this representation, the power (active or reactive) is defined as a function of the voltage magnitude and/or the system frequency for the same instant. Examples of this type of model are described in the following.

a.   Constant Impedance Model:

In this model, the power of the load bus directly varies with the square of the voltage magnitude ($P = f(V^2)$) and this is the reason why this model is called constant impedance. This model is also called a constant admittance model.

b.   Constant Current Model:

In this case, the power of the load bus directly varies with the voltage magnitude, justifying the name adopted for the model ($P = f(V)$).

c.   Constant Power Model:

In this model, the power of the load bus does not vary with the voltage magnitude. This model is widely used in almost all power system analysis functions ($P = constant$).

d.   Polynomial Model:

In this representation, the power and the voltage of the load bus are related by polynomial equations of the following type:

$$P = P_0.[a + b.(V/V_0) + c.(V/V_0)^2]$$

$$Q = Q_0.[d + e.(V/V_0) + f.(V/V_0)^2],$$

(1)

where: a, b, c, d, e and f are parameters of the model. Normally, P_0, Q_0 and V_0 are the initial conditions of the system under analysis. However, if this model is representing a specific load, P_0, Q_0 and V_0 are, respectively, the nominal values of the active and reactive powers and the voltage. Its is important to emphasize that this model is also called ZIP model, since the components of the polynomial equation represent a constant power (P), a constant current (I) and a constant impedance (Z).

e.   Exponential Model:

In this model, the power and the voltage of the load bus are related by exponential equations of the following type:

$$P = P_0.(V/V_0)^a$$

$$Q = Q_0.(V/V_0)^b,$$

$$(2)$$

where *a* e *b* are the model parameters. The same considerations about P_0, Q_0 and V_0 can be adopted for this model. Sometimes, two or more components with different exponents are included in these equations. It can be easily noted that one can derive a polynomial model from these equations, if they are described by the sum of three exponential components with exponents 2 (constant impedance), 1 (constant current) and 0 (constant power).

f.   Frequency Dependent Model:

This model represents the relation between the power and the frequency of the load bus. The equation that describes the exponential or polynomial model is multiplied by a factor that depends on the bus frequency. This factor can be represented by the following equation:

$$Factor = [1 + a_f.(f - f_0)], \qquad\qquad (3)$$

where a_f is the frequency sensitivity parameter; *f* is the frequency of the bus voltage and f_0 is the nominal frequency.

However, it is important to emphasize that if the ZIP model is used combined with the frequency dependent model, this means the equations that represent the ZIP model multiplied by the factor described in Eq. (3), the obtained load model has no physical meaning anymore, since the component related to the constant impedance is now dependent on the frequency.

g.   EPRI LOADSYN Model:

This model is used in the EPRI LOADSYN computational program that is employed in many facilities to assess the behavior of electrical power systems [3]. This model comprises the characteristics described in the previous models (polynomial, exponential and frequency dependent) and can be represented by the following equations:

$$P = P_0.\{Pa_1.(V/V_0)^{KPV1}.[1 + KPF1.(f - f_0)] + (1 - Pa_1).(V/V_0)^{KPV2}\}$$

$$Q = P_0.\{Qa_1.(V/V_0)^{KQV1}.[1 + KQF1.(f - f_0)] + $$
$$+ ((Q_0/P_0) - Qa_1).(V/V_0)^{KQV2}.[1 + KQF2. (f - f_0)]\},$$

$$(4)$$

where:

- Pa$_1$ is the load component (active power) that depends on the frequency;
- KPV1 is the voltage exponent of the load component (active power) that depends on the frequency;

- KPV2 is the voltage exponent of the load component (active power) that does not depend on the frequency;
- KPF1 is the frequency sensitivity coefficient of the load (active power);
- f is the frequency of the bus voltage;
- f_0 is the nominal frequency;
- Qa_1 is the relation between the initial reactive power (without compensation) and the initial total active power (P_0);
- KQV1 is the voltage exponent of the reactive power (without compensation);
- KQV2 is the voltage exponent for the reactive compensation;
- KQF1 is the frequency sensitivity coefficient of the load for the reactive power (without compensation);
- KQF2 is the frequency sensitivity coefficient of the load for the reactive compensation;
- Q_0 is the initial reactive power, obtained in the power flow solution.

From these equations, one can note that the active power is represented by two distinct components: one that depends on the voltage frequency and one that does not depend on this parameter. One can also observe that the reactive power is described by a first component that represents the load reactive consumption and by a second component that is related to the reactive consumption of feeders and transformers necessary to provide the initial reactive power injection (Q_0) for the bus.

2.1.2. Dynamic Models
In these models, the power (active or reactive) at an instant is represented as a function of the voltage magnitude and frequency of the load bus defined for the current instant and also for past instants. These models correspond to the base of many identification techniques used to model and forecast the load behavior, as described in the following.

2.1.3. Physical or White Box Models
The white box modeling method is also known as phenomenological, physical or conceptual modeling. These models are based on the deep knowledge about the physical behavior of the load and also on mathematical relations that describe the phenomena involved with the functioning of devices. It is an analytical modeling method, based on Physical laws as the Newton or the Ohm laws. If the physical laws that describe the load behavior are well known, then the model parameters are also known or previously specified. References [4, 5, 6, 7] deal with the phenomenological modeling of individual electrical equipment as motors, lamps, heating devices and air conditioners.

However, it is not always possible to know so well the behavior of the load to be modeled and, in these cases, one can use the black box modeling, described in the sequence.

2.1.4. Black Box Models and Identification Techniques
The black box modeling method is also known as identification method or empirical modeling and has as a main characteristic the fact that it is not necessary a deep knowledge about the load behavior to obtain an adequate mathematical model. This means that the model does not have physical meaning. The usage of these methods has been growing since 1990 due to (a) the complexity of the systems under analysis and the consequent difficulty to

develop physical models for them and (b) the availability of computers with processing power that allow the usage of a great amount of data and measurements directly collected from the system [8].

Some examples of this kind of model are the stochastic models AR (autoregressive), ARX (autoregressive with exogenous inputs), ARMAX (autoregressive moving average with exogenous inputs), MA (moving average), among others.

The models previously described in 2.1.1 and 2.1.2 correspond to equations composed by variables and parameters. The variables are related to power system quantities that can be measured and monitored as, for example, active and reactive power, voltage magnitude, frequency, etc. The parameters, which basically correspond to the coefficients and exponents of the equations, cannot be directly measured. However, they can be estimated through the usage of the identification techniques. Besides, these same identification methods can be used to determine the parameters of the mentioned stochastic models. Examples of these techniques are the Least Squares Method and the Kalman Filter [9].

2.2. Heuristic Models

The heuristic modeling is based on rules that associate some conditions (antecedents) to specific conclusions (consequents). The determination of the model consists on determining the adequate rules to represent the load behavior and these rules are commonly obtained from the experience of experts, allowing the development of a non-mathematical function to model the load. This modeling approach is usually employed in the implementation of expert systems.

2.3. Artificial Neural Network Based Models

Artificial Neural Networks (ANN) have the ability to adequately deal with nonlinear problems and complex systems. Hence, nowadays, they are used to model systems in many knowledge areas, including the load modeling area, as reported in [10, 11].

An ANN is composed by a set of processing units, called artificial neurons, which intend to simulate the functioning of biological neurons. These units are interconnected by weights (called connection weights, that are in fact mathematical parameters) that represent the synapses that take place inside the human brain. The connection weights can be modified along the ANN operation, generating an adaptation process that provides a great adaptability degree. In a generic way, the ANNs are trained using a succession of input/output patterns [10] and this training process results on the final values of the connection weights that determine the load model.

There are many architectures of ANNs and they are usually categorized depending on their learning process type. There are, basically, two types of learning processes:

- the supervised learning, that generates the values of the connection weights, through the minimization of a function that represents the difference between the desired output and the output obtained by the ANN. This learning process is used

in many types of ANNs, such as Hopfield networks, Multilayer Perceptron networks, etc.

- the self-organized or unsupervised learning, where the ANN organizes itself based on a function of its inputs. This learning process is used in other types of ANNs as, for example, Herault and Jutten networks and Kohonen networks.

The determination of the ANN load model consists on determining the set of parameters that leads to accurate results as the adequate number of neurons, the necessary number of hidden layers, the connection weights values, etc. Usually, the topology of the network is made in an empirical way and the weights are determined through the training process. It is important to emphasize that due to its characteristics, the ANNs can be considered as black box models.

2.4. Load Forecasting Methods

Load forecasting is of great value in power system operation and analysis. The knowledge of the load behavior at future instants can be used to prevent stability problems and also to plan the expansion of the electrical network (short-term and long-term load forecasting). All the models described 2.1, 2.2 and 2.3 can be used to predict the load behavior in the near or distant future. The way to do this is to adapt them and/or the identification techniques.

For example, the ZIP load model can be determined using the Least Squares identification method for a specific instant (static parameter estimation). However, the same identification method can be employed to dynamically estimate the load parameters. This dynamic estimation allows the knowledge of the load behavior for future instants (load forecasting). For the heuristic load models, the load forecasting can be implemented by choosing ruled that represent the load behavior in the future and for ANNs models, if the network is adequately trained, it becomes able to forecast the load. References [10, 12] illustrate some examples of load forecasting implementations.

3. NEW TRENDS ON LOAD MODELS AND FORECASTING TECHNIQUES

The models and methods described in section 2 can all be applied to represent the load behavior at normal operation conditions. This means that they are adequate to the way electric power systems are conventionally analyzed during simulations and studies: as a snapshot of the network operation, usually maintaining constant values for load demand and power generation. However, if the power system is subject to some kind of disturbance the load models must be able to display an adequate response to such perturbations. Commonly, for each type of perturbation, a specific approach of analysis is used. Consequently, specific models are adopted for each different kind of analysis. So, if one intends to assess the power system with respect to distinct phenomena nowadays, different models must also be used. In these cases, it is assumed that the load models associated to each bus, despite being able to

present an adequate response to the perturbation under analysis, are somewhat homogeneous from the dynamic point of view.

The problem is, power systems are becoming more and more complex each day; new generation equipment are being integrated to the network (as FACTS devices in the transmission subsystem), to work with well established technology equipment already in operation. Electronics is also playing an important role in the end-user load behavior due to its strong penetration in electrical appliances today. Hence, the aggregated load is not constituted by a homogeneous set of devices anymore. Moreover, the behavior of these devices is not unique, but depends on the fabricant and technology adopted in their construction. So, the models used in power system analysis today are not able to represent this multitude of different devices and appliances connected to the power network, specially under abnormal operation conditions (as overvoltage situations or blackouts).

In this context, the new trends on load modeling consist of the usage of models and techniques able to represent the impact of a large number of different devices (and their corresponding response) in the load behavior, during normal and abnormal operation conditions. Three major tendencies (discussed in the following sections) can be emphasized as suitable for this task: the usage of gray box models, artificial intelligence based models using a top-down approach and bottom-up approach with very detailed individual models. This last approach corresponds to a non-intrusive load modeling method that can be applied, especially in distribution networks, avoiding inconvenient related to the monitoring of residential consumption (in terms of total active power and the type of appliance).

3.1. Gray Box Identification

As mentioned before, the white box models have the advantage of bringing physical meaning to the load model and have the disadvantage related to the necessity of previous knowledge about the behavior of the load to be modeled. On the other hand, the black box models have no physical meaning, what represents a disadvantage, but present a significant advantage when compared to white box models: they can be determined without previous knowledge about the load behavior, using only measurement data and identification techniques.

The gray box identification is neither a white box nor a black box identification. In fact, it combines the advantages of these two identification methods and, consequently, is based on system measurements and also on auxiliary information, usually associated to some knowledge (but not deep knowledge) about the physical laws that describe the load behavior [8].

For example the polynomial model previously described can be considered as a black box model, because it has no physical meaning. However, some previous knowledge can be extracted from physical laws (based on the Ohm's law) that can describe more accurately the load behavior. This information is associated to the relation that can be found between the power and the voltage magnitude at the load bus: the power can depend on the square of the voltage magnitude, can depend directly on the voltage magnitude or can be constant. So if the exponents of the polynomial model are adequately chosen, generating a ZIP model, the load model is not a black box model anymore. However, since the exact physical law that represents the load is not known, this model can not be considered as a white box model

neither: it corresponds to a gray box model that combines the advantages of both conventional models.

This modeling technique corresponds to a new identification area and, since it provides more global models even with a limited number of measurements, it is a tendency to be used on load modeling and forecasting.

3.2. Artificial Intelligence Techniques with Top-Down Modeling

As previously mentioned, the top-down approach consists of models that directly represent the aggregated behavior of all systems' components. In the electric load modeling problem, it permits the determination of a model associated to the aggregated load on a feeder, area or substation.

The usage of artificial intelligence techniques combined with this approach can lead to interesting results because the model parameters can be modified based on heuristic rules stored in a knowledge database. Examples of these techniques are Artificial Neural Networks, Expert Systems, Fuzzy Logic associations [18], etc.

3.3. Bottom-up Modeling with Detailed Individual Models

In a general way, the bottom-up approach is based on the determination of individual models for each component of the actual system. In the load modeling problem, this approach permits the determination of individual models that represent the behavior of different equipment: lamps, showers, heating devices, air conditioners, among others. With the advances on the programming and computer fields, this approach can be easily combined with the usage of object-oriented programming (or agents) to represent very distinct individual models as different instances in a same computational context.

As an example, the next section presents an application of Bottom-up models to determined the behavior of thermostatically controlled loads in an extremely abnormal condition: the energization after a power system outage.

3.4. Application of Bottom-up Modeling in Determining
Load Restoration Behavior

The electric power systems around the world are all vulnerable to great outages. Recent examples of extensive interruptions in the energy supply are the blackouts that occurred in the United Kingdom, USA, Denmark, Sweden and Italy in 2003. In addition to new investments in research and device deployment to reduce the risk of new outages, the elaboration of reliable restoration plans is crucial to minimize the enormous social and economical possible impacts of blackouts [13].

In this context, the load behavior at the reenergization moment and the subsequent instants (cold load pick-up) is a fundamental issue that must be considered when developing restoration plans. If an attempt is made to pick-up a larger load block than the one that can be

accommodated by the response capability of the on-line generators, then blocks of load already restored can be lost, due to under-frequency load shedding. If the frequency swing is severe, some generators may trip precipitating another shutdown.

Problems of this nature can occur specially when reenergizing thermostatically controlled devices (TCDs), since this type of load can present an unexpected power demand increase at the reenergization instant [14]. Consequently, forecasting the behavior of these devices can prevent overload problems and improve the reliability of power system restoration plans. Reference [13] presents some methods that deal with this problem assuming aggregated loads constituted by groups of the same type of devices. However, their inherent difficulty in estimating the aggregated behavior of a feeder containing appliances of different types can be considered as a limitation of these methods.

Hence, this section presents a methodology to forecast the post-outage behavior of the aggregated load in a feeder, considering thermostatically controlled loads of distinct types and based on object oriented programming techniques. This method makes possible the identification of external variables impacts on the post-outage load behavior (as the environment temperature and the blackout duration influences).

3.4.1. Thermostatically Controlled Loads

Thermostatically controlled loads can be defined as appliances and devices which states (turned on or off) are controlled by a thermostat or by temperature sensors. Examples are refrigerators, air-conditioners, freezers and heating devices. In the following, the behavior of a single TCD under normal operation conditions and for outage situations is described.

a.   Behavior Under Normal Operation Conditions

The behavior of a thermostatically controlled appliance under normal operation conditions is illustrated in figure 1. In this figure, the first curve represents the thermostatic cycle of an individual cooling device and the second curve characterizes its active power demand under normal operation conditions. When the device is turned off, it begins to warm-up and its internal temperature begins to rise. This fact can be observed in the regions marked with a "*W*". Similarly, when the device is turned on, it begins to cool-down and its internal temperature decreases (region marked with a "*C*"). It should be noted that this behavior is inverted for heating devices.

Hence, in steady-state conditions, the appliance internal temperature remains between two specified values: a superior limit (T_{max}) and an inferior limit (T_{min}). The temperature control is made through a thermostat that changes the appliance state (on or off) at these temperature limits. The device active power demand can be considered *constant* (P_{on}) when it is turned on. It must be emphasized that this hypotheses corresponds to a simplification of the real behavior of this type of load. According to reference [14], after the appliance energization, there is an overshoot in the power consumption during the first thermostatic cycle, caused by a set of physical phenomena related to the device functioning. However, the consideration of such phenomenon increases the complexity of the methodology, requiring additional efforts during the resolution process. So, this overshoot was not considered and the power demand of each device is set as constant during its "on" state, simplifying the aggregated power demand evaluation.
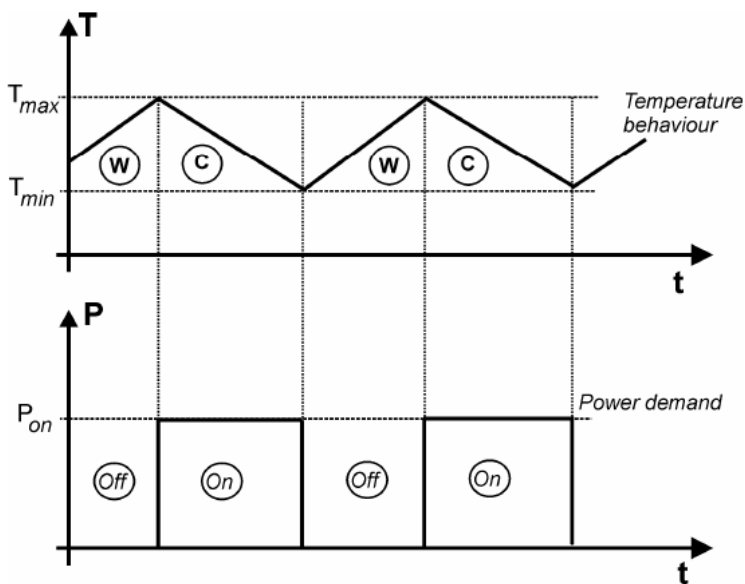
Figure 1. Thermostatic cycle and power demand behavior under normal operation conditions (cooling device).

b. Behavior Under Outage Conditions

The behavior of a cooling device under outage conditions is illustrated in figure 2. In this figure, the first curve represents the thermostatic cycle of a cooling device for an outage situation and the second curve characterizes its active power demand.
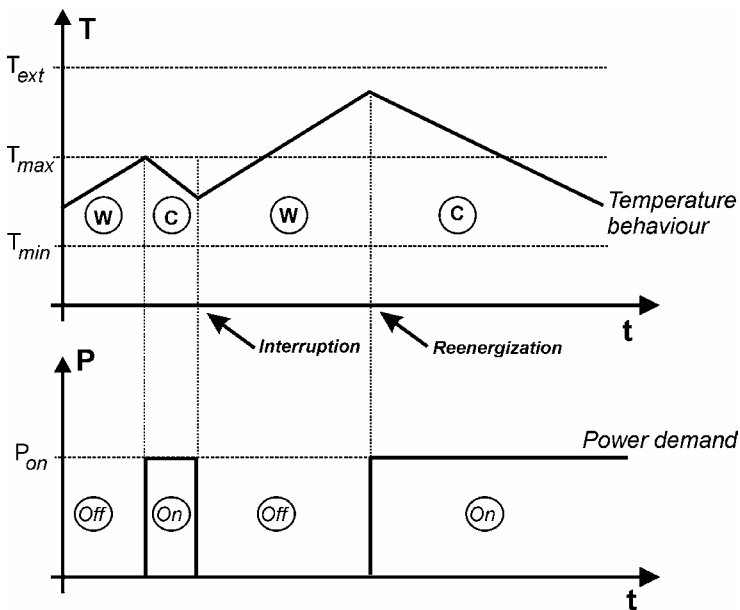
Figure 2. Thermostatic cycle and power demand behavior for outage conditions (cooling device).

Depending on the outage duration, the appliance can be unable to turn-on when the upper temperature limit is reached, due to the electric power interruption. This means that the TCD is unable to execute its on cycle when the temperature upper limit is violated, modifying the thermostatic cycle as shown in figure 2.

In fact, the interruption causes a thermal energy loss inside the TCD and this loss is the responsible for the alterations in the thermostatic cycle. So, when reconnected, the cooling TCD will execute an "on" cycle with longer duration when compared to its normal "on" cycle, in order to recover its lost thermal energy. This altered cycle will last until the temperature reaches the lower temperature limit T_{min}.

Considering a large number of TCDs (as, for example, the quantity that can be found connected to a distribution feeder), these modifications on the thermostatic cycle of each TCD can imply on a very peculiar phenomenon that corresponds to the *loss of diversity* among the states of the TCDs. This means that, due to the alterations on the thermostatic cycle, there will be more TCDs (when compared to normal operational conditions) ready to execute their "on" cycle at the reenergization instant. This fact results in an increase on the power demand when the load is restored. The load level in this case can be much greater than the steady-state power demand. This characteristic is commonly associated to the post-outage behavior of aggregated thermostatically controlled loads and is illustrated in figure 3 [14], that shows the actual behavior of a residential area, constituted by many TCDs, after a blackout of 3 hours.
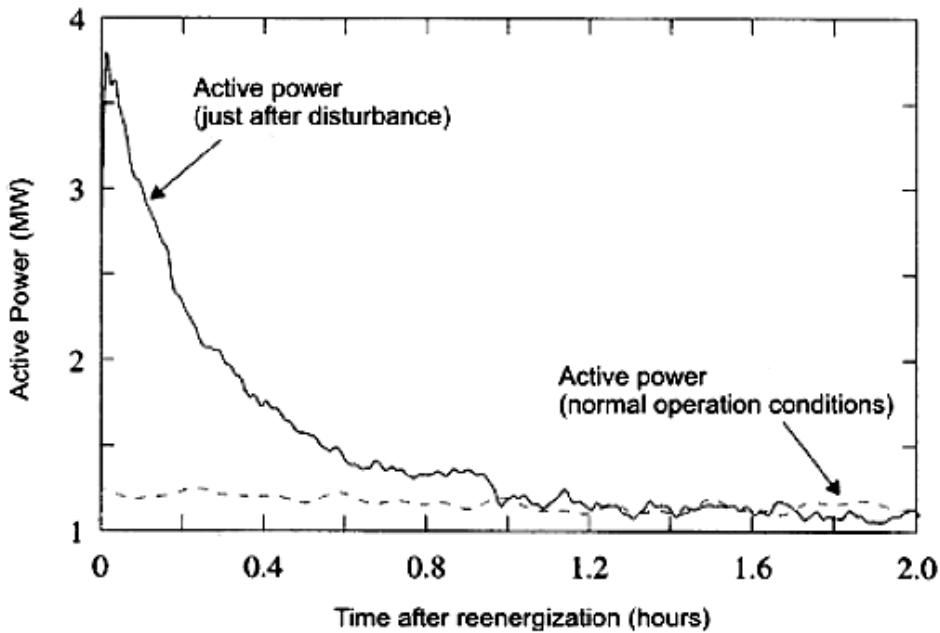
Figure 3. Post-outage behavior of a residential area constituted by many TCDs.

If this power peak is not considered during the load pick-up, it can cause serious operational problems that can even lead the system to a new blackout. In this context, this section is focused on the determination of the aggregated load in a feeder with significant

penetration of TCDs, based on models that define the behavior of a single TCD. The proposed methodology can be summarized as follows.

- Determination of an individual model for each type of appliance. In this work, the TCDs considered for modeling were refrigerators, freezers and air-conditioners. The models were implemented using object oriented programming techniques as described in Section 3.4.2.
- Aggregation of the individual post-outage behavior of each appliance to obtain the aggregated behavior of a feeder constituted by many thermostatically controlled loads. This aggregation procedure is described in Section 3.4.3.

3.4.2. Object Oriented Models

Object oriented programming (OOP) techniques are very suitable to implement the individual models of the devices, since they value the physical modeling through the definition of a small number of classes [17]. The behavior of a TCD is illustrated in figure 1 (normal operation conditions) and 2 (outage conditions) as mentioned before, and can be described by Eq. (5) and (6) [15].

$$T_i = T_{i-1} + \alpha . \Delta t \tag{5}$$

$$P_i^e = m_i . P_{on} \tag{6}$$

In these equations:

- T_i is the device internal temperature at instant i.
- α is a thermal coefficient (cooling or heating), given in degrees per time-units.
- P_i^e is the active power demand of device e at instant i.
- m is a logical variable that represents the state of the device e. If $m = 0$, the device is turned off; if $m = 1$, the device is turned on.
- P_{on} is the power demand of the device e when turned on.
- Δt is the time interval between instant i and instant $i\text{-}1$.

However, the coefficient α is dependent on the temperature difference between the device internal temperature T_i and the external or environmental temperature T_{ext}. To avoid the presence of this dependence in problem formulation, Eq. (5) can be rewritten as:

$$T_i = T_{i-1} + \alpha_N . \Delta T . \Delta t = T_{i-1} + \alpha_N . (T_{ext} - T_{i-1}) . \Delta t \tag{7}$$

where α_N is the normalized thermal coefficient, given in inverse time units (s^{-1}, m^{-1} or h^{-1}). Eq. (7) can be used to describe the behavior of both heating and cooling devices, by choosing adequate values for the normalized thermal coefficient. So, although this section is focused on cooling loads (refrigerators, freezers and air-conditioners), the proposed methodology can also be applied to model heating devices.

Using object oriented techniques, the individual models of freezers, refrigerators and air-conditioners were derived from a superclass named *Thermostatically Controlled Device*. The main characteristics of this class are represented in the diagram illustrated in figure 4.

Class Thermostatically Controlled Device

MOST IMPORTANT FIELDS:

α_w: normalized heating coefficient {double}
α_c: normalized cooling coefficient {double}
T: device current temperature {double}
T_{max}: maximum temperature {double}
T_{min}: minimum temperature {double}
T_{ext}: external or environmental temperature {double}
P: active power demand when turned on {double}
t: device current time {double}

MOST IMPORTANT METHODS:

public void setTime(double time)
public boolean getState()
public void setReferenceTemperature(double T_{ext})
public double getDemand()

Figure 4. Class diagram of the superclass *Thermostatically Controlled Device*.

The basic attributes (fields) of this class are:

- α_w: a *double* value that represents the *normalized thermal coefficient* of the device when it is warming-up (internal temperature rising), given in inverse time units (h^{-1}, m^{-1}, s^{-1}).
- α_c: a *double* value that represents the *normalized thermal coefficient* of the device when it is cooling-down (internal temperature lowering), given in inverse time units (h^{-1}, m^{-1}, s^{-1}).
- T: a *double* value that represents the current temperature of the device, given in Celsius or Fahrenheit degrees (^{o}C or ^{o}F).
- T_{max}: a *double* value that represents the maximum temperature of the device, given in Celsius or Fahrenheit degrees (^{o}C or ^{o}F).
- T_{min}: a *double* value that represents the minimum temperature of the device, given in Celsius or Fahrenheit degrees (^{o}C or ^{o}F).
- T_{ext}: a *double* value that represents the external temperature of the device, given in Celsius or Fahrenheit degrees (^{o}C or ^{o}F).
- P: a *double* value that represents the active power demand of the device at its "on" state, given in Watts (W).

- t: a *double* value that represents the device current time (given in hours, minutes or seconds).

The basic procedures (methods) defined in this class are:

- Method "setTime": by using this method, it is possible to update the TCD current time (t).
- Method "getState": this method returns the current state of the device (turned on or turned off).
- Method "setReferenceTemperature": method responsible for the update of the TCD current temperature (T).
- Method "getDemand": this method returns the power demand (P) of the device at its current time (t).

Based on this superclass and on the mechanism of inheritance related to OOP techniques, one can derive other classes that correspond to different TCDs. The definition of these other classes through inheritance is illustrated in figure 5.
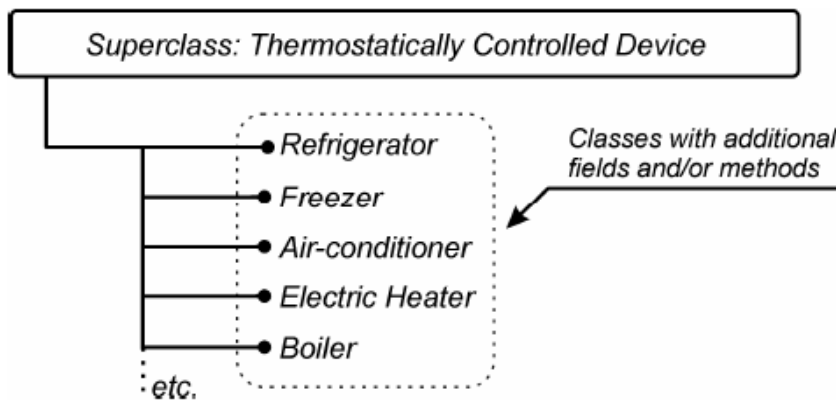
Figure 5. Inheritance mechanism.

These classes have the basic functionalities defined in the superclass and also can use other methods that characterize each class individually. For instance, one can define that the "air-conditioner" class behaves as a TCD only during daylight hours; at night, its power demand is null. So, it is possible to define different classes to represent freezers, refrigerators and air-conditioners.

The classes correspond to containers that are used to define computational instances, the *objects*. These objects are the computational entities that really correspond to the electrical devices. For instance, in order to represent 100 freezers connected to the distribution system, 100 objects of the class *Freezer* must be instantiated. Moreover, OOP techniques make the definition of devices with different characteristics (all connected to the same feeder) an easy task. This allows the adoption of distinct values for the individual power demand and dynamic parameters for each device, even when they are instances (objects) of the same class.

3.4.3. Aggregation

Disregarding power distribution losses, the aggregated power demand of thermostatically controlled loads in a feeder, for an instant i, can be obtained using Eq. (8):

$$P_i^F = \sum_{e=1}^{Ndev} P_i^e \tag{8}$$

where:

- P_i^e is the power demand of device e at instant i ;
- P_i^F is the power demand of feeder F at instant i;
- *Ndev* is the total number of TCDs connected to feeder F.

Based on this concept, another class was defined to determine the aggregated behavior of the thermostatically controlled loads. This class, *TCD Feeder*, is illustrated in figure 6.

Class TCD Feeder
MOST IMPORTANT FIELDS: P: Aggregated active power {double}   V: Feeder voltage magnitude {double}   FT (Feeder Time): Feeder reference time {double}
MOST IMPORTANT METHODS: public void setTime(double time)   public void setInterruptionTime(double IT)   public void setInterruptionDuration(double ID)   public double aggregateDemand()
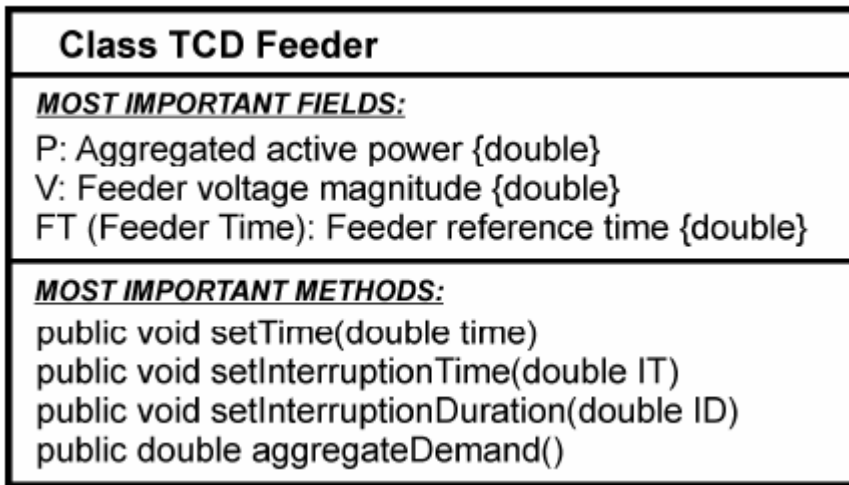
Figure 6. Class diagram of the class *TCD Feeder*.

The basic attributes (fields) of this class are:

- P: a *double* value that represents the total power demand related to the feeder at its current time, given in Watts (W).
- V: a *double* value that represents the voltage level related to the feeder, given in Volts (V).
- FT (Feeder Time): a *double* value that represents the feeder current time, given in hours, minutes or seconds.
- The basic procedures (methods) defined in this class are:
- Method "setTime": by using this method, it is possible to update the current feeder time (FT).

- Method "setInterruptionTime": this method permits to define the moment when the outage begins in the simulation timeline (given in hours, minutes or seconds).
- Method "setInterruptionDuration": this method is used to define the duration of the outage (given in hours, minutes or seconds).
- Method "aggregateDemand": once called, this method simulates the aggregated load behavior during the outage an in the subsequent instants.

The process activated by the method "aggregateDemand" is the basis of the aggregation methodology and is detailed in the following. The procedure consists of an iterative signal traffic among the feeder object and its corresponding TCDs objects [19]. Each iteration of this traffic can be divided in two sequential stages. In the first, illustrated in figure 7-A, the current feeder time (FT) and its voltage level (V) are broadcasted to a set of TCDs. When receiving this signal, the TCDs update their internal states. This causes the update of all internal TCD variables, basically ruled by the dynamics defined in Eq. (5), (6) and (7).

In the subsequent stage, illustrated in figure 7-B, each selected TCD sends the updated value of its power demand P to the feeder F. The feeder waits for all demand values and then aggregate them using Eq. (8).
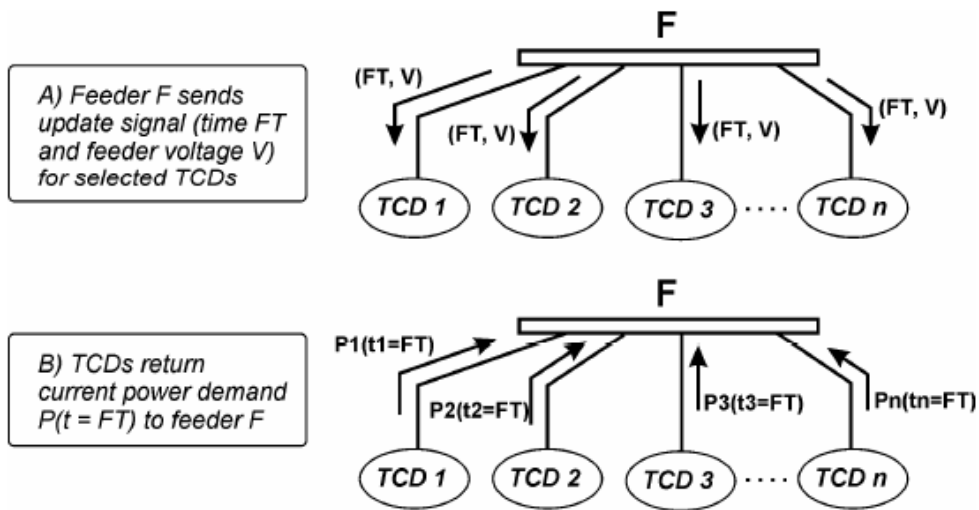
Figure 7. Stages of load behavior determination: broadcast (A) and aggregation (B).

A great advantage of this approach is the heterogeneous nature of the TCDs set. A huge number of different TCDs with very distinct dynamics can be aggregated using the signal coordination of a single feeder. Moreover, the TCDs with faster dynamics can be updated separately at a rate higher than the updates required by the remnant set of devices. This means that the updates signaling can be asynchronous. So, the computational burden associated to feeders with huge dimensions can be attenuated by breaking the TCDs set into smaller groups and then processing them one by one, taking advantage of the asynchronous updating characteristic.

3.4.4. Results

In order to test the proposed methodology, simulations were made considering two hypothetical feeders containing different TCDs, named feeder F1 and feeder F2. Table 1 shows the TCDs parameters. Table 2 summarizes the composition of both feeders. The objective of these simulations was to observe the influence of two important variables, the *outage duration* and the *environment temperature*, on the aggregated behavior of the thermostatically controlled loads. These influences are described in the next subsections.

Table 1. TCDs Simulation Parameters

TCD	Parameters
Refrigerator	$\alpha_w = 8.10^{-3}$ s^{-1} ; $\alpha_c = 24.10^{-3}$ s^{-1} ; $T_{max} = 12$ ^{o}C; $T_{min} = 6$ ^{o}C; $P = 50$ W.
Freezer	$\alpha_w = 5.10^{-3}$ s^{-1} ; $\alpha_c = 15.10^{-3}$ s^{-1} ; $T_{max} = -5$ ^{o}C; $T_{min} = -10$ ^{o}C; $P = 80$ W.
Air-conditioner	$\alpha_w = 1,7.10^{-3}$ s^{-1} ; $\alpha_c = 5.10^{-3}$ s^{-1} ; $T_{max} = 22$ ^{o}C; $T_{min} = 20$ ^{o}C; $P = 500$ W.

Table 2. Composition of the Feeders Under Analysis

Feeder	Description	Total of TCDs
F1	500 refrigerators, 125 freezers and 30 air-conditioners	655
F2	1504 refrigerators and 310 freezers	1814

The internal temperature (T) of all TCDs was initially set as an uniform random distribution varying between T_{min} e T_{max}, for each class of objects. The default external temperature (T_{ext}) is 25 ^{o}C. The initial state (turned on or turned off) was also determined in a random way for each TCD. The simulation begins at 06:00 h and finishes at 14:00 h.

a. Impacts of Outage Duration

The first test condition corresponds to the simulation of outages with *short duration* ($\leq$ 30 minutes), in feeder F1, starting at 10:00 a.m. Figure 8 illustrates the aggregated behavior of the TCDs after outages of 5, 10, 20 and 30 minutes. This figure shows that the active power level at the reenergization instant is proportional to the outage duration. The greater this duration, the greater the power peak. In other words, as the outage duration increases, there is also an increase on the diversity loss among the states of these devices, i.e., at the restoration moment, more devices will be ready to start their "on" cycles, consuming nominal power and, consequently, causing an increase on the aggregated power level.

The second test condition corresponds to the simulation of outages with *long duration* (> 30 minutes), starting at 10:00 a.m., also in feeder F1. Figure 9 illustrates the aggregated behavior of the TCDs after outages of 60, 90 and 120 minutes.

Comparing this test condition to the first one, a different characteristic can be noted in figure 9: a constant power plateau starting at the restoration moment. This plateau corresponds to the maximum power demand level and is caused by the total loss of diversity among the TCDs states (all devices turned on at the same time). It is important to observe that the duration of this plateau is also proportional to the outage duration.

As expected, the feeder power demand at the reenergization instant is much greater than its demand at normal operation conditions in both tested situations. This characteristic must be taken into account when planning the restoration process.
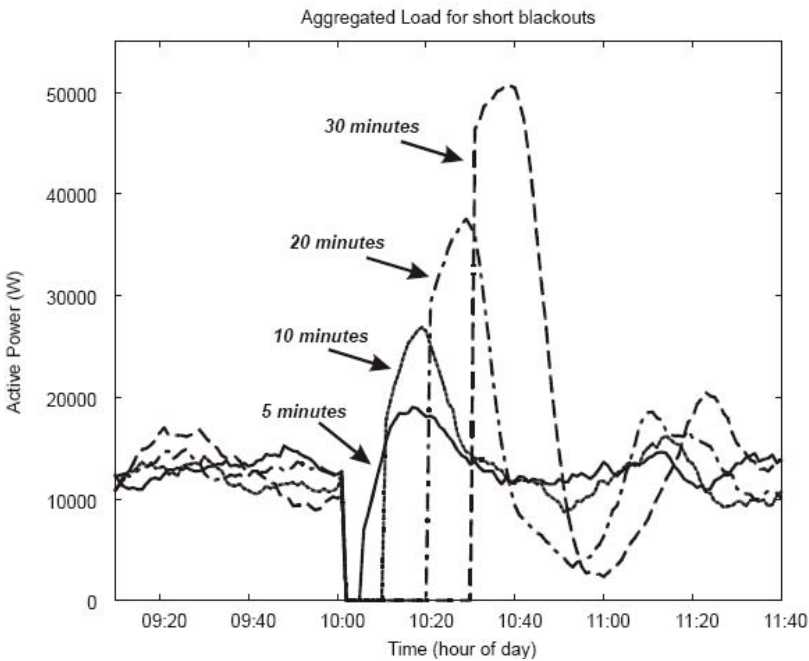
Figure 8. Aggregated behavior of the TCDs after short outages (≤ 30 minutes).
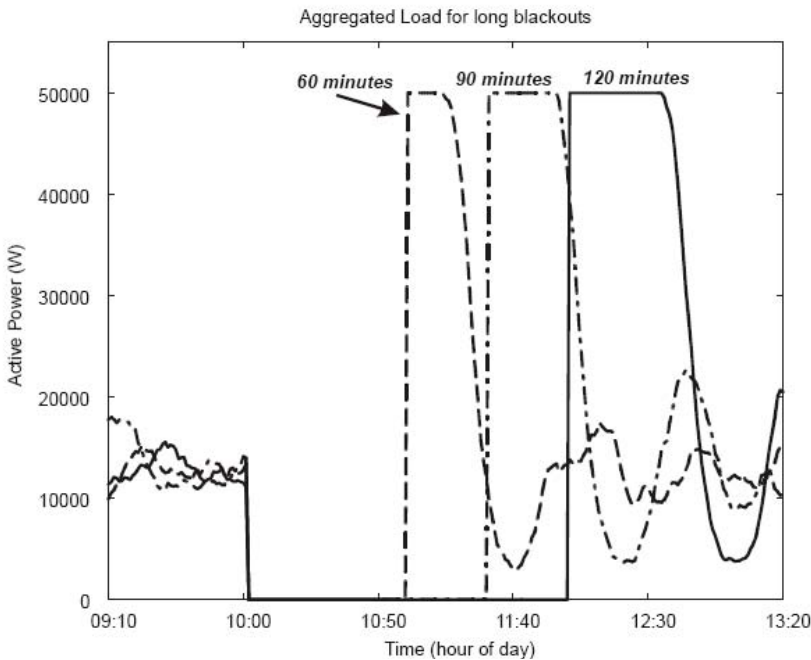
Figure 9. Aggregated behavior of the TCDs after long outages (> 30 minutes).

b.   Impacts of Environment Temperature

This test condition, now in feeder F2, corresponds to the simulation of an outage of 60 minutes, starting at 10:00 a.m.. Figure 10 illustrates the aggregated behavior of the TCDs considering different environment temperatures: 15, 25 and 32 Celsius degrees.

From figure 10, it is possible to observe that the environment temperature influences the pre-outage aggregated load behavior. An increase on the temperature causes an increase on the steady-state power level before the interruption. This behavior was expected since the loads simulated are all cooling devices. Concerning the post-outage behavior, one can also observe that an increase on the temperature also implies on an increase on the plateau duration. This was also expected since the environment temperature modifies the thermostatic cycle of the devices and consequently is directly related to the diversity among the states of these loads and to the plateau duration [13].
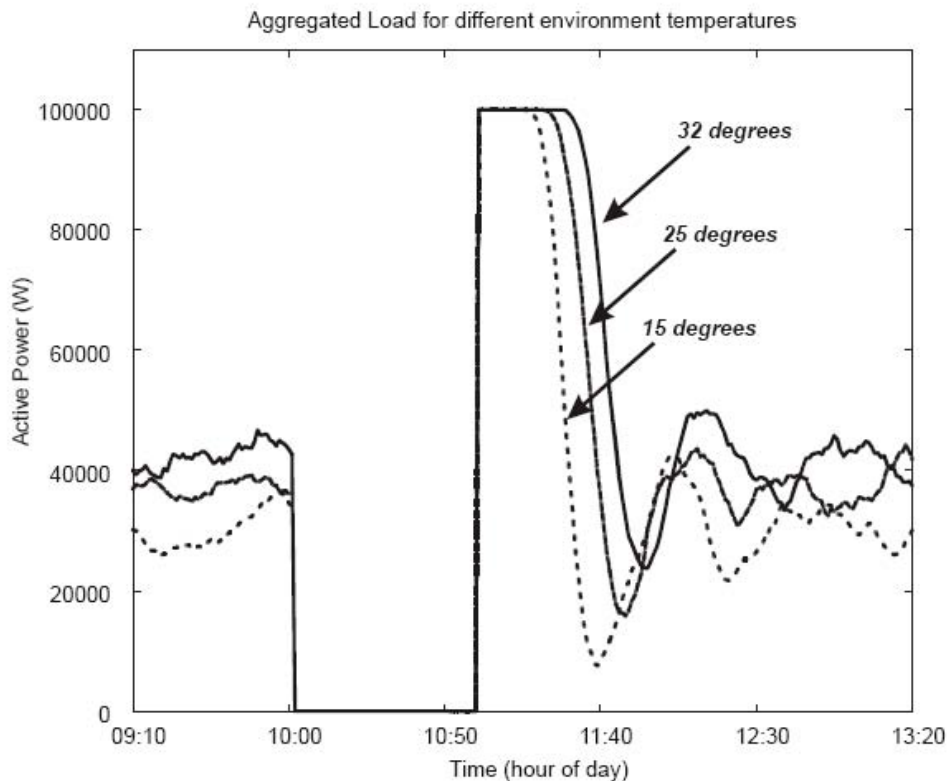
Figure 10. Aggregated behavior of the TCDs considering different environment temperatures.

3.4.5. Discussion

In this section, a methodology to model the behavior of thermostatically controlled loads was proposed. This method allows the determination of the aggregated post-outage behavior of TCDs, based on object oriented programming techniques.

Simulations were carried out considering two hypothetical feeders with a large number of TCDs. They agree with field measurement data presented in the literature [13, 17]. The obtained results permit to observe the influence of the blackout duration and of the environment temperature on the aggregated behavior of these devices. The differences among the resulting load curves enforce that these impacts must be considered in the development of restoration plans, in order to avoid inadequate load pick-up.

4. CONCLUSION

As power systems around the world become more complex, the operation and maintenance of the electrical network relies more and more on simulation results, that are very dependent on the adopted load models. The models commonly used in power systems analysis and simulations are not able to represent individual phenomena associated to energy end-users. This can lead to inaccuracies in the results assessment and, thus, can significantly affect the reliability of power system operations. However, artificial intelligence techniques, gray-box modeling approach and bottom-up aggregation procedures present characteristics that can be advantageous to eliminate these deficiencies in the electric load models. As computer hardware and software evolve, the computational burden associated to these approaches becomes less significant. So, these approaches can represent, in the near future, a very suitable way to predict and analyze the aggregated load in electric power feeders more accurately.

REFERENCES

[1]   Coker, M.L.; Kgasoane, H.; Load Modeling. *IEEE AFRICON,* 1999. Volume 2, 28 Sept.-1 Oct. 1999 Page(s):663 – 668.

[2]   Price, W.W.; Bibliography on Load Models for power Flow and Dynamic Performance Simulation - IEEE Task force on Load Representation for Dynamic Performance, *IEEE Transactions on Power Systems,* pp 523 – 538, Feb. 1995.

[3]   Ju, P., Handschin, E., Karlsson, D.; Nonlinear Dynamic Load Modeling: Model and Parameter Estimation, *IEEE Transactions on Power Systems,* pp 1689 – 1697, Nov, 1996.

[4]   Nehrir, M.H., Dolan, P.S., Gerez, V., Jameson, W.J.; Development and Validation of a Physically-Based Computer Model For Predicting Winter Electric Heating Loads, *IEEE Transactions on Power Systems,* Vol. 10, No. 1, Feb. 1995.

[5]   Chen, C-S., Wu, T-H., Lee, C-C., Tzeng. Y-M.; The Application Of Load Models Of Electric Appliances To Distribution System Analysis, *IEEE Transactions on Power Systems,* Vol. 10, No. 3, Aug. 1995.

[6]   Dolan, P.S., Nehrir, M.H., Gerez, V.; Development Of A Monte Carlo Based Aggregate Model For Residential Electric Water Loads, Electric Power Systems Research, 36, pp. 29 – 35, 1996.

[7]   Tomiyama, K., Daniel, J.P., Ihara, S.; Modeling Air Conditioner Load For Power System Studies, *IEEE Transactions on Power Systems,* Vol. 13, No. 2, May 1998.

[8]   Aguirre, L. A., An Introduction to Systems Identification (in portuguese), *Editora UFMG,* Brazil, 2000.

[9]   Mota, L.T.M.; Mota, A.A. ; Monticelli, A.J.; Dynamic Load Parameters Estimation in Electric Power System State Estimation. *IEEE PES T and D2002 Latin America,* 2002.

[10]   Baumann, T.; Germond, A.J.; Application of the Kohonen network to short-term load forecasting. *Proceedings of the Second International Forum on Applications of Neural Networks to Power Systems - ANNPS '93.,* Page(s):407 – 412, 1993.

[11]   Lee, J.P. et al.; A Load Modeling Using ANN for Power System Analysis. *IEEE TENCON,* 1999.

[12]   Gross G., Galiana F., Short-term load forecasting, *Proceedings of the IEEE, Vol.* 75, No. 12, pp.1558-1573, Dec. 1987.

[13]   M. M. Adibi, Power System Restoration – Methodologies and Implementation Strategies. New York: IEEE Press, 2000.

[14]   E. Agneholm, "Cold load pick-up," Ph.D. *Thesis,* Chalmers University of Technology, Göteborg, Sweden, 1999.

[15]   R. E. Mortensen, and K. P. Haggerty, "Dynamics of Heating and Cooling Loads: Models, Simulations and Actual Utility Data," *IEEE Transactions on Power Systems,* vol. 5, pp.243-249, 1990.

[16]   J. Rumbaugh, (1991), *Object Oriented Modeling and Design,* Prentice-Hall, 1991.

[17]   E. Agneholm, J. Daalder, "Cold load pick-up of residential load" *IEE Proceedings on Generation, Transmission and Distribution,* vol. 147, n.1, pp.44-50, jan. 2000.

[18]   Mota, L.T.M.; Mota, A.A.; Morelato, A.; Load behaviour prediction under blackout conditions using a fuzzy expert system. *Generation, Transmission and Distribution, IET.* Volume 1, Issue 3, Page(s):379 – 387, May 2007.

[19]   Mota, L.T.M.; Mota, A.A.; Morelato, A.; Physically-based aggregation of object-oriented load models in power systems restoration. *IEEE/PES Transmission and Distribution Conference and Exposition Latin America,* 2004. Page(s):155 – 160. Nov. 2004

Chapter 6

APPLICATION OF FTIR TO ANALYSIS CHEMICAL STRUCTURE OF SOOT AEROSOL EMITTED FROM LUMP-COAL PYROLYSIS IN FIXED BED

Q. C. Wang and Y. H. Luo

Institute of Thermal Engineering, School of Mechanical and Power Engineering,
Shanghai Jiao Tong University, Dongchuan Road 800,
Shanghai 200240, People's Republic of China

ABSTRACT

The objectives of this work are to study the primary chemical structure of soot aerosol derived from lump-coal pyrolysis in different experimental conditions in fixed bed. A laboratory-scale movable fixed bed, water-cooled soot aerosol collection system, and electric reactor have been designed and used in the process. Three kinds of coals, sized at 3-5 mm, have been heated in the experiments. Fourier Transform Infrared Spectroscopy has been employed to test functional groups of soot aerosol samples. Infrared spectra from 400 to 4000 cm^{-1} and semiquantitative analysis have been employed. The results of experiments show that contents of hydrogen-bonded are increased, contents of unsaturated hydrocarbons are decreased, and contents of aromatic hydrocarbons are increased with temperature increase; contents of hydrogen-bonded are increased, contents of unsaturated hydrocarbons are increased, and contents of aromatic hydrocarbons are increased early and decreased late with residence time extension; and the contents of unsaturated hydrocarbons derived from soot aerosol samples are higher than those from original coal samples, and contents of hydrogen bonded and aromatic hydrocarbons are different depending on chemical structure of original coals.

Keywords: lump-coal; pyrolysis; soot aerosol; fixed bed; FTIR.

1. INTRODUCTION

Small submicron carbonaceous particles, known as soot, are commonly observed in the pyrolysis and combustion of hydrocarbons and coals. Soot does great harm to people's health for carcinogenic effect [1]. There are many organic pollutants with soot, such as polycyclic aromatic hydrocarbons (PAH), polychlorinated dibenzo-p-dioxins and polychlorinated dibenzo-p-furans (PCDD/Fs), hydrocarbon of C1~C20, hydroxybenzene, amine, Bap, and so on. On the other hand, soot suspended in flames is important to combustion systems because it will significantly enhance radiative heat transfer due to their large surface area [2,3].

Based on the simulation and experiments conducted on simple hydrocarbon flames, it is found that soot is usually formed when conditions are sufficiently fuel rich to allow condensation or polymerization reactions [4-7]. Besides, soot formation has also been observed in many pulverized coal utilization processes, including coal gasification and combustion. In a conventional pulverized coal combustion boiler, the PAH have been believed to be the precursor of coal-derived soot. These tar molecules are therefore much larger and more chemically diverse than the simple hydrocarbon fuels generally used for studies of soot formation [4]. Nenniger [8], Wornat [9], and Chen [10] et al studied the soot yields from pulverized coal in Pyrolysis environment. They indicated that the sum of tar plus soot remained constant; increases in soot yields coincided with decreases in tar yields. Ma [2] researched soot formation in oxygen-containing presence condition, and revealed that the soot yields obtained were much lower than those obtained in inert condition.

However few relevant investigations on lump coal-derived soot in fixed bed have been detected. In China, the total installed capacity of the industrial boilers in the industrial and service sectors is more than 1.2 millions t/h (tons per hour), and more than 400 million tons of lump coals have been consumed a year [11]. In these industrial boilers, stoker firing is the dominating combustion style. The combustion mode of lump coal is different from that of pulverized coal, because of poor mixing between volatiles of coal and air, and pyrolysis reaction exists in local areas in the hearth of stoker-fired boiler. The study on soot aerosol in pyrolysis process in local areas of stoker-fired boiler is significant in terms of environmental impact. FTIR used widely in analysis chemistry is employed to study the formation mechanism of soot aerosol by measuring change of functional groups.

2. EXPERIMENTAL SECTIONS

2.1. Experimental Apparatus

In order to study conveniently the soot formation in the stoker-fired boiler, a fixed-bed reactor has been designed, avoiding the expensive and difficult work in the large-scale equipment. The lump coal pyrolysis experiments were conducted in an electric reactor to simulate the coal pyrolysis on the grate in the hearth of stoker-fired boilers. It was configured to provide a simulation of the fuel-bed pyrolysis regime in a coal-fired, mass-burning stoker. A thick fuel bed composed of lump coal was heated in a batch experimental process. In this configuration, the coal in the fixed bed was heated in a transient mode to provide a Lagrangian simulation of the time, temperature, and environmental history experienced by

small section of coal, which is in the fuel bed in a large travelling-grate stoker. In this way, the pyrolysis processes that occur simultaneously in a full-scale unit can be simulated in the fixed bed [12-14]. Figure 1 is a conceptual illustration of this facility built at the Shanghai Jiaotong University after extensive discussions.

There were three sections in the whole experimental apparatus: the core parts were electric reactor and two-section temperature control equipment; other parts were soot collection system and coal supply system.

The electric reactor was a cylindrical alundum tube heated by silicon-carbon pole. The temperature in the tube could be controlled by regulating the electric current and voltage of the two-section temperature control equipment, and could be controlled by either one section or two sections depending on the experimental objectives. The range of temperature was from room temperature to 1500 .The diameter of the cylindrical tube was 45 mm, and the height of invariable-temperature tube section was 300 mm. A fixed bed was in the tube, supported by coal supply system.

The gases from the tube reactor flowed through a water-cooled and 30mm-diameter tube, in which, a glass fiber or polycarbonate filter with 0.05 µm pore diameters was fixed, so it is feasible that the gas went through the filter and the soot samples were collected on the filter [4]. The filters marked Whatman® were made in England. Vacuum pump and valve were set for sustaining stable pressure in the tube, and a columned port was provided to sample the gases.
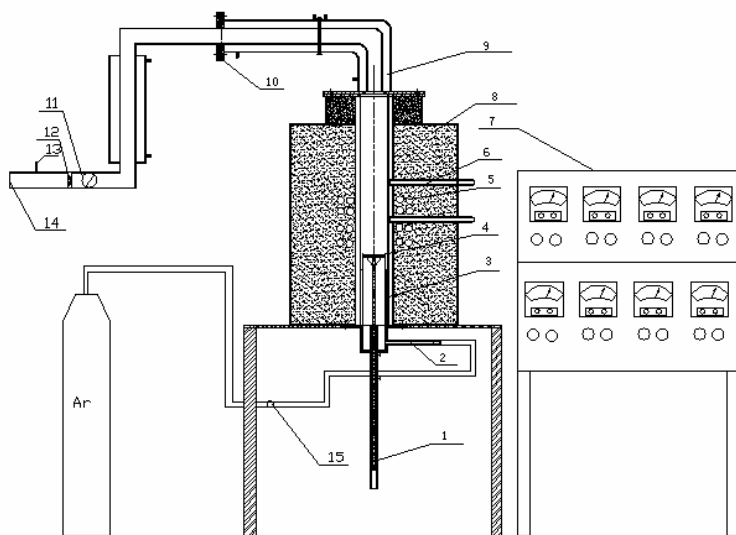
Figure 1. The conceptual illustration of lump coal pyrolysis experiments.

The fixed bed was a plate-shape refractory, whose diameter was 44 mm. Many 2 mm-diameter holes were arranged in the bottom of the fixed bed. A moveable refractory pole was used to support the fixed bed, and a larger-diameter stainless tube was connected to the foundation of the reactor by bolts. The different heights of the fixed bed in the reactor depend on the position of the moveable refractory pole, and then the residence time of gases in the reactor could be adjusted.

Table 1. Ultimate and proximate analysis of Shenmu, Datong, and Zibo coal

Coal type	Proximate analysis %					Ultimate analysis %				
	M_{ar}	Aar	Var	V_{daf}	FCar	Car	Har	Sar	Nar	Oar
Shenmu	12.13	5.65	29.36	35.70	52.86	65.22	4.44	0.17	1.37	11.02
Datong	8.42	16.28	24.98	33.17	50.32	61.67	3.99	0.72	1.28	7.64
Zibo	2.85	20.95	29.79	39.10	46.41	62.97	4.12	0.63	1.24	7.24

2.2. Experimental methods

Three effect factors including temperature, residence time of gases, and type of coal, were considered before this experiment.

In the experimental facility described in figure 1, temperature was adjusted to different point according to individual experimental objectives firstly. The lump coals were introduced onto the bed outside of the reactor secondly. A batch of coal particles was 7 g. The coal was classified with sieves by its diameter, that is, particles ranging in size from 3 to 5 mm were used to be pyrolyzed, particles ranging in size less than 3 mm or more than 5 mm were dismissed. The selected coals, whose ultimate and proximate analysis are shown in table 1, were the three coals widely used in industrial boiler in China. The bed with coals were sent to the hearth from the bottom of the reactor thirdly, supported and fixed by the coal supply system (figure 1). The volatiles were released and expanded away from the coal particles. Argon flows measured with a rotameter were introduced to the reactor from the bottom immediately; at the same time, the vacuum pump was turned on. The variation of Argon amount was necessary for the case of different temperature in different conditions in the reactor in order to achieve the residence time projected.

The soot aerosol samples were scraped from the filter and prepared for FTIR analysis using the traditional procedure for this type of materials. Thus, an extract (residue): KBr mixture at 1: 12 ratio was ground in an agate container. Pellets were then vacuum-pressed and dried at 100 overnight. More than one pellet was prepared for each sample. The pellets were pressed to form slice; and subsequently the sample slices were tested by FTIR.

Table 2. Characteristic infrared absorption frequencies

Bond	Compound Type	Frequency range, cm^{-1}
C-H	Alkanes	2960-2850(s) stretch
		1470-1350(v) scissoring and bending
C-H	Alkenes	3080-3020(m) stretch
		1000-675(s) bend
C-H	Aromatic Rings	3100-3000(m) stretch
	Phenyl Ring substitution Bands	870-675(s) bend
	Phenyl Ring Substitution Overtones	2000-1600(w) - fingerprint region
C-H	Alkynes	3333-3267(s) stretch
		700-610(b) bend
C=C	Alkenes	1680-1640 (m,w) stretch
C≡C	Alkynes	2260-2100(w,sh) stretch
C=C	Aromatic Rings	1600, 1500(w) stretch
C-O	Alcohols, Ethers, Carboxylic acids, Esters	1260-1000(s) stretch
C=O	Aldehydes, Ketones, Carboxylic acids, Esters	1760-1670(s) stretch
O-H	Monomeric, Alcohols, Phenols	3640-3160(s,br) stretch
	Hydrogen-bonded -- Alcohols, Phenols	3600-3200(b) stretch
	Carboxylic acids	3000-2500(b) stretch
N-H	Amines	3500-3300(m) stretch
		1650-1580 (m) bend
C-N	Amines	1340-1020(m) stretch
C≡N	Nitriles	2260-2220(v) stretch
NO_2	Nitro Compounds	1660-1500(s) asymmetrical stretch
		1390-1260(s) symmetrical stretch

v- variable, m- medium, s- strong, br- broad, w- weak.

2.3. Analysis Methods

The interpretation of infrared spectra involves the correlation of absorption bands in the spectrum of an unknown compound with the known absorption frequencies for types of bonds. Table 2 will help to become more familiar with the process. Significant for the identification of the source of an absorption band are intensity (weak, medium or strong), shape (broad or sharp), and position (cm^{-1}) in the spectrum. Characteristic examples are provided in the table below to assist becoming familiar with the intensity and shape absorption bands for representative absorptions.

The semiquantitative analysis of samples used in transmission studies is based on the Beer-Lambert law, which gives the relation between the observed intensity of a given band and the content of the corresponding functional group [15].

The Beer-Lambert Law can be described as equation (1)

$$A = \varepsilon bc \tag{1}$$

The transmission and absorption are related by equation (2)

$$A = 2 - \log_{10}(\%T) \tag{2}$$

Content of chemical bonds can be computed by equation (3)

$$c = \frac{2 - \log_{10}(\%T)}{\varepsilon b} \tag{3}$$

where
A   absorbance
ε molar absorbtibity L mol^{-1}cm^{-1}
b   path length of the sample cm
c   content of the compound in solution mol L^{-1}

3. EXPERIMENTAL RESULTS AND DISCUSSION

Many factors that impact the soot formation in stoker-fired boiler have been studied in this fixed bed, such as temperature, residence time of gases, and coal type. The results of these experiments are recited as below.

3.1. Effects of Temperature

Temperature is a critical parameter in soot formation. The higher temperature is, the easier that the volatiles are released from the coal surface is. Large PAH (soot precursors) start to form soot at about 1300 K for simple hydrocarbon flames, and for a complicated

compound like coal, the incipient temperature for soot precursor formation can be as low as 1100 K [16]. The viscosity of the aerosol samples changed from viscous to incompact as the temperature increased. The sample viscosity serves as a rough visual metric for the extent of soot formation. Figure 2 shows Infrared spectra from 400 to 4000 cm^{-1} derived form original coal and soot aerosols with the temperature increase. The infrared spectra of sample gases, in the certain experimental conditions that the residence time of gases is 1.76 seconds and the Shenmu coal is selected, have been examined by FTIR and analyzed. Analysis of chemical bond of original coal and soot aerosol in different temperature are listed in table 3.

1. Hydrogen-Bonded, Such as C-H, O-H and N-H in Stretching Vibration in 4000-2500 Cm^{-1}

According to Beer-Lambert Law, contents of chemical bonds can be computed. Hydrogen bonds, such as O-H and N-H in stretching vibration, are in 3750-3000 cm^{-1}. The content of O-H and N-H in original coal sample is 0.0107 mol L^{-1}; the content of O-H and N-H in soot aerosol samples are high. The content of soot aerosol derived from lump-coal pyrolysis in 1600 K is 0.0278 mol L^{-1}; the content of soot aerosol derived from lump coal pyrolysis in 1400 K is 0.0258 mol L^{-1} ; and the content of soot aerosol derived from lump coal pyrolysis in 1100 K is 0.0238 mol L^{-1}

C-H bond is in 3000-2700 cm^{-1} bands, for example CH$_3$-. The transmission of original coal sample by infrared ray is high; however, the transmission of soot aerosol sample derived from lump-coal pyrolysis in different temperature by infrared ray is low.

Table 3. Analysis of chemical bond of original coal and soot aerosols in different temperature

Chemical bond	Original Coal	Soot aerosol (1100K)	Soot aerosol (1400K)	Soot aerosol (1600K)
hydrogen-bonded ; -O-H, -N-H stretching vibration	0.01069179	0.0237895	0.02577849	0.02782832
C≡C, -C≡N; stretching vibration	0.01084181	0.02314466	0.01711051	0.0151355
aromatic, unsaturated aliphatic bond; C=C	0.12484481	0.04776454	0.0964091	0.18396389

Content of C-H in original coal is more than that in soot aerosols; and content of O-H and N-H in original coal is less than that in soot aerosols. It may be caused by that bond energy of C-H in volatiles is low (414KJ/mol); the bond is easy to be unfolded and release H$^+$; some of H$^+$ are emitted as hydrogen; and some of H$^+$ are united with oxygen or nitrogen elements to form O-H and N-H.
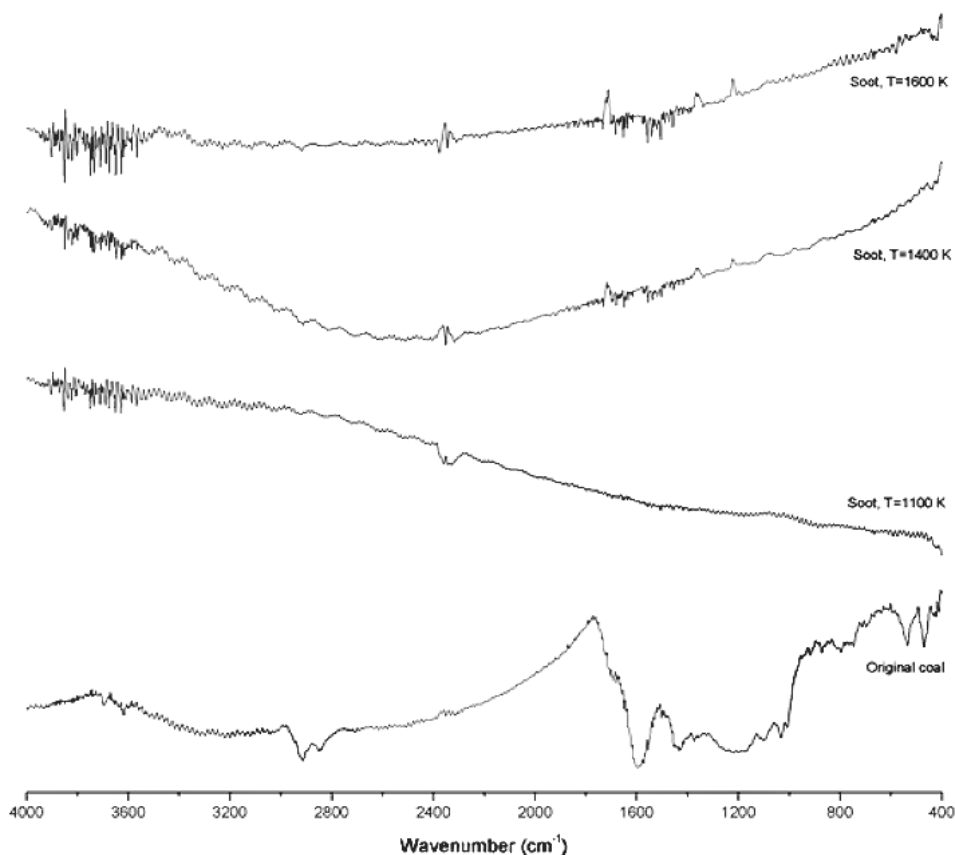
Figure 2. Infrared spectra for original coal and soot aerosols in different temperature.

The contents of O-H and N-H in soot aerosol are increased with temperature increase. There is a different degree about connection between H+ and nitrogen, oxygen elements. The bond energy of O-H is 464KJ/mol, so unfold the chemical bond needs more energy. The change of O-H and N-H contents is an integrative competition result of rupture and connection of chemical bonds.

2. C≡C and C≡N in Stretching Vibration in 2500-2000 Cm⁻¹

The content of C≡C and C≡N bonds derived from original coal sample is low, and the content of C≡C and C≡N bonds derived from soot aerosol samples are higher than that from original coal sample. The content of C≡C and C≡N bonds derived from soot aerosol sample in 1100K is higher than that in 1400K and 1600K. The content of C≡C and C≡N is decreased with temperature increase, because some of unsaturated hydrocarbon react with other chemical bonds and can be consumed in higher temperature.

3. Aromatic Rings in 1600-1450 Cm⁻¹

The content of aromatic rings derived from original coal is high, and aromatic rings are released in coal pyrolysis process. The content of aromatic rings derived from soot aerosol

samples is 0.0478, 0.0964, and 0.1839 mol L^{-1} in 1100K, 1400K, and 1600K respectively. The content of aromatic rings derived from soot aerosol samples is increased with temperature increase. That may be caused that more hydrocarbons have been transformed to aromatic rings by cyclization reaction.

3.2. Effects of Residence Time of Gases

Residence time of gases is also an effect factor on soot formation. The residence time have been altered by regulating the height of fixed bed in the reactor and the inlet gases velocity. Figure 3 shows Infrared spectra from 400 to 4000 cm^{-1} derived form original coal and soot aerosols with residence time of gases extension. The infrared spectra of sample gases, in the certain experimental conditions that the temperature is 1600K and the Shenmu coal is selected, have been examined by FTIR and analyzed. Analysis of chemical bond of original coal and soot aerosol in different temperature are listed in table 4.

1. Hydrogen-Bonded, Such as C-H, O-H and N-H in Stretching Vibration in 4000-2500 Cm^{-1}

According to Beer-Lambert Law, contents of chemical bonds can be computed. Hydrogen bonded, such as O-H and N-H in stretching vibration, are in 3750-3000 cm^{-1}. The content of O-H and N-H in original coal sample is 0.0107 mol L^{-1}; the content of O-H and N-H in soot aerosol samples are high. The content of soot aerosol derived from lump-coal pyrolysis is 0.0145 mol L^{-1} when residence time of gases is 1.52s; the content of soot aerosol derived from lump-coal pyrolysis is 0.0278 mol L^{-1} when residence time of gases is 1.76s; and the content of soot aerosol derived from lump-coal pyrolysis is 0.0312 mol L^{-1} when residence time of gases is 2.00s.

C-H bond is in 3000-2700 cm^{-1} bands, for example CH_3-. The transmission of original coal sample by infrared ray is high; however, the transmission of soot aerosol sample derived from lump-coal pyrolysis in different residence time by infrared ray is low.

Content of C-H in original coal is more than that in soot aerosols; and content of O-H and N-H in original coal is less than that in soot aerosols. It may be caused by that bond energy of C-H in volatiles is low (414KJ/mol); the bond is easy to be unfolded and release H^+ in longer residence time; some of H^+ are emitted as hydrogen; and some of H^+ are united with oxygen or nitrogen elements to form O-H and N-H.

The content of O-H and N-H are increased of soot aerosols derived from lump-coal pyrolysis with residence time extension. It may be the result that hydrogen is connected with oxygen and nitrogen in longer residence time.

2. C≡C and C≡N in Stretching Vibration in 2500-2000 Cm^{-1}

The content of C≡C and C≡N bonds derived from original coal sample is low, and the content of C≡C and C≡N bonds derived from soot aerosol sample are higher than that from original coal sample. It is 0.0135, 0.0151 and 0.0244 mol L^{-1}in 1.52s, 1.76s and 2.00s respectively. The contents of C≡C and C≡N bonds derived from soot aerosol samples are increased with residence time of gases extension because of more chemical bonds rupture and unsaturated hydrocarbons formation in longer residence time.
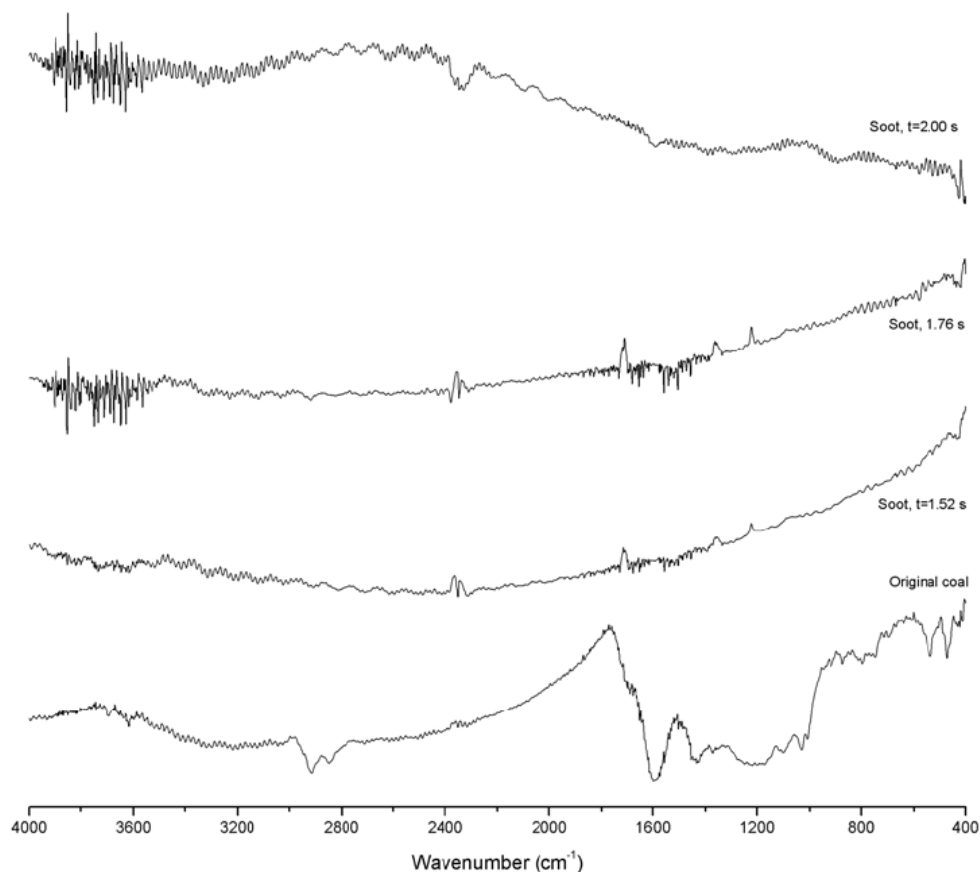
Figure 3. Infrared spectra for original coal and soot aerosols in different residence time.

**Table 4. Analysis of chemical bond of original coal and
soot aerosols in different residence time**

Chemical bond	Original Coal	Soot aerosol (1.52s)	Soot aerosol (1.76s)	Soot aerosol (2.00s)
hydrogen-bonded ; -O-H, -N-H stretching vibration	0.01069179	0.014519	0.02782832	0.031227178
C≡C, -C≡N; stretching vibration	0.01084181	0.013463	0.0151355	0.024474877
aromatic, unsaturated aliphatic bond; C=C	0.12484481	0.162823	0.18396389	0.100757655

3. Aromatic Rings in 1600-1450 Cm^{-1}

The content of aromatic rings derived from original coal is high, and aromatic rings are released in coal pyrolysis process. The content of aromatic rings derived from soot aerosol samples is 0.1628, 01839, and 0.1008 mol L^{-1} in 1.52s, 1.76s, and 2.00s respectively. The

content of aromatic rings derived from soot aerosol samples is increased first and decreased late with residence time of gases extension. That may be caused that more hydrocarbon have been transformed to aromatic rings; and then some of aromatic rings aggregate and form PAH; and the contents of aromatic rings C=C is decreased.

3.3. Effects of Coal Type

From the proximate analysis of the three coals shown in table 1, volatiles content of dry-ash-free coal, including Shenmu, Datong, and Zibo are 35.70%, 33.17%, and 39.10% respectively. At the same heating rates, more volatiles are released from the high-volatiles coal. The experimental condition is that the residence time of gases is 1.76 seconds, and the temperature in the reactor is 1600 K. Figure 4 shows Infrared spectra from 400 to 4000 cm^{-1} derived from original coal and soot aerosols with coal type change. Analysis of chemical bond of original coal and soot aerosols in different temperature are listed in table 5.
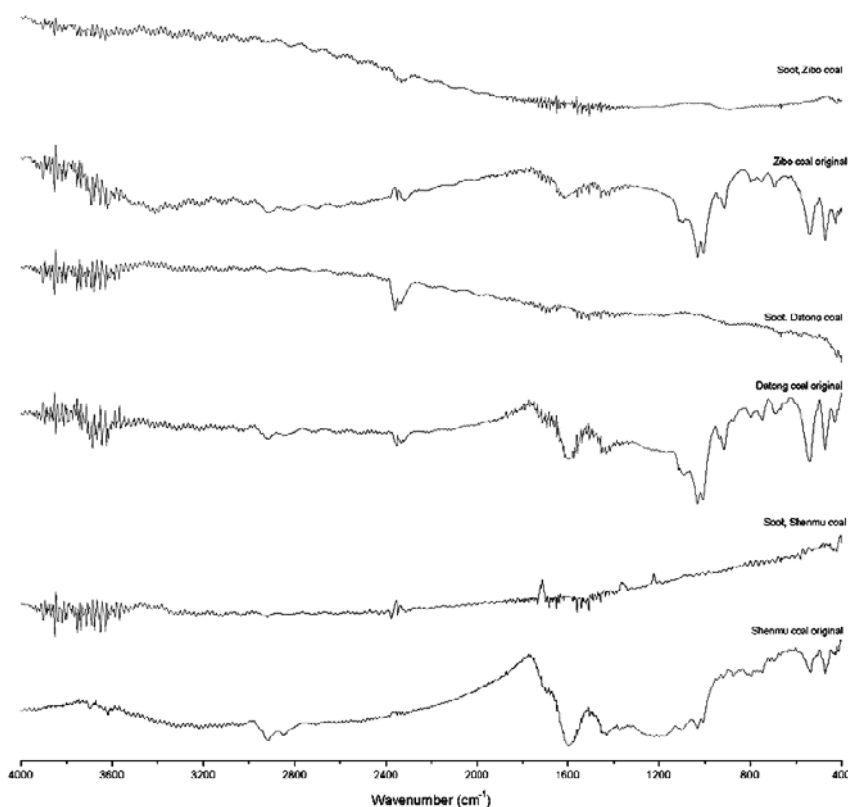
Figure 4. Infrared spectra for original coals and soot aerosols in different coal type.

1. Hydrogen-Bonded, Such as C-H, O-H and N-H in Stretching Vibration in 2500-2000 Cm^{-1}

According to Beer-Lambert Law, contents of chemical bonds can be computed. Hydrogen bonded, such as O-H and N-H in stretching vibration, is in 3750-3000 cm^{-1}. The

contents of Shenmu, Datong, and Zibo coal samples are 0.0107, 0.0291, and 0.0236 mol L^{-1} respectively; the contents of O-H and N-H of soot aerosol samples derived from Shenmu, Datong, and Zibo coal is 0.0278, 0.0256, and 0.0160 respectively.

C-H bond is in 3000-2700 cm^{-1} bands, for example CH$_3$-. The transmission of original coal sample by infrared ray is high; however, the transmission of soot aerosol samples derived from lump-coal pyrolysis in different coal type by infrared ray is low.

Content change of C-H, O-H and N-H in original coals and in soot aerosols are connected with coal type. There is a difference on rupture of alkyl group in different coal. That some of H$^+$ are united with nitrogen or oxygen elements to form O-H and N-H in different coal type is different.

Table 5. Analysis of chemical bond of original coals and soot aerosols in different coal type

Chemical bone	coal (S)	Soot aerosol (S)	coal (D)	Soot aerosol (D)	coal (Z)	Soot aerosol (Z)
hydrogen-bonded ; -O-H, -N-H stretching vibration	0.01069	0.027828321	0.02914	0.025617812	0.02361	0.01604796
C≡C, -C≡N; stretching vibration	0.01084	0.01711051	0.01453	0.023692205	0.01325	0.016876428
aromatic, unsaturated aliphatic bond; C=C	0.12484	0.18396389	0.11846	0.08606508	0.05066	0.06299007

2. C≡C and C≡N in Stretching Vibration in 2500-2000 Cm^{-1}

The contents of C≡C and C≡N bonds derived from Datong and Zibo coal sample are higher than that from Shenmu coal for the case of different chemical structure of coals. The contents of C≡C and C≡N bonds derived from soot aerosol sample are higher than that from original coal samples. The cause may be that more chemical bond rupture and unsaturated hydrocarbon formation in pyrolysis process.

3. Aromatic Rings in 1600-1450 Cm^{-1}

Aromatic rings are released in coal pyrolysis process. The content of aromatic rings derived from soot aerosol samples is different. That may be caused that more hydrocarbon have been transformed to aromatic rings. The result can be interpreted by that parts of aliphatic hydrocarbons have been possibly converted to aromatic hydrocarbons by reason of chemical bond reforming. And it is affected by chemical structure of coals

4. CONCLUSION

The present study examines the lump coal-derived soot formation during pyrolysis in a fixed bed. Based on the experimental study on functional groups of original coals and soot aerosols in the experimental facility by FTIR, the conclusion can be drawn: contents of

hydrogen-bonded are increased, contents of unsaturated hydrocarbons are decreased, and contents of aromatic hydrocarbons are increased with temperature increase; contents of hydrogen-bonded are increased, contents of unsaturated hydrocarbons are increased, and contents of aromatic hydrocarbons are increased early and decreased late with residence time extension; and the contents of unsaturated hydrocarbons derived from soot aerosol samples are higher than those from original coal samples, and contents of hydrogen bonded and aromatic hydrocarbons are different depending on chemical structure of original coals.

ACKNOWLEDGMENT

REFERENCES

[1]   He BL, Song Q, Chen CH, Xu XC. (2003). *Investigations on mechanism of soot formation during combustion and control of soot emission.* The 5th international symposium on combustion, China. 1-5.

[2]   Ma JL, Fletcher TH, Webb BW. (1995). *Effect of flame environment on soot formation in coal combustion.* International conference on coal science, 869-872.

[3]   Zhang HF. (2001). Nitrogen Evolution and Soot Formation during Secondary Coal Pyrolysis Ph.D. dissertation. Department of Chemical Engineering, Brigham. Young University.

[4]   Fletcher TH, Ma JL, Rigby JR, Brown AL, Webb BW. (1997). Soot in coal combustion systems. *Progress in Energy and Combustion Science.* 23: 283-301.

[5]   Lee EJ, Oh KC, Shin HD. (2005) Soot formation in inverse diffusion flames of diluted ethane. *Fuel.* 84(5): 543-550.

[6]   Inal F, Tayfur G, Melton TR, Senkan SM. (2003) Experimental and artificial neural network modeling study on soot formation in premixed hydrocarbon flames. *Fuel.* 82(12): 1477-1490.

[7]   Aksit IM, Moss JB. (2005) Model fuels to reproduce the sooting behaviour of aviation kerosene. *Fuel.* 84(2-3): 239-245

[8]   Nenniger RD. Nenniger, (1986) *Aerosols produced from coal pyrolysis,* M.S. thesis, Department of chemical Engineering, Massachusetts Institute of Technology.

[9]   Wornat MJ, Sarofim AF, Longwell JP. (1987). *Energy and fuel.* 1(5):431-437.

[10]   Chen JC. (1991). Effect of secondary reactions on product distribution and nitrogen evolution from rapid coal pyrolysis, Ph. D. thesis, Stanford University.

[11]   M. J. Tan, J. X. Mao, (2004) *The advance technologies of coal-fired industrial boiler in China.* The conference of the China-America advanced technologies of industrial boiler, Beijing, pp1-17

[12]   Sun D, Choi S. (2000). The Combustion of Simulated Waste Particles in a Fixed Bed. Combustion and flame; 121(1, 2): 167–180.

[13]   Ford N, Cooke MJ, Pettit MD.(1992) The Use of Laboratory Fixed-Grate Furnace to Simulate Industrial Stoker-Fired Plant. Institute of Energy. 65(9): 137–143.

[14]   Staley GP, Bradshaw FW, Carrel CS, Pershing DW. (1985) The Influence of Bed-Region Stoichiometry on Nitric Oxide Formation in Fixed-Bed Coal Combustion. Combustion and flame. 59(2): 197–211.

[15]   M. D. Guillen, M. J. Iglesias, A. Dominguez, and C. G. Blanco. (1992) Semiquantitative FTIR Analysis of a Coal Tar Pitch and Its Extracts and Residues in Several Organic Solvents. *Energy and Fuels,* 6, 518-525.

[16]   Solum MS, Sarofim AF, Pugmire RJ, Fletcher TH, Zhang HF. (2001) ^{13}C NMR Analysis of Soot Produced from Model Compounds and a Coal. *Energy and Fuels* 15(4): 961-971.

INDEX

D

E

J

K

L

W

Y