



***CENTER FOR ADVANCED PROCESS
DECISION-MAKING***

New Directions for Process Systems Engineering

**Lorenz T. Biegler
Ignacio E. Grossmann
Nikolaos V. Sahinidis
B. Erik Ydstie**

***Department of Chemical Engineering
Carnegie Mellon University
Pittsburgh, PA 15213***

January, 2008

TABLE OF CONTENTS

General News	4
Executive Summary	7
 Status of Research Projects	
 <u>Larry Biegler's Group</u>	
<i>General Frameworks for Large Scale Optimization Strategies and Applications</i>	10
<i>Mathematical Programs with Equilibrium Constraints (MPECS)</i>	11
<i>Simultaneous Optimization of Differential-Algebraic (DAE) Systems</i>	12
<i>Large-Scale Optimization for Fuel Cell Models</i>	14
<i>Dynamic Optimization for Semi-closed Process Models</i>	14
<i>Reduced Order Models for PDE-based Units in Power Plant Flowsheets</i>	15
<i>Optimization and Control of Periodic Adsorption Processes</i>	16
 <u>Ignacio Grossmann's Group</u>	
<i>Open Source Code for MINLP Optimization</i>	16
<i>Algorithms for Nonlinear Disjunctive Programming</i>	17
<i>Global Optimization of Integrated Process Water Systems</i>	18
<i>Energy Optimization of Corn-based Ethanol Plants</i>	19
<i>Optimal Synthesis of Integrated Gasification Combined Cycle (IGCC) Systems</i>	19
<i>MINLP Flowsheet Optimization with Process Simulators</i>	21
<i>Synthesis of Crystallization Processes</i>	21
<i>Design for Sustainability</i>	22
<i>Design and Planning of Deep-Water Oilfield Development under Uncertainty</i>	23
<i>Simultaneous Planning and Scheduling of Multiproduct Plants</i>	24
<i>Design and Planning of Responsive Supply Chains</i>	26
<i>Optimal Scheduling of Crude Oil Operations</i>	27
<i>Planning of Refinery Operations</i>	28
<i>Planning for Petroleum Allocation</i>	29
<i>Scheduling of Batch Multiproduct Plants</i>	30
<i>Scheduling of Paper Machines with Replacement of Felts</i>	30
<i>Simultaneous Scheduling and Control</i>	31
<i>Software for MINLP Optimization in Design and Scheduling</i>	32
 <u>Nick Sahinidis' Group</u>	
<i>THRUST 1 – Optimization Algorithms, Theory and Software</i>	
<i>Algorithms and Software for Global Optimization of NLPs and MINLPs</i>	33
<i>Global Optimization for Nonlinear Least Squares Problems</i>	33
<i>Global Optimization for Machine Learning Problems</i>	34
<i>Algorithms and Software for Black-box Optimization</i>	34
<i>Algorithms and Software for Linear Optimization Problems (LP)</i>	35
 <i>THRUST 2 – Applications of Optimization in Biology, Chemistry, Engineering and Medicine</i>	
<i>Inverse Imaging Problems in X-ray Crystallography</i>	36

<i>Protein Side-chain Conformation Problem</i>	36
<i>Protein Structural Alignment</i>	37
<i>Combinatorial Library Design</i>	37
<i>Protein-Ligand Docking</i>	38
<i>Steady State Optimization with Guaranteed Robust Stability under Parametric Uncertainty</i>	38
<i>DNA Sequencing by Hybridization</i>	38
<i>Protein Binding Site Identification</i>	38
<i>A Branch-and-Bound Algorithm for the Continuous Facility Layout Problem</i>	39
<i>Portfolio Optimization for Wealth-dependent Risk Preferences</i>	39
<i>Medical Diagnosis and Prognosis</i>	39
<i>Design of Novel, Environmentally Benign Chemicals</i>	40

Erik Ydstie's Group

<i>Real Time Optimization and Distributed Adaptive Control</i>	41
<i>Passivity Based Control of Multi-Phase Reactor Systems</i>	41
<i>Modeling the Vapor Recovery Reactor in Carbothermic Aluminum Production using Multi-scale Modeling Methods</i>	41
<i>Process Networks</i>	42
<i>Multi-Scale Modeling of Particulate Processes with Fluid Flow</i>	42

Publications	43
---------------------	----

Reprints	44
-----------------	----

GENERAL NEWS

Gary Powers. A memorial for Gary Powers took place at the Singleton Room, Carnegie Mellon, on September 13, 2007. The memorial was very well attended and provided a fitting memory for our dear colleague. On October 2, 2007, Jeff Siirola gave a Special PSE Seminar in which he described the pioneering work that he did with Gary Powers as part of their Ph.D. work with Dale Rudd at the University of Wisconsin. Also, a nice obituary on Gary was published in Chemical Engineering Progress, p.18, Dec. 2007, which was written by Cindy Mascone, a CMU alumnus that took the course Introduction to Chemical Engineering with Gary Powers. An undergraduate scholarship has been established in the memory of Gary. Donations may be sent for the Gary J. Powers Scholarship Fund, c/o Toni McIlTrot, Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15222-5488.

AIChE Meeting. The four faculty of the CAPD gave a total of **27 presentations** at the Annual AIChE Meeting in Salt Lake City, November 5-9, 2007. These are listed in the CAPD e-Newsletter that we sent on October 30, 2007.

Larry Biegler gave a talk at the AIChE Meeting in Salt Lake City as recipient of the 2007 CACHE Award for Excellence in Computing in Chemical Engineering Education. Larry was recognized for leadership in the development of strategies and methods for process optimization, particularly for the pioneering implementation of the successive quadratic programming (SQP) method in the process simulator FLOWTRAN that has allowed many students to optimize process flowsheets.

In December, Larry spent two weeks in India as the IIT Bombay Golden Jubilee Distinguished Lecturer. In addition to seminars at IIT Bombay, UICT/Mumbai and IIT Madras, he was an invited keynote speaker at the International Conference on Advanced Energy Research in Mumbai and at the Advances in Optimization Workshop at the GE Global Research Center, Bangalore, India.

Ignacio Grossmann received on September 18, 2007, the degree Doctor Honoris Causa from the University of Maribor in Slovenia for his work in process systems engineering and mixed-integer nonlinear programming. Ignacio also gave the ExxonMobil Lecture, "Research Challenges in Process Systems Engineering and its Potential Impact in the Process Industry," at the University of Massachusetts, Amherst on December 4, 2007. He also gave seminars at the Center for Operations Research and Economics, University of Louvain, Belgium, on August 30, and several talks at the Universidad Iberoamericana on December 14. Finally, he chaired the meeting of section 3 (Chemical Engineering) at the meeting of the National Academy of Engineering on October 1.

Nick Sahinidis who joined the Department of Chemical Engineering and the CAPD on August 1, 2007, was named John E. Swearingen Professor of Chemical Engineering, effective January 1, 2008. This past semester, Nick gave seminars on global optimization in the Department of Mathematics at the Massachusetts Institute of Technology (October 2007), the Department of Industrial Engineering of the University of Pittsburgh (October 2007), and the Enterprise-Wide Optimization Seminar Series at Carnegie Mellon (October 2007). Nick is initiating his research program at CMU, where he picked seven new graduate students to join his group this past semester. He is also supervising four graduate students at the University of Illinois.

Erik Ydstie received the 2007 AIChE Computing in Chemical Engineering CAST Award at the AIChE Meeting in Salt Lake City, where he gave the after dinner talk "What is a Model?" This award recognizes outstanding contributions in the application of computing and systems technology to chemical engineering. He also presented a paper with the title "New Vistas for Process Control: Integrating Physics and Communication Networks". Last semester Erik had a leave from teaching while he focused on developing his start-up company. There are at present 3 full time and 2 part employees in ILS working on adaptive and self-optimizing solutions for the power generation industries.

Erik accepted a three year engagement as Professor II of Electrical Engineering in the Institute of Cybernetics at the Norwegian University of Science and Technology. He will be working on Integrated Operations in the Oil and Gas field production in the North Sea with Statoil/Hydro and the IO Center at NUST. He presented a plenary lecture on "Optimality of Integrated Process and Financial Networks" at the IO conference in Trondheim on September 1.

It is a pleasure to welcome the following new graduate students to the PSE group: **Rodolfo Lopez Negrete de la Fuente** from Universidad Iberoamericana, will be working for his Ph.D. degree with Larry Biegler; **Sebastian Terrazas**, from Universidad Iberoamericana will be working for his Ph.D. degree with Ignacio. **Keith Zorn** from Rensselaer Polytechnic Institute will be working for his Ph.D. degree with Nick Sahinidis. **Shweta Shah** from Indian Institute of Technology, Bombay, will also be working for her Ph.D. with Nick. **Saurabh Awasthi** (from Indian Institute of Technology, Delhi), **Deepak Channamariyappa** (from National Institute of Technology Karnataka, Surath), **Apurva Samudra** (from Indian Institute of Technology Roorkee), **Xuan Shi** (from Shanghai Jiao Tong University), and **Yiqi Zhu** (from Shanghai Jiao Tong University) will be working for their M.S. with Nick. Master students **Harshad Ghodke** and **Siddarth Raman** have joined Erik's research group to work on PPG related projects.

Congratulations to Ignacio's students, **Muge Erdirik** who defended her Ph.D. degree on August 28, and to **Ramkumar Karuppiiah** who defended his Ph.D. degree on September 13. Muge has joined Amazon in Seattle and Ram has joined The Dow Chemical Company in Freeport.

Congratulations to **Anshul Agarwal**, student of Larry, and **Fengqi You**, student of Ignacio, who passed their Ph.D. qualifying exam.

Dr. Gonzalo Guillen, who collaborated with Ignacio in the area of supply chain optimization with sustainability concerns, and gave the last EWO seminar on agent-based systems, completed his visit on December 31. He has joined the faculty of the Chemical Engineering Department of the Universitat Roviri e Virgili in Tarragona, Spain.

Professor Sebastian Engell from the University of Dortmund, will visit the PSE group and CAPD in February and March, 2008. He will give the Bayer Lecture on Process Systems Engineering on March 6, 2008, and will also participate at the special session on EWO at the CAPD meeting on March 11.

CAPD e-News. We sent on October 30 our e-newsletter, which we included all the CAPD presentations at the AIChE Meeting in Salt Lake City. We issue on a yearly basis two extensive newsletters with detailed descriptions of the projects in August and January, while the two others are brief e-Newsletters in April and October that contain short announcements as well as information that we think will be useful to the CAPD members. We would appreciate receiving feedback or suggestions for improving the e-Newsletter.

2008 ANNUAL REVIEW MEETING

The Annual Review Meeting will be held on **March 10-11, 2008**. The first day of the meeting will consist of overviews given by Larry, Ignacio, Nick and Erik, followed by a discussion with industrial participants, and a poster session by the students. As we did last year, we will have a reception on Sunday evening, March 9, at the Danforth Lounge, University Center, and a group dinner on Monday, March 10, at Monterey Bay Inn. The second day is devoted to final year student presentations. As a special feature this year there will be a session on Enterprise-wide Optimization on Tuesday, March 11, 2:00-4:30PM, which will feature as speakers Prof. Sebastian Engell from the University of Dortmund, Prof. Tom Marlin from McMaster and Prof. Sridhar Tayur from the Tepper School of Business and Smartops. This session is actually part of the Enterprise-wide Optimization project (see next section), but will be open to all CAPD members as we did last year.

The detailed agenda of the meeting will be sent very soon. Also, if you are interested in giving a short presentation, please let us know. We promise we will give more time than what we did last year. Also, if you have any suggestions for the format of the meeting, please let us know.

ENTERPRISE-WIDE OPTIMIZATION INTEREST GROUP

The special interest group of the CAPD on Enterprise-wide Optimization is part of the project "**Computational Models and Algorithms for Enterprise-wide Optimization (EWO) of Process Industries**" that has been funded by the Pennsylvania Infrastructure Technology Alliance. The current participating companies are ABB, Air Products, BP America, Dow Chemical, ExxonMobil, NOVA Chemicals and Total. These companies have supplied case studies that are being undertaken by the EWO faculty and students. The project involves a total of 7 Ph.D. students, and 5 faculty (CMU: **Larry Biegler, Ignacio Grossmann, John Hooker**; Lehigh: **Larry Snyder**; UPitt: **Andrew Schaeffer**; Wisconsin: **Jeff Linderoth**);). The next meeting of this group will take place on **March 11-12**, immediately following the CAPD Meeting. As indicated above, the EWO meeting will start at 2:00 PM with the special session on Enterprise-wide Optimization. Companies that might be interested in joining this group in fiscal year 2008, please contact Ignacio Grossmann. The membership fee to join this group is \$12,500 for members of the CAPD. A description of the EWO project can be found at <http://egon.cheme.cmu.edu/ewocp/>

NEW 2008 CAPD SHORT COURSE <http://capd.cheme.cmu.edu/shortcourse.html>

The short course, "Optimization Modeling and Integrated Process Operations," which is scheduled this year on **June 12-18, 2008**, has been reorganized and now involves Larry, Ignacio, Nick and Erik. There are a total of six modules that can be taken in any combination over the 6-day period. The short course will consist of two parts: *Optimization Modeling* (Modules A - C) that focuses on modeling and algorithms with applications to process optimization, process synthesis and molecular design, and *Integrated Process Operations* (Modules D-F) that focuses on the three major decision levels in plant and enterprise-wide optimization. The modules are offered are the following: A: Nonlinear Programming (Biegler, June 12), B: Mixed Integer and Disjunctive Programming (Grossmann, June 13), C: Global Optimization and Optimization Under Uncertainty (Sahinidis, June 14), D: Mixed-integer Models for Planning and Scheduling (Grossmann, June 16), E: Process Dynamics and Control (Ydstie, June 18), F: Differential / Algebraic Models for Real Time Optimization (Biegler, June 18). Recall that CAPD members receive a 25% discount. The course includes workshops where participants obtain hands-on experience formulating problems and using various software packages. Course materials include extensive notes, the GAMS software, and our textbook "Systematic Methods of Chemical Process Design." If you are interested in attending this course, please contact Toni McIltrout at 412-268-3573, or e-mail: tm21@andrew.cmu.edu.

PAN AMERICAN ADVANCED STUDY INSTITUTE ON EMERGING TRENDS IN PSE

Ignacio Grossmann is organizing with Frank Doyle, Jaime Cerda and Argimiro Secchi the Pan American Advanced Studies Institute (PASI) on emerging trends in Process Systems Engineering on August 12-21, 2008, at Hotel Costa Galana, Mar del Plata, Argentina. The proposed workshop is aimed at advanced graduate and post-doctoral students, and will emphasize the latest developments and research challenges in the emerging areas of *sustainability, energy, biological systems, multiscale systems, and enterprise-wide optimization*. This workshop is being funded by NSF and the Department of Energy. We are grateful for the additional support that we received from GAMS Corporation, Cargill, Pfizer, PPG and Procter & Gamble. This will allow the participation of 50 Ph.D. students from the Americas (50% from U.S.). The PASI workshop will be a follow-up to the successful PASI on Process Systems Engineering that took place on in Iguazú, Argentina, in 2005. The website for this meeting is: <http://cepac.cheme.cmu.edu/pasi2008>

VIRTUAL LIBRARY ON PROCESS SYSTEMS ENGINEERING

The Pan American Study Institute on Process Systems Engineering that was organized by **Ignacio Grossmann** took place in Iguazu, Argentina on August 16-25, 2005. The workshop addressed four major areas: Optimization, Process and Product Design, Process and Supply Chain Operations, and Process Dynamics and Control. A major outcome of the PASI conference was the development of virtual library on Process Systems Engineering covering the four areas. The library consists of the Powerpoint slides of the presentations, background articles, as well as exercises and MATLAB and GAMS computer files for various applications, and a practice exam with solutions. The virtual library can be found in, <http://cepac.cheme.cmu.edu/pasilectures.htm>

WEBSITES/PROCESS SYSTEMS DIRECTORY

We are happy to report that the CAPD newsletter is distributed in electronic form. We would appreciate receiving feedback from our member companies of our CAPD website, <http://capd.cheme.cmu.edu>. This website provides a number of interesting items including information about individual groups, industrial success stories, software available from the group, etc. Work is under way to update and modify our website. Other websites of interest are Erik's <http://mongol.cheme.cmu.edu/>, Ignacio's <http://egon.cheme.cmu.edu>, Larry's <http://dynopt.cheme.cmu.edu>, and Nick's <http://www.andrew.cmu.edu/user/ns1b/>. Also, the department maintains the website of CEPAC (Chemical Engineering Pan American Collaboration), the organization that has been formed to promote collaboration between the US and Argentina, Brazil and Chile. You can find in <http://cepac.cheme.cmu.edu/country.htm> a listing of all universities and researchers with links in the process systems area in the US, Argentina, Brazil, Canada, Mexico and Chile.

EXECUTIVE SUMMARY

Larry Biegler's Group

Since the last newsletter, a number of advances have been made in application of large-scale optimization, particularly for dynamic systems. The past few months have seen the following applications.

Dynamic optimization tools, ultimately for on-line use, have been applied that are centered on IPOPT, which continues to be developed at IBM. These have seen a number of applications ranging from detailed fuel cell models, advanced chromatographic separation, polymer reactors and processes, gas pipeline networks and nonlinear model predictive control. **Victor Zavala** has made strong advances in parameter and system identification for large-scale polymerization processes, and has developed a very efficient parallel computation approach. This is described in a preprint below. In addition, Victor has developed optimization formulations for Nonlinear Model Predictive Control (NMPC) and Moving Horizon Estimation (MHE) that exploit the simultaneous approach as well as recent NLP features in IPOPT. As a result, large dynamic optimization problems used for NMPC and moving horizon estimation (MHE) have *on-line computation costs reduced by two orders of magnitude!*

Related to this work is the development of specialized decomposition strategies within the IPOPT framework. **Brian Baumrucker** has developed interesting optimization formulations that incorporate complementarity constraints and allow a class of discrete decisions to be modeled within NLP problems. This approach is not restricted to particular NLP solvers and has been demonstrated on a variety of process applications. A preprint that describes this approach is listed below.

Finally, reduced order modeling strategies have been developed for large-scale scale PDE-constrained optimization. These are based on a snapshot-based decomposition and the isolation of principal components (i.e., eigenfunctions) that constitute the dominant elements of the solution space. Using these components as basis functions, one can create accurate reduced order models (ROMs) that are much smaller and easier to solve than the original. This approach, originally due to Karhunen and Loeve has been

used in two ways in this research group. In the first, we have applied ROM strategies to Fluent-based models to create reduced order optimization models for gasifiers and combustors. As described below and implemented by **Yi-dong Lang**, they make use of empirical input-output relationships described by neural network models. In the second approach, ROM strategies have been applied to pressure-swing adsorption (PSA) models. Here **Anshul Agrawal** has used these basis functions in a Galerkin framework to develop much more efficient models for gas adsorption. Both lead to much more efficient optimization strategies for dynamic systems.

Enabling Software From Larry Biegler's Group

Highlights of **Larry Biegler's group** include further development of the interior point nonlinear programming algorithm, IPOPT, in a number of directions. **Andreas Wächter**, now at IBM, developed a novel and very efficient barrier (interior point) NLP method, based on many of our earlier SQP concepts. In the general purpose framework, it adopts a sparse decomposition of the KKT matrix, and has been tested on thousands of applications with up to a quarter of a million variables. It has also been used to solve problems that incorporate a number of specialized decomposition strategies and for a wide variety of applications with several million variables. IPOPT is available under an Open Source software license and the source code can be downloaded from <http://www.coin-or.org/Ipopt/index>.

Carl Laird, now at Texas A&M, and **Andreas Wächter** have rewritten IPOPT as an object-oriented code written in C++. The resulting C++ package was released for public distribution last August and is seeing regular updates by Andreas. The code contains a much simpler interface to modeling environments as well as a more efficient implementation of the algorithm. Finally, the object-oriented code allows for greater flexibility, especially in adopting additional linear solvers and advanced decomposition strategies.

The IPOPT code has already been interfaced to a number of modeling systems including:

- **GAMS** –This system has been a standard optimization platform at CMU and is widely used in industry. IPOPT has been linked to GAMS by colleagues at GAMS development using the COIN-OR interfaces. The MUMPS sparse linear solver was used in the standard release, although versions are available that allow any sparse matrix solvers to be applied (subject to 3rd party licenses). Moreover, GAMS provides exact first and second derivatives and exploits the full capabilities of IPOPT.
- **AIMMS** –This GAMS-like system has a number of extended features including a full graphical user interface for the development of specific user applications. Moreover, it provides exact first and second derivatives and exploits the full capabilities of IPOPT. **Yi-dong Lang** has nearly completed the interface to this system and is currently testing the interface. A similar interface to GAMS is also planned for January.
- **AMPL** – This GAMS-like system is the standard development platform in the Biegler group, has been available for the past two years and can be downloaded from the COIN-OR webpage. It has extensive features for linking to MATLAB and other environments in Unix and Windows. It also provides exact first and second derivatives and exploits the full capabilities of IPOPT.
- **CAPE-OPEN** – IPOPT has been developed under the MINLP object class and has been demonstrated with the CO-Tester.
- **Bonmin** – As described below, joint work with researchers at IBM and Tepper as well as Ignacio's and Larry's groups have integrated IPOPT and CBC to develop a public domain MINLP code. Developed principally by Dr. Pierre Bonami, the code has been tested on a suite of convex and nonconvex MINLP problems and is being public released. It is also part of the COIN-OR interface incorporated within GAMS.

In addition, Carl has extended this approach to apply internal decompositions that take advantage of linear decomposition schemes of the KKT matrix in IPOPT. This allows a seamless interface to different sparse linear solvers as well as advanced decomposition schemes. This has been prototyped on Schur complement decompositions of the KKT matrix for block bordered diagonal structures. This strategy has been demonstrated on a large-scale Errors-in-Variables-Measured (EVM) parameter estimation problem, which is described in a preprint below. Finally **Victor Zavala** has extended the IPOPT framework to deal with

NLP sensitivity analysis for perturbed optimization problems. This has proved to be very successful for the advanced-step NMPC and MHE strategies developed below.

Yi-dong Lang has developed a user-friendly version of dynamic optimization methods. In particular, they have extended these codes to exploit sparsity and are using them to tackle dynamic optimization problems including batch polymerization, crystallization and parameter estimation. This software, called *DynoPC*, runs under Windows with a GUI interface. It has been distributed to several CAPD members and parts of it can now be downloaded on the same website as IPOPT. Currently, this approach is being extended to include updated versions of IPOPT and will be enhanced further by the development of a number of interfaces to optimization modeling platforms including CAPE-OPEN. The reduced-space, FORTRAN version of the IPOPT code is also being extended by **Kexin Wang** to include recent developments in the IPOPT project.

Ignacio Grossmann's Group has been involved in the following developments:

In the area of optimization **Juan Pablo Ruiz**, in a new project, has been developing a new solution approach for on the global optimization of Bilinear Generalized Disjunctive Programming in which the lower bound is strengthened based on intersecting disjunctions following the recent PhD thesis of Nick Sawaya. In collaboration with **Francois Margot, Pietro Belotti and Larry Biegler**, the open source code *Couenne* for the global optimization of MINLP problems continues concentrating on branching rules and tightening of bounds for nonconvex NLP problems. **Aldo Vecchiatti**, work on LOGMIP for disjunctive programming has been incorporated as an option in GAMS (<http://www.ceride.gov.ar/logmip/>).

Ramkumar Karupiah completed his Ph.D. thesis, submitted the paper on bioethanol in collaboration with **Cargill** and **Mariano Martin**, and joined Dow Chemical. **Ravi Kamath**, in a new project, has developed a comprehensive superstructure for the optimization of IGCC plants, and is completing the development of a model for air separation unit so as to integrate them with the models for the gasifier and utility plant that he has developed previously. **Jose Maria Ponce Ortega** and **Arturo Jimenez** have completed the paper for extending the MINLP synthesis model for heat exchanger networks so as to handle isothermal streams as well as streams with change of phase. **Jose Caballero** has completed a manuscript using kirging models for the optimization of process flowsheets in order to handle implicit functions. **Ricardo Lima** has been completing a manuscript for optimizing superstructure for separation processes based on crystallization using aggregated models, and has compared his solutions with flowsheets described in the patent literature. **Gonzalo Guillen** is completing two manuscripts for the bicriterion optimization of process networks in which the two criteria being optimized are economics (Net Present Value) and the Eco-indicator 99, a measure of sustainability, given uncertainties in the emission and damage data .

Bora Tarhan, incorporated nonlinear production profiles for wells in his Lagrangean branch and bound method for solving a multistage stochastic programming model for the design and planning of oil infrastructure planning under uncertainties in the sand quality, size of the reservoir and breakthrough time. **Muge Erdirik** completed her Ph.D. thesis, joined Amazon, and submitted her paper in collaboration with John Wassick from Dow on a scheduling model of parallel reactors connected with finishing lines, and where the reactors work simultaneously as groups. **Sebastian Terrazas** is the new Ph.D. student who will follow Muge's work. **Fengqi You** successfully passed his Ph.D. qualifier and submitted a paper that incorporates probabilistic inventory models with safety stocks for optimizing the economics and responsiveness of process supply chains. He also has developed a Lagrangean decomposition for a joint location-inventory problem. **Sylvain Mouret**, in a new project, has developed symmetry breaking constraints through a constraint programming approach for a novel continuous time model for crude oil scheduling that relies on time slots, and obtains very quickly optimal or near optimal solutions for the original MINLP model. **Abdul Attalas**, in a new project, has been developing nonlinear aggregated CDU models and initialization schemes for refinery planning models in order to go beyond the conventional fixed yield and swing cut models. **Roger Rocha**, a postdoctoral fellow from PETROBAS, has joined our group and is addressing the solution of very large-scale MILP models for petroleum allocation. **Pedro Castro** submitted a manuscript for the simultaneous batching and scheduling of batch processes with sequence dependent changeovers. **Daniela Drummond**, a visitor from University of Campinas in Brazil,

has been developing an MINLP model for production scheduling and replacement of felts in paper machines. **Antonio Flores-Tlacuahuac** from Mexico and **Sebastian Terrazas** produced a manuscript for simultaneous design, scheduling, and optimal control of a methyl-methacrylate continuous polymerization reactor. Finally, **Rosanna Franco** has completed the PC and web interfaces for the new MINLP model for heat exchanger networks for isothermal streams and continues with the global optimization of integrated water systems by Ramkumar Karupiah.

Nick Sahinidis' Group

Nick's work falls into two main thrusts: (1) optimization theory, algorithms, and software and (2) applications of optimization to problems in biology, chemistry, engineering, and medicine. Research in optimization addresses the development of algorithms and software for: (a) global optimization of NLPs and MINLPs, (b) optimization of black-box models, and (c) linear programming. On the applications side, current activities address: (a) crystallographic computing, (b) protein structural alignment, (c) metabolic network modeling and design, and (d) design of compounds with desired properties.

Enabling Software From Nick Sahinidis' Group

Nick continues the development of BARON, primarily in collaboration with his former student **Mohit Tawarmalani**, who is currently an Associate Professor at Purdue's Krannert School of Management. In addition, students in Nick's group address theoretical and computational issues in global/local optimization. Their results are, from time to time, incorporated into BARON. The BARON software is currently available commercially under GAMS and AIMMS. In addition, a full blown version of the GAMS/BARON system is available entirely for free under the NEOS server for optimization. More information about BARON can be found at <http://www.andrew.cmu.edu/user/ns1b/baron/baron.html>.

Software for bioinformatics is also developed and maintained by Nick's group:

- CMOS: software implementing the Xie and Sahinidis algorithm for solving the contact map overlap problem for aligning the 3D structures of proteins
- R3: software implementing the Xie and Sahinidis residue-rotamer reduction for predicting the structure of protein side chains
- SBH: software implementing the Chang and Sahinidis algorithm for finding all near-optimal solutions of the combinatorial problem in DNA sequencing by hybridization

Nick's web site at <http://www.andrew.cmu.edu/user/ns1b/group/biosoftware.html> provides these codes as on-line solvers.

STATUS OF RESEARCH PROJECTS

Larry Biegler's Group

General Frameworks for Large Scale Optimization Strategies and Applications

Researcher: Victor M. Zavala (Ph.D. started Fall, 2004)

Visitor: Kexin Wang, Zhejiang University

As a result of several generations of Ph.D. students, we now have a number of very efficient NLP algorithms including rSQP, full-space SQP, parameter estimation, data reconciliation and multiperiod optimization. As described above, the most recent incarnation is IPOPT, with continuing development by Andreas Wächter. This code has become a core algorithm for many of the research projects described below. In his PhD thesis, Carl Laird has applied IPOPT and the AMPL interface to develop a novel source detection strategy for municipal water networks as an important application for barrier methods. This optimization problem is developed to detect attacks on networks and to determine optimal control and remediation strategies. By exploiting the structure of these networks, discretized problems that represent

actual municipal networks with up to 300,000 variables could be solved in only a few CPU seconds. More recently, this approach has been extended, by using an MIQP strategy, to a more robust one that promotes unique solutions. Using an efficient pre-processing step, small subproblems can be derived and attack scenarios can be identified more precisely and efficiently. This approach is described in the preprint below.

Finally, the new version of IPOPT continues to be updated and maintained. Designed in C++, this code is object oriented and flexible to deal with a wide variety of linear solvers, new barrier algorithms, and a variety of globalization strategies and decomposition approaches. Moreover, as noted above, the new code provides a much easier interface to modeling environments. Carl has expanded the IPOPT code to accommodate in the following ways:

- Multi-scenario problems are essential to deal with process optimization under uncertainty and have a structure that can be decomposed through Schur complements and can be completely parallelized on distributed memory machines. Carl has obtained some excellent results with this decomposition that validate these properties.
- More recently, Victor has demonstrated this approach on a large-scale parameter estimation problem, executed on parallel processors and show essentially perfect speedups with this approach. A preprint that describes this approach is listed below.
- The structure of IPOPT has been expanded to deal with NLP sensitivity. This approach very quickly provides an estimate of the solution of a perturbed NLP problem. As a result, it is very useful for quickly solving a sequence of problems that change only slightly from the previous one. This is applied below to develop a very fast strategy for Nonlinear Model Predictive Control (NMPC).
- The reduced-space, FORTRAN version of the IPOPT code is also being extended by **Kexin Wang** to include recent developments in the IPOPT project. This will also allow the incorporation of recent penalty-based barrier methods developed by Chen and Goldfarb to be adapted to the reduced space approach.

Mathematical Programs with Equilibrium Constraints (MPECS)

Researchers: **Juan Arrieta (Ph.D. completed April, 2007)**
Brian Baumrucker (Ph.D. started Fall, 2004)
Gaurav Bansal (MS started Fall, 2006)

MPECs represent an exciting new field in mathematical programming. While applications of MPECs have long been recognized in game theory, transportation planning, economics and engineering design, little work has been done in developing efficient algorithms for their solution. In process engineering, these problems stem from bilevel and multilevel optimization problems as well as optimization of hybrid (discrete and continuous) systems. Included in this class are optimization problems with phase equilibrium constraints, as in equilibrium stage processes.

The MPEC problem is essentially a nonlinear programming problem augmented by complementarity conditions. The addition of these conditions leads to a violation of convergence assumptions for most nonlinear programming algorithms (e.g., linear independence of the constraint gradients) and can cause these algorithms to fail. On the other hand, the barrier method described above has a straightforward extension to these problems. As a result, complementarity constraints can be incorporated with little additional cost in the algorithm. In fact, all of the features that contribute to the efficient performance of IPOPT can be invoked directly for the solution of the MPEC problems. The resulting algorithm based on IPOPT has been developed further and refined for a number of challenging dynamic optimization applications in process engineering. Our work in this area includes the following:

This project builds on the Ph.D. work of Arvind Raghunathan. In his research, he has incorporated the MPEC formulation into the IPOPT code, now called IPOPT-C, along with algorithmic modifications to treat the complementarity constraints more efficiently. This code has been interfaced to the AMPL modeling system and has been tested on a set of library MPEC problems. Our results show that this implementation compares well against competing barrier algorithms such as LOQO. Arvind tested this

approach on a number of process optimization problems with phase equilibrium constraints. These include distillation optimization problems with disappearing vapor and liquid phases on equilibrium stages, dynamic systems with switching conditions, such as overflows and phase disappearance in distillation columns and metabolic network systems.

For the latter, we have considered parameter estimation problems for dynamic fermentation systems for wine production. These results show that this approach leads to better and more uniform performance than the smoothing approach used in our previous studies. Related work initiated by **Juan Arrieta** deals with modeling and optimization of dynamic biological systems that are based on cellular models. Such systems are large-scale extensions of MPECs. Moreover, they have interesting modeling features since state dependent constraint functions with discontinuities also need to be treated through MPEC formulations. **Gaurav Bansal** has recently taken over this project and is expanding the capabilities of these formulations along with the consideration of larger, extended models for yeast fermentation. In his MS thesis, he demonstrated the use of the penalty formulation for MPECs on these models.

In extending MPEC optimization formulations, **Brian Baumrucker** is currently considering the development of good MPEC formulations that model discrete decisions. He has explored these with the ROMeo real-time optimization package along with formulations for complementarity conditions within real-time optimization models. In particular, he has shown that penalty-based NLP formulations have worked very well on these models. Moreover, with these formulations, one is able to decouple problem-specific optimization algorithms from the optimization model and rely on general purpose optimization solvers. In addition to the ROMeo comparison described above, he has investigated complementarity formulations in distillation columns with mass transfer limitations.

Simultaneous Optimization of Differential-Algebraic (DAE) Systems

Researchers: **Victor Zavala (Ph.D. started Fall, 2004)**
Rui Huang (Ph.D. started Fall, 2006)
Weijie Lin (Ph.D. started Fall, 2006)
Rodrigo Lopez Negrete de la Fuentes (Ph.D. started Fall, 2007)
Yi-dong Lang (Jiansu Research Institute, Nanjing, China)

The topic on simultaneous solution and optimization strategies for large-scale, dynamic optimization problems incorporates a number of projects. In the past, we have applied this approach to optimization problems for batch reactors, process dynamics and tray-by-tray batch distillation and large-scale dynamic flowsheeting models. Moreover, we have tackled dynamic examples successfully with this approach that cause any sequential dynamic optimization method to fail. A survey of simultaneous solution approaches is described in a reprint below. In particular, this can occur in open-loop unstable systems that arise in optimal control and nonlinear MPC, dynamic parameter estimation and the optimization of runaway reactors. Prof. **Antonio Flores**, a research colleague from the University of Iberoamericana, has demonstrated the advantages of this approach on a number of challenging polymerization processes. Two preprints that describe this work are listed below. Moreover, the following research projects are currently being addressed in our group.

Dynamic Optimization Platforms

We have developed two complementary modeling frameworks for dynamic optimization, one based on reduced space and one on full space. First, over the past four years Yi-dong Lang has developed a stand-alone Windows modeling system for dynamic optimization called *DynoPC*. Written for Windows with a Visual Basic user interface and with heavy integration of graphical, simulation and automatic differentiation tools, this software package incorporates our advances in dynamic optimization, particularly IPOPT and the elemental decomposition. This program continues to be distributed to interested research groups both in academia and industry and can be downloaded from <http://coin-or.org>. Current developments with *DynoPC* include ESO interfaces that are compliant with recently developed CAPE-Open protocols. This leads to an interface compatible with a number of existing process models and modeling environments. In the last year, **Euclides Almeida** worked in our group as a visiting researcher

from Petrobras. He combined the EMSO interface, a modeling interface similar to gPROMS, with DynoPC. The resulting package will be used to solve challenging industrial problems in dynamic real-time optimization (DRTO). A key feature of this approach is the diagnosis of dynamic optimization solutions, along detailed explanation of algorithmic performance, probing and exploration of solutions and an extensive examination facility to deal with elucidation and correction of possible failures of DRTO problems.

Second, we have been using IPOPT in a full space framework linked to the AMPL modeling environment. As detailed in many of the studies listed below, this framework has been extremely powerful and flexible in solving a wide variety of problems. A MATLAB framework has been coupled to the AMPL modeling language and domain specific prototypes have already been developed and work very well. In addition, Johan Akesson, a previous visitor who recently received his PhD from Lund University, has developed a Modelica framework that works with IPOPT and AMPL. In a recent study, he was able to optimize an open-plate reactor with over 100,000 variables in about an hour of CPU time.

Large-scale Parameter Estimation for Polymerization Processes

In a project funded by ExxonMobil Chemicals, **Victor Zavala** has begun to develop multi-stage dynamic optimization problems for grade transition and nonstandard operations. For this, he has developed a detailed reactor model of a polyolefin reactor, which includes a complete kinetic description that is used for parameter estimation using actual process data. Using AMPL and IPOPT, Victor has extended this model to multiple data sets as well as EVM problems (where errors in both inputs and outputs are considered). Applying this approach to a large number of data sets, Victor was able to estimate kinetic parameters quite well; the scatter of the data was reduced by an order of magnitude, compared to previously fitted models. Moreover, using the NLP sensitivity approach described above, Victor has recently shown that the covariance matrix from a maximum likelihood formulation can be extracted easily from IPOPT using a few additional back-solves of the KKT matrix. This allows statistical inference to be added efficiently to parameter estimation studies.

Weijie Lin has recently joined the group and is working on process model development for an advanced polymer product. The approach is similar to the moment models developed by Victor, but with more complex features dealing with the polymer structure. Weijie's preliminary work has shown the importance of parameter estimation for this large-scale model in order to demonstrate structurally dependent features (e.g., gel effect, cage effect, glass effect) on the polymer reactions. This also allows her to develop MW distributions of the complex polymer network.

Nonlinear Model Predictive Control Strategies

In less than two decades, Nonlinear Model Predictive Control (NMPC) has evolved from a conceptual framework to an attractive, general approach for the control of constrained nonlinear processes. These advances were realized both through better understanding of stability and robustness properties as well as improved algorithms for dynamic optimization. The above dynamic optimization algorithms have been applied extensively in Nonlinear Model Predictive Control (NMPC). This strategy is also being extended to large-scale models for polymerization processes.

In addition, we recently adapted the *real-time iteration* approach developed in the context of multiple shooting (Diehl et al. (2006); Bock et al. (2006)) to a collocation-based approach with a full space nonlinear programming solver. Over the past year, we show that straightforward sensitivity calculations from the KKT system also lead to a real-time iteration strategy, with both shifted and non-shifted variants. This leads to an NMPC strategy called the Advanced Step NMPC Controller. Demonstrated on a large-scale polymer process, the Advanced Step Controller leads to on-line calculation effort that is reduced by over two orders of magnitude. As described in a preprint below, we have been developed a general Lyapunov-type stability analysis for this approach that demonstrates nominal stability, input to state stability and robustness margins to noise and model mismatch. This recent development has been extended to moving horizon estimation (MHE) for parameter and state identification and yields the same dramatic improvements in performance. This result was recently presented in a keynote paper at the DyCOPS meeting and is listed as a reprint below. Finally, **Rodrigo Lopez** has recently joined the group and is

extending this approach to stochastic processes with uncertainties in states and outputs. Also, **Rui Huang**, has recently joined the group to work on NMPC distillation control and advanced power plant systems. His preliminary work has demonstrated that the Advanced Step NMPC also reduces on-line calculation time by about two orders of magnitude.

Large-Scale Optimization for Fuel Cell Models

Researchers: Parag Jain (Ph.D. started Fall, 2005, joint with Prof. M. S. Jhon)

In his thesis project, Cong Xu initiated the investigation of optimization problems arising in fuel cell applications, including Direct Methanol Fuel Cells (DMFCs) and Hydrogen PEM fuel cells. In particular, he has extended simple DAE models for fuel cells to consider optimal flow patterns and minimization of contamination effects due to membrane crossover, and he has extended these models to consider sensitivity and optimization calculations. In particular for DMFCs, we have compared and critically assessed several literature models and have explored the crossover phenomenon. By applying our dynamic simulation and optimization tools, they were able to develop dynamic feeding strategies for these systems that greatly reduce contamination due to methanol crossover. The results show that the optimal design of DMFCs can be achieved without pulsed operation. Here crossover is significantly reduced and the power density is correspondingly increased. The efficiency of these processes can also be improved through flowsheet optimization strategies as well as heat integration of the process units. This approach is based on conventional flowsheet optimization schemes within Aspen/Plus along with simultaneous optimization and heat integration strategies. The application to fuel cell based processes leads to interesting trade-offs between the fuel cell, the rest of the flowsheet and the heat recovery system.

Dynamic optimization has also been applied by Shiva Kameswaran to develop ramping strategies for fuel cell power plants. In a project funded by Fuel Cell Energy, Inc. and DOE, Shiva has demonstrated significant improvements in maintaining power levels with more energy efficient operation. Using AMPL and IPOPT, he has incorporated detailed models of the entire power plant and has demonstrated this approach both for parameter estimation and optimal control. Finally, **Parag Jain** has extended previous modeling and optimization work for Hydrogen PEM Fuel Cells. He has applied a comprehensive two-dimensional computational model of a PEM fuel cell that accounts for major transport processes including membrane electrode assembly (MEA) and gas distribution channels. Additionally a water transport model in the membrane has been developed and incorporated into the full PEM fuel cell model. Using AMPL and IPOPT, Parag is using this model to develop a detailed optimization case study. The results of this work are listed in a preprint below.

Dynamic Optimization for Semi-Closed Process Models

Student: Sree Ram Raju Vetukuri (PhD started Fall, 2006)

Collaborator: Prof. Andrea Walther (Technical University of Dresden)

This project explores the application of recent developments in automatic differentiation and nonlinear programming to semi-closed process models, including systems solved with DAE integrators. This sequential optimization approach for dynamic systems has been the competing strategy to the simultaneous optimization approach adopted in our group. It has advantages for problems where error control is especially important, where many time steps are needed and where these need to be adjusted frequently during the integration. Recent years have seen the development of efficient sensitivity strategies to obtain accurate first and even second derivatives for use in dynamic optimization algorithms. However, this approach becomes inefficient when there are many degrees of freedom. For instance, in the optimization of Pressure Swing Adsorption units, Ling Jiang developed an efficient sequential dynamic optimization approach. However, when several hundred inputs were considered in the model, the optimization required hundreds of CPU hours.

This bottleneck has recently been addressed by a recent SQP-type algorithm that uses exact right hand sides of the KKT equations, along with Jacobian-vector and Hessian-vector products, but computes only inexact Jacobians. Using leading edge technology in automatic differentiation, these quantities can be computed orders of magnitude faster than with the standard sequential approach. Profs. Andreas Griewank and

Andrea Walther have recently developed a globally convergent SQP that uses inexact Jacobians, called TR1. Based on a composite trust region strategy, this approach has recently been applied to periodic adsorption processes as well. **Sreeram Vetukuri** has recently joined the group and has extended this approach using ADOL-C for automatic differentiation and CVODES for sensitivity calculations of first and second derivatives, and interfacing to TR1. Preliminary results on a small SMB system show the potential of this algorithm for reducing CPU time.

Reduced Order Models for PDE-based Units in Power Plant Flowsheets

Researcher: **Yi-dong Lang (Jiansu Research Institute, Nanjing, China)**

Students: **Anshul Agarwal (Ph.D. started Fall, 2005)**

Adam Malacina (Ph.D. started Fall, 2006)

Process simulation has become an essential capability that spans most tasks in process engineering, design and analysis. Simulation programs such as Aspen Plus, Pro/II and HYSYS are integrated into core engineering design activities. However, due to computational limitations, current process simulators usually deal with lumped parameter models that often do not capture detailed spatial characteristics of process units, including fluid flow and transport processes. As a result, they can suffer accuracy limitations and the loss of predictive capability. To overcome this difficulty, we are working with NETL on the Advanced Process Engineering Co-Simulation (APECS) project. Within this framework, process flowsheeting models (in Aspen Plus) have been successfully combined with computational fluid dynamics (CFD) models (using Fluent) to model multiphase systems, such as gasifiers and turbines within power plants. Current efforts are also focused on the inclusion of other distributed parameter custom models. In particular, an important, and challenging, application for this activity is the FutureGen Cycle, a 275 MW coal-fired power plant that includes hydrogen generation and CO₂ recovery, and requires multi-phase, spatially distributed models within its simulation environment.

To extend these NETL activities, we are developing and applying integrated optimization strategies that provide powerful tools for model development and process improvement. So far, these optimization problems are small-scale and efficient, and robust optimization algorithms are available that can be implemented easily. These features have led to the relatively easy and widespread adoption of optimization strategies. On the other hand, optimization strategies for CFD and distributed parameter models have seen far less development and this currently represents an important and challenging research activity. For APECS, which incorporates these models, we believe that optimization strategies will perform a key role in the development of advanced energy processes, such as the FutureGen Cycle. To provide a vehicle for the research tasks in this project, we consider the model components of the FutureGen Cycle flowsheet. This process includes a number of unit operations for coal gasification and byproduct (e.g., ash/slag and sulfur) separation; multiphase CFD models are needed for this purpose. Yi-dong Lang has been working with colleagues at Fluent Corp. to develop reduced order CFD models based on Proper Orthogonal Decomposition (POD) and Principal Components Analysis (PCA). This approach allows for a tremendous reduction in computing effort for the process flowsheet. Yi-dong has considered a PCA based approach coupled with an efficient experimental design approach along with a neural network solver to determine an input output model. Using an NETL-based Fluent combustor model, he has developed accurate reduced order models (ROMs) that capture the behavior of the Fluent model over a wide input range, but with require about two orders of magnitude less execution time. As a result, these models are quite suitable to incorporate within FutureGen process flowsheet. **Adam Malacina** has recently joined the group and is currently extending Yi-dong's approach to a FutureGen gasifier model, also developed in Fluent.

Further downstream of the gasifier in the FutureGen process, the byproduct gases CO₂ and hydrogen are separated and CO₂ is sequestered. These separations can be performed effectively by pressure swing adsorption (PSA) units, which operate dynamically and require the calculation (and subsequent optimization) of cyclic steady state behavior. Such models were considered in the PhD thesis of Ling Jiang. While her optimization approach was very successful for large-scale PSA units, it still required many hours of CPU time, and therefore requires custom reduced order models (ROMs). Over the past year, **Anshul Agarwal** has been developing reduced order models for these units as well using a novel space-time discretization to develop the ROM using a POD-based approach. A key advantage of this approach is that very small models are developed that can be discretized in time (using collocation on finite elements) and

solved with AMPL and IPOPT. Preliminary results by Anshul show that this strategy can lead to very efficient optimization strategies. This approach is described in a preprint listed below.

Optimization and Control of Periodic Adsorption Processes

Student: Yoshi Kawajiri (Ph.D. completed July, 2007)

Simulated Moving Bed (SMB) chromatography was developed in the 1960s as a realization of continuous chromatographic separation. Since then, SMB has been gaining attention in food, sugar, and petrochemical applications. In more recent years, SMB has been recognized as an attractive technology for enantiomeric separations in the pharmaceutical industry. An SMB system is realized by connecting multiple columns in a cycle. The feed and desorbent are supplied continuously between columns, and at the same time extract and raffinate products are withdrawn. Counter-current operation is “simulated” by switching these inlet and outlet streams intermittently in the direction of the liquid flow. As with PSA this system is characterized by a Cyclic Steady State (CSS), where the concentration profiles in the columns change dynamically but repeat the same pattern in every cycle.

Yoshi Kawajiri recently completed his PhD. Prior to starting a faculty position at Georgia Tech, he is a Humboldt Fellow at the Max Planck Institute in Magdeburg, Germany. In his research, he has been investigating efficient optimization strategies for SMB systems. Here optimization methods can be based on the single-discretization approach to discretize the Partial Differential Equations (PDEs) only in the spatial domain and then integrate the resulting Differential Algebraic Equations (DAEs) in time. On the other hand, the full-discretization approach discretizes the PDE both in time and spatial domain, leading to a large-scale Nonlinear Programming (NLP) problem. Yoshi has compared both approaches and found that the latter approach has shown tremendous improvements in the productivity of SMB systems. Moreover, these were obtained almost two orders of magnitude faster than with the single discretization (sequential) method used in gProms. The approach was applied to optimization of SMB systems with steady and time-varying feeds (i.e., PowerFeed), both for linear and nonlinear isotherms. This efficient approach also allows us to consider the design and control much more complex SMB systems. The results of Yoshi’s approach have been detailed in a number of papers included with previous newsletters. In particular, Yoshi has developed two superstructure SMB systems, one for a single step and the second over a full cycle, that allow for the optimization of a number of novel, recently reported designs, as well as some new ones. These superstructures lead to well-defined and easily solved problem formulations and lead to innovative designs with improvements in productivity. Moreover, through the formulation and solution of multi-objective optimization problems, we confirm that our approach has the potential to find more advantageous operating schemes than the standard SMB, PowerFeed, or other advanced, but fixed, operating schemes. These results are described in two reprints and a preprint listed below.

Ignacio Grossmann’s Group

Open-Source Code for MINLP Optimization

New developments: Open source code *Couenne*

Post-doctoral fellow: Pietro Belotti (Tepper)

Collaborators: Larry Biegler, Francois Margot (Tepper)

This project has been a collaboration with researchers at IBM (Andreas Wächter, Joantahn Lee, Pierre Bonami). The main objective has been to take the next step beyond the open source code *bonmin* for MINLP problems (see <http://egon.cheme.cmu.edu/ibm/page.htm>). In particular the goal has been to develop a global optimization code for solving MINLP problems.

In this project Pietro Belotti is developing the *Couenne* open-source code for solving nonconvex NLP and MINLP problems. The goal is to develop both approximate as well as rigorous methods that rely on the use of convex envelopes and bound tightening strategies. The rigorous method relies on a spatial branch-and-bound method for bilinear, linear fractional and concave separable functions. Efforts for developing code

generating parametric linear upper and lower convex envelopes are under way. Parametric linear convex envelopes are obtained by breaking a nonlinear expression into compositions of simple operators for which convex linearizations are known. By introducing additional variables, a linear convex envelope of the original expression can be obtained. This polyhedral envelope is a function of the upper and lower bounds on the variables and it becomes a tighter approximation when upper and lower bounds on the variables are moved closer. This can be exploited in a branch-and-cut algorithm by branching on continuous variables, i.e. splitting the domain of its possible values. These parametric convex envelopes can also be used to generate an infinite number of valid linear cuts, raising interesting questions regarding cut selection and cut strategy when integrated in a branch-and-cut framework. These ideas are being implemented in the software package *Couenne* with input from Leo Liberti at the Ecole Polytechnique in Paris. Pietro has written the code *CouenneSolverInterface* that allows the implementation of various bound tightening techniques, including a recent one for quadratic expressions. Pietro has also been exploring the idea of an “aggressive” bound tightening procedure in which a three-way branching is performed by considering a smaller interval around the relaxed NLP solution. This approach has produced reasonable reductions in CPU time on several test problems, which are still relatively small. Pietro is currently exploring the use of this aggressive tightening for quickly generating a feasible integer point for a nonconvex MINLP.

Algorithms for Nonlinear Disjunctive Programming

New Developments: **Criteria for intersecting disjunctions for bilinear GDP problems**

Student: **Juan Pablo Ruiz [started Jan 2007]**

Research collaborator: **Aldo Vecchietti [Researcher at INGAR]**

Juan Ruiz

The project of Juan is concerned with the global optimization of Generalized Disjunctive Programs that involve bilinearities in the constraints. These problems arise for instance in the design of pooling problems, in the synthesis of integrated water treatment networks, or generally, in the synthesis of process networks with multicomponent flows. Juan’s work has as a major objective to improve the computational efficiency of disjunctive spatial branch and bound methods by making use of the recent theoretical developments of Nick Sawaya to obtain tighter relaxations. The basic approach consists of first relaxing the bilinear terms in the GDP using the McCormick convex envelopes, and introducing them as part as the disjunctive set. Since the corresponding relaxation leads to a linear GDP problem, the next step then consists in performing basic steps which involve intersecting disjunctions in order to obtain tighter relaxations as outlined in Balas’ theory. Of course there are many options that are possible on how to perform the intersections. To address this issue Juan has proved a sufficient condition for a Theorem by Balas that states some conditions under which basic steps are not necessary. The conditions in this theorem are difficult to verify since they establish that no new extreme points are created as part of the basic steps. The sufficient condition that Juan has developed states that this Theorem holds true if two-term disjunctions involve bounded variables that partition for each term. Using this proposition as a basis as well as other properties, he has developed a procedure that relies on the following rules: 1. Basic steps are applied between disjunctions with at least one variable in common (The more variables in common the more tightening can be expected); 2. If bilinearities are outside the disjunctions, apply basic steps by introducing them in the disjunctions (if bilinearities are inside disjunctions less tightening can be expected); 3. Basic steps between improper disjunctions and proper disjunctions do not add new variables to the formulations. The solution method then consists in first applying this reformulation procedure, then applying the bound contraction method by Juan Zamor, and finally applying a branch and bound procedure similar to the one that Sangbum Lee had developed in which branching is performed first on the discrete variables.

Following the method described above, Juan has applied his reformulation procedure to two examples, In the case of a water treatment network the bilinearities arise in equations outside the disjunctions, so that the intersection of disjunctions with the convex envelopes leads to potentially stronger relaxations. This was confirmed in an instance where Juan managed to improve the lower bound by 25% which in turn translated in a 50% reduction in the spatial branch and bound search (399 nodes to 204 nodes). In the case of a

pooling network, where the bilinearities arise within the disjunctions the strength of relaxations is expected to be lower. This was again confirmed with an instance in which the lower bound was improved by only 1% and the number of nodes was reduced from 748 down to 683. Juan presented this work at the AIChE Meeting in Salt Lake City.

Aldo Vecchietti: LOGMIP and DICOPT

Aldo and his students at INGAR in Argentina have been developing the LogMIP code, which is as an extension of the GAMS modeling language for posing logic/disjunctive problems. The syntax developed involves the use of IF THEN ELSE statements to represent the disjunctions, including embedded ones. The definition of disjunctions over sets and subsets can also be specified. As for the propositional logic, the special constructs ATLEAST, ATMOST, EXACTLY are defined to specify that the sum over a subset of variables with a given cardinality is at least, at most or equal to one. Also, LogMIP can now accept logic propositions in symbolic form without the need of translating them into linear inequalities.

For solving linear disjunctive/hybrid problems LOGMIP can automatically generate the convex hull relaxation and big-M transformation for linear problems. Aldo has also automated the logic-based OA algorithm by Metin Turkay, which is suitable for nonlinear GDP process network problems. In collaboration with GAMS, Aldo has been developing a version that will be commercialized later in the fall. In the meantime LogMIP Website provides the possibility of downloading the code: <http://www.ceride.gov.ar/logmip/>.

Global Optimization of Integrated Process Water Systems

New developments: **Publication on multiscenario synthesis**

Students: **Ramkumar Karupiah [Ph.D. started Jan 04; completed September, 2007]**

Ram defended on September 13 and has joined The Dow Chemical Company. His project has dealt with the global optimization for the synthesis of process water systems that integrate the water reuse section for processes with the distributed water treatment section for a given number of pollutants. Ram assumed for the processes units that they are either characterized by fixed contaminant loads, or by variable loads that obey mass exchange constraints in terms of concentrations. For the treatment units fixed recoveries are specified. Ram developed a superstructure that involves for a given number of process units and treatment units all possible alternatives for interconnection of process and treatment units, and mixing and splitting of streams. The model corresponds to an NLP model with bilinearities in the mixers given by the product of flow times concentrations in ppm of the contaminants. Ram developed a global optimization algorithm that relies on combining spatial branch and bound search, with piecewise linear approximations. To strengthen the quality of the lower bound, Ram derived a valid cut that represents overall mass balances for each of the contaminants. This cut proved to be extremely effective to a point where adding it to the original formulation and using BARON requires similar times as the special purpose spatial branch and bound algorithm.

As a next step in his research, Ram considered the extension of the above problem to the case when the water system must operate under multiple scenarios in which the loads in the units as well as the recoveries in the treatment units are uncertain, and therefore change in each scenario. This problem gives rise to a two-stage stochastic programming MINLP problem. The first stage costs include the investment cost for piping which depends on the maximum flowrate allowable in a pipe, and the design cost of each treatment unit, which is dependent on the maximum flow of wastewater handled by that treatment unit. The operating costs of the network appear in the second stage, which include the cost of obtaining freshwater for use in the process units, the cost of pumping a certain flow of water through the pipes and the operating costs in the treatment units. The difficulty of the global optimization for the nonconvex multiscenario problem is that the MINLP becomes much larger, and 0-1 variables must be introduced for the piping in order to control the potential complexity in the configurations for each period. Ram developed a solution method

that is based on a branch and cut algorithm. The basic idea consists in performing a spatial branch and bound where cuts are generated at each node using a Lagrangean decomposition scheme. These cuts are obtained by globally optimizing each scenario independently, and are added to the original problem that is convexified by constructing convex envelopes for the non-convex nonlinear terms leading to an MILP that predicts a rigorous lower bound. A heuristic is used for the generation of good upper bounds. These lower and upper bounds are converged within a specified tolerance in a spatial branch and bound algorithm. Ram considered a problem with 2 process units and 2 treatment units and 10 scenarios, and another one with 5 process units, 3 treatment units and 3 scenarios. These problems were solved in only few minutes within 1% of global optimality.

Ram was able to generalize the above cited branch and cut algorithm for generic multiscenario problems, which was submitted for publication and recently accepted in JOGO. The paper contains rigorous proofs on the validity and strength of the lower bounds generated by the proposed method.

Energy Optimization of Corn-based Ethanol Plants

New developments: **Manuscript submitted on 65% reduction in steam consumption**

Students: **Ramkumar Karupiah [Ph.D. started Jan 04; completed September, 2007]**

Visitor: **Mariano Martin (Salamanca), Andreas Peschel (RWTH Aachen)**

This project was initiated with the exchange student Andreas Peschel and Ramkumar Karupiah in collaboration with Cargill (Wade Martinson and Luca Zullo). Mariano Martin continued the work with the help of Ramkumar. The objective was two-fold: a) Develop short-cut models for corn-based ethanol plants; (b) Develop a superstructure optimization model to determine the extent to which the design of these plants can be improved.

In the first phase of this work, a simplified model was developed for the “dry-grind” process for the corn-based bio-ethanol plant. In such plants, fuel ethanol is produced using corn-kernels as the feedstock. Fuel grade ethanol has to be 100 % pure before it can be blended with gasoline to be used in automobiles. However, conventional distillation columns produce an azeotropic mixture of ethanol and water (95 % ethanol – 5 % water), which has to be purified further for making fuel ethanol. The main challenge in the way of producing fuel ethanol commercially is that the process is very energy intensive and requires large amounts of steam and the use of expensive molecular sieves to get 100 % pure ethanol. Furthermore, the waste spillage from the fermentation units, which is rich in volatile organic compounds, and the other wastewater streams in the plant have to be treated before disposal, which, in turn requires energy intensive units like centrifuges and dryers.

In order to optimize the design of a bio-ethanol plant, Andreas, and Ram developed a model to predict the performance of the flowsheet that includes grinding, saccarification, fermentation, centrifugation and drying operations. A superstructure was also postulated in which some of the major alternatives include separation by molecular sieves and or corn grits, different ways to accomplish the drying for the dried grains solids, the cattle feed by-product, use of multi-effect columns and heat integration. The objective was to minimize the energy requirement of the overall plant while trying to maximize the yields. The optimization produced a 65% reduction in steam consumption, which lead to a decrease of manufacturing cost from \$1.50/gal to \$1.28/gal. This was largely achieved by using multi-effect distillation in the “beer” column and in the azeotropic column, as well as heat integration. The manuscript describing this work has been submitted for publication and is available to CAPD members.

Optimal Synthesis of Integrated Gasification Combined Cycle (IGCC) Systems

New developments: **Model for air separation unit**

Student: **Ravi Kamath [Ph.D. started Jan 2007]**

Ravi successfully passed his Ph.D. qualifying exam. This is a new project within the newly established Institute of Energy Solutions that is being funded through NETL. The objective is to develop a comprehensive synthesis model for IGCC plants, with the possibility of polygeneration and CO₂ sequestration.

Integrated Gasification Combined Cycle (IGCC) technology for future power plants has attracted great interest because of its capacity to not only achieve a higher thermal efficiency but also capture CO₂ more economically than a conventional coal-fired plant. However, IGCC technology has higher capital costs than coal-fired plants and is economically viable only if operated with CO₂ capture. Ravi's project will aim at evaluating the techno-economic performance of an IGCC plant. We assume that we are given a type (rank, quality, composition, physical state) of coal, net output power, location of site, ambient conditions for utilities like air and water and other requirements like co-production of hydrogen or chemicals and extent of carbon capture. The objective is to determine the optimal structural configuration and operating parameters of the IGCC plant that minimize the investment and operating cost, while meeting the constraints for environmental emissions. An important related decision will be the minimum amount of CO₂ to be captured.

As a first step, Ravi developed a comprehensive superstructure that includes potential alternative technologies. The goal is to determine the optimal configuration and operating parameters for the IGCC flowsheet by formulating a mixed-integer nonlinear programming (MINLP) problem. Because of the large number of alternatives, we intend to apply the methodology developed by Mark Daichendt that involves optimizing the entire flowsheet at different levels of complexity using a combination of aggregated and detailed models. As indicated in the last newsletter, Ravi developed simplified models for coal gasification and for the utility section. For coal gasification he developed a model based on Gibbs energy minimization that can be modified to match the performance of real gasifiers. This is done by adding equilibrium reactions as constraints, or additional specifications to match outlet compositions. He has obtained very good results for the Shell, GE and ConocoPhillips entrained gasifiers. The prediction of the main eight species is in close agreement with published data. Ravi also developed an MINLP model for combined cycles that extends the MINLP model by Bruno. The basic idea consists of postulating a superstructure for the specified power and heating demands. The superstructure is modeled as an MINLP in which pressures and temperatures of steam, enthalpies, and turbine efficiencies are treated with nonlinear models. For an example with an electricity demand of 500 MW, 2 mechanical power demands, and steam demands at high, medium and low pressure, Ravi was able to obtain good results from the MINLP model that involved 44 0-1 variables, 1275 continuous variables and 1309 constraints.

In the last few months Ravi has concentrated on the development of the flowsheet and optimization model for the Air Separation Unit (ASU), which supplies high pressure oxygen (> 95% purity) to the gasifier and medium pressure nitrogen to the gas turbine. A real industrial flowsheet for air separation involves 2 complex columns which are thermally integrated in such a way that heat load of the condenser of high pressure (HP) column matches that of the reboiler of the low pressure column. Also included in the flowsheet are multiple feeds and side-draws that are heat-integrated using two multi-stream heat exchangers. Ravi has developed a model based on Kremser's equations and modifications by Edmister in order to be able to simulate and optimize this system in equation form in GAMS. Ravi has developed approximate models for counter-current cascades whose performance models closely resembles that of rigorous models of Aspen Plus. The step-wise construction of the flowsheet can help in understanding why such a complicated heat integration involving multiple streams is preferred over a simple configuration by evaluating material and energy costs at each step in the flowsheet construction. The flowsheet model of Ravi can be solved in GAMS as a simulation problem or also as an optimization problem. The current task of Ravi is to develop models for the multi-stream heat exchangers using the pinch location method (Duran and Grossmann, 1986) which does not require the definition of temperature intervals. Ravi also plans to add extra elements to the flowsheet such as side-streams and additional feeds to the HP and/or the LP columns and accordingly modify the models for the complex columns..

MINLP Flowsheet Optimization with Process Simulators

New developments: **Submission of paper for flowsheet optimization using kriging model**

Collaborators: **Jose Caballero [Associate Prof., Univ. Alicante]**

While most of the work with Jose Caballero has concentrated on thermally integrated separation sequences, one additional piece of work has been the incorporation of process simulators for MINLP optimization in process flowsheets. Beyond the work reported in the last newsletter for solving Process Synthesis problems with GDP and MINLP models, Jose has concentrated his efforts in the NLP optimization of flowsheets implicit models.

Jose's work has addressed the rigorous optimization of nonlinear programming problems in which the objective function and (or) some constraints are represented by noisy implicit black box functions. The special application considered is the optimization of modular process simulators in which the derivatives are not available and some units operations introduce noise preventing the calculation of accurate derivatives. The basic idea in the proposed approach relies on substituting the black box modules by metamodels based on a kriging interpolation that assumes that the errors are not independent but a function of the independent variables. A kriging metamodel uses a non Euclidean measure of distance that avoid sensitivity to the units of measure. It includes adjustable parameters that weight the importance of each variable getting a good model representation, and it allows to calculate errors that can be used to establish stopping criteria and provide a solid base to deal with 'possible infeasibility' due to inaccuracies in the metamodel representation of objective function and constraints. Jose has proposed a detailed algorithm that continues with refining stages and successive bound contractions in the domain of independent variables with or without kriging recalibration until an acceptable accuracy in the kriging metamodel is obtained. While in general it is not possible to guarantee convergence to a local optimum of the rigorous nonlinear model, Jose has identified sufficient conditions that rely on a trust region approach, which however is computationally only feasible for low dimensionalities (up to 5 decision variables). Jose has successfully applied his method to a small analytical example, two distillation problems that make use of the simulator HYSYS and optimize the vapor flows, and a flowsheet for producing phthalic anhydride from o-xylene where a plug flow reactor is replaced by a kriging model. In all cases encouraging results were obtained. The paper describing this work is listed in the section on references.

Synthesis of Crystallization Processes

New developments: **Completing manuscript**

Post-doctoral fellow: **Ricardo Lima (started July 2006)**

Ricardo has been studying the development of superstructures for the optimization of the separation of p-xylene from a mixture of mixed xylenes using crystallization. Ricardo has proposed a superstructure that considers feasible flowsheet alternatives for the all process configurations. This superstructure is a generalization of the flowsheets studied by Carlos Mendez et al. (2005) and includes alternatives for feed compositions of p-xylene ranging from 21% to 98%. Ricardo has embedded all the flowsheets studied by Carlos and considered new connections between stages. The superstructure includes a sub-superstructure for the first crystallization stage and a new sub-superstructure for the set of centrifuges with wash, which allowed establishing new connections between the different process stages so as to consider flowsheets that have been proposed in US patents. In the MINLP model the integer variables are associated with equipment and with disjunctions related with solubility predictions. To solve this problem Ricardo has developed a two-level decomposition consisting of the iterative solution of an aggregated and a detailed model. The two key ideas in the aggregated model are: a) merging the units in centrifuge blocks and slurry drums into single input-output blocks so that the aggregated model is defined in the space of interconnection of major blocks; b) elimination of the constraints that set an upper bound on the inlet flowrate of each centrifuge unit. In this way a large number of equations and variables are eliminated because groups of individual units are replaced by a single equivalent unit. However, in order to meet the same production targets, the

constraints that impose operating ranges for the centrifuges are relaxed in the aggregated model. For the definition of the detailed model, the number of units in each aggregated block is calculated by the ceiling of the ratio between the calculated capacity of the block and the upper bound of the size for each unit. At each iteration of the two-level decomposition an integer cut is added to the aggregated model to eliminate previous combinations of number of units. This requires the introduction of binary variables in the aggregated model in order to calculate the number of units associated with each aggregated block. The solution of the aggregated model is used to initialize the detailed model as well as to define a reduced superstructure for the detailed model. While the proposed approach has the advantage of providing an effective solution method, its limitation is that convergence cannot be guaranteed in terms of lower and upper bounds. Therefore, the approach we take is to simply iterate over a fixed number of major iterations.

The results obtained show that the aggregated MINLP model is easier to initialize and to solve than the MINLP model of the overall superstructure. Furthermore, the solution from the aggregated model provides a good initial point for the detailed model. The optimum flowsheet that was obtained yields a new design alternative compared to previous flowsheets studied by Carlos Mendez. Comparing the configuration of the optimum flowsheet obtained by these authors with the one obtained in this work, the latter has three fewer slurry drums, one less melting stage, four fewer centrifuges, but one more crystallizer. This led to a reduction of 12% in the total annualized investment cost but at the expense of a small increase in the total annual operating cost, resulting in a modest net improvement in the total annual profit. Ricardo is completing the manuscript on this work.

Design for Sustainability

New development: Completing two manuscripts

Post-doctoral fellow: Gonzalo Guillen

Gonzalo, a postdoc with a Fulbright fellowship, has tried to incorporate sustainability considerations in the synthesis and design of chemical processes. The aim of this project is to develop a set of quantitative tools based on mixed-integer modeling techniques that will facilitate the adoption of more sustainable alternatives for a wide range of problems arising in PSE.

Gonzalo has developed a holistic framework for the design of sustainable chemical supply chains that takes into account the uncertainty associated with the impact assessment model. The environmental impact is measured through the Eco-Indicator 99, which reflects the advances in the damaged oriented method recently developed for Life Cycle Impact Assessment. This strategy covers the entire life cycle of the product, process or activity, including extracting and processing of raw materials; manufacturing, transportation and distribution; reuse and maintenance; recycling and final disposal. The strategy proposed by Gonzalo explicitly considers the uncertainty associated with the damage model based on the Life Cycle Assessment principles. Specifically, the variability of the Eco-Indicator 99 under the uncertain environment is controlled by reducing the probability of exceeding a specified target level. Gonzalo formulated this problem as a bi-criterion mixed integer linear program. The objectives are maximizing the net present value (NPV) and minimizing the probability of an Eco-Indicator 99 value below a given target. The inclusion of the latter criteria gives rise to a chance-constraint whose deterministic equivalent is obtained by applying concepts from chance-constrained programming. The resulting multi-objective model is reformulated as a parametric MILP. The probability associated with the Eco-Indicator 99 is treated as the main objective whereas the NPV is constrained to be greater than an epsilon value. By parametrically changing this epsilon value, the entire set of Pareto optimal solutions of the problem can be generated. The parametric MILP is solved by decomposing it into two sub-problems and iterating between them. The first sub-problem is a parametric LP obtained by fixing all the binary variables of the original formulation. This problem is solved by performing a sensitivity analysis with respect to the target imposed to the NPV. This yields a parametric profile that is intersected with the current best one providing an approximation to the Pareto curve of the original problem. The second sub-problem is a MILP where the epsilon parameter is relaxed as a variable. This MILP includes parametric cuts and logic cuts that are derived from previous explored solutions. The

parametric cuts force the model to seek solutions improving the best current parametric profile in at least one point. The logic cuts are added to exclude the solutions found in previous iterations.

Using the proposed model and solution procedure Gonzalo has solved several case studies related to a supply chain in Tarragona, Spain, with possible expansions and distributions in Central Europe. The results show how environmentally friendlier solutions in the face of uncertainty in the damage model can be attained by systematically trading-off the economic benefit of the process. These robust solutions are achieved by structural modifications in the SC and also by properly adjusting the production rates and transportation flows between the SC nodes. Furthermore, the proposed decomposition strategy is able to provide the whole set of Pareto optimal solutions in a fraction of the CPU time required in the standard epsilon-constraint method. Gonzalo has extended this technique to allow the control of the individual impact measures included in the Eco-indicator 99. He is completing two manuscripts, which should be available in the next newsletter.

Design and Planning of Deep-Water Oilfield Development under Uncertainty

New Developments: **Incorporation of nonlinear production profile**

Students: **Bora Tarhan (Ph.D. started January 2005)**

The project of Bora Tarhan deals with the design and planning of gas and oil fields under uncertainty.

As a first step Bora addressed the following stochastic optimization problem that in principle is simpler to model. A network of candidate processes is given over a specified time horizon that is described by multiple time periods in which product demands are specified. It is assumed that major uncertainties are involved in the yields of each process, which are described by discrete probability distribution functions. The uncertainty can be reduced with the potential investment in pilot plants, which delays the introduction of a process but provides more accurate process information on the yields. In addition, once a process is installed, the uncertainties in the yields decrease with time to reflect the experience and improved understanding in the operation of the process. The problem then consists in making planning decisions for process selection, capacity expansions, and possible investment in pilot plants in order to maximize the expected net present value over the specified time horizon. In order to capture all the complex trade-offs, Bora developed a new mixed-integer/disjunctive programming model that is composed of scenario linking and non-anticipativity constraints that are similar in spirit to Vikas Goel's model. One importance difference is the time varying uncertainty, which for simplicity Bora considered is revealed over two time periods. Furthermore, he assumed that the time required for a pilot plant study corresponds to one time period, meaning that uncertainty is reduced to only one time period. In order to solve this problem Bora developed a dual Lagrangean branch and bound method. The corresponding subproblems are obtained by relaxing the disjunctions and transferring the first period non-anticipativity constraints to the objective function with Lagrange multipliers. The resulting model can be rewritten as independent subproblems for each scenario. The overall objective is to find the minimum upper bound by updating the multipliers. The branch and bound involves branching over both discrete and continuous variables that are involved in the non-anticipativity constraints. Bora has successfully applied this method to a 3 process network, with one existing process and two with new technologies and for which investment of pilot plants can be considered over a 10 year horizon. Reformulating this problem as a single MILP involves 7,360 0-1 variables, 8,841 continuous variables and 85,139 constraints. After one week CPLEX obtained a net present value of \$61.5 million. With the proposed method a solution of \$80.14 was found in 10 hours of CPU time with a gap of 3%. The proposed solution did not select pilot plants to reduce the uncertainty, and proposed expanding Process I up to a capacity of 10 tons/day and making an additional expansion of 4.49 tons/day at time period 1 if the yield turns out to be 69%. If the yield for Process I is found to be 81% then an expansion of 2.98 tons/day is made also at the time period 1.

For the deep-water oilfield development problem, Bora has considered an oil field consisting of a number of reservoirs where each contains several possible well sites. Some of these well sites have to be drilled and exploited for oil over a planning horizon. The oil field infrastructure can be composed of Floating

Production Storage and Offloading (FPSO) and/or Tension Leg Platform (TLP) facilities. The FPSO facility can be either a small FPSO, converted from a retiring oil tanker, or a large FPSO, newly constructed grassroots facilities. An FPSO facility can produce, store and offload the produced oil to other tankers. Unlike FPSO, a TLP facility cannot produce oil; it possesses only drilling and oil recovering capability. TLP and FPSO facilities can be connected to each other through pipes. There are two options for drilling wells. Each well can be drilled as a sub-sea or a TLP well. Drilling ships are used to drill sub-sea wells, so there is no need to have an FPSO or a TLP facility present to drill a sub-sea well. A sub-sea well has to be connected to an FPSO facility, whereas a TLP well has to be connected to a TLP facility. The problem consists of selecting investment and operation decisions such as selection of the type, location of facilities, time to build them, selection of wells to drill, the time to drill each well, time to connect each well to the facilities and production from the wells. There are uncertainties in the sand quality, size of the reservoir and breakthrough time. Given the above assumptions, the goal is to maximize the expected net present value of the project. A simplified reservoir model has been considered where the oil rate decreases linearly, and the water-to-oil ratio is follows a nonlinear function with the cumulative oil production.

Bora has developed an MINLP model and solution method that he has largely completed in his second internship at ExxonMobil last summer. He developed a model in which facilities and wells of the original problem are aggregated. Instead of deciding which well to drill or which facility to build, the decisions are how many wells or facilities to build. Bora initially modeled the production profile (deliverability vs. cumulative production) as a linear function. In the last few months he has incorporated a more realistic nonlinear model that in principle one might fit with the results of a detailed reservoir simulation model. To account for the fact that uncertainties are not revealed immediately, Bora incorporated a number of rules that specify when the uncertainty is revealed, either in terms of number of items (e.g. number of wells) or in terms of time of production. Because of the aggregation, variables for connection of facilities and wells are disregarded, and lengths to wells are underestimated in order to guarantee that the model predict an upper bound on the detailed model. This model is similar to the process networks model with the exception that each scenario is a non-convex MINLP. The steps of the algorithm are the same, except that each subproblem is solved using BARON as the global optimization algorithm, although Bora is also testing the use of DICOPT as it is much faster but without guarantee of global optimality. The relaxation yields an upper bound and the lower bound is generated by a simple heuristic. At this point Bora has implemented the method using a special scheme for subgradient optimization to update the Lagrange multipliers. He has recently considered the case of one reservoir with linear production profile over a 10 year period. The best feasible solution is obtained on an example had an expected NPV of $\$6.5 \times 10^9$. In the optimal solution, the model predicts start building two small FPSO facilities and drill 9 subsea wells. Depending on the future outcomes the investments range from building up to 21 TLPs and 8 additional small FPSOs (best case) to only drilling subsea wells (worst case). This solution was shown to have a significantly higher expected NPV compared to the deterministic solution (NPV= $\$4.4 \times 10^9$). Bora is obtaining results for the case of a nonlinear production profile, and is close to completing the manuscript of this work.

Simultaneous Planning and Scheduling of Multiproduct Plants

New developments: **Submission manuscript for scheduling of reactors with finishing lines**

Students: **Muge Erdirik [Ph.D. started Jan 2004; completed August, 2007]**

New student: **Sebastian Terrazas Moreno [Ph.D. started on Jan 2008]**

Muge's project defended her Ph.D. thesis on August 28. Her project dealt with the simultaneous planning and scheduling of multiproduct plants. Sebastian Terrazas, a new Ph.D. student from Universidad Iberomericana in Mexico, will continue with Muge's project focusing on the scheduling of multisite problem.

The initial objective of Muge's work has been to consider the case of a single processor on which a number of products must be produced continuously over a given number of time periods. Sequence-dependent changeovers are given and due dates are assumed to take place at the end of each time period. The objective is to determine a production plan and schedule in order to maximize the profit (income minus

inventories and changeovers). She has also considered recently the extension to the case of parallel processors, for which she applied ideas of the parallel batch reactors described below. To address the scheduling problem Muge developed a detailed slot-based MILP scheduling model. Since this model cannot be used to solve planning problems with long time horizons, she developed an iterative scheme that relies on aggregated models and that is guaranteed to converge to the same solution as if we had solved the detailed scheduling model. Muge developed an aggregated MILP model that underestimates the changeover times and includes 0-1 variables for deciding what products may be produced in a detailed schedule. The iterative scheme consists in solving a sequence of aggregated MILP planning models and detailed scheduling models with fixed assignment of products to be produced at each time period. In addition to using a simple integer cut, Muge developed superset, subset and capacity cuts that eliminate a larger number of alternatives. The results that Muge obtained were quite encouraging.

Based on an internship at Dow Chemical in Midland, Muge addressed the planning and scheduling problem for parallel batch reactors, a case study that was selected by Dow for the Enterprise-wide Optimization project. In this problem we are given a unique set of raw material costs and availability, storage tanks with associated capacity, reactors with associated materials it can produce and batch sizes and times for each material it can produce, as well as operating costs for each material, and sequence dependent clean out times. We are also given a set of customers, each with a set of demand and prices for desired products. Finally, specified are the materials produced: process intermediates and final products. Dedicated storage tanks are considered. Another issue is that a final product of one process may be used as a raw material for another process. However, once a final product is fed to the dedicated storage tank, it can not be retrieved back to the plant. The problem is then to determine the monthly production quantities for each reactor and the assignment of materials to storage tanks to maximize profit. Muge proposed a novel continuous time MILP optimization model for scheduling that is based on slot time representation that overcomes some of the major difficulties faced by the STN and RTN discrete and continuous time models. While effective for short-term scheduling, the proposed model becomes computationally very expensive to solve for long planning horizons. Therefore, as in the case of the single continuous processor, Muge devised a rigorous bi-level decomposition algorithm. The aggregate upper level planning model determines the products to be produced at each time period as well as number of batches of each product, production levels and product inventories. This model is based on a new relaxed STN model where the detailed timing constraints and changeovers are replaced by time balances yielding a tight upper bound on the profit. A major new development here has been the *incorporation in the aggregate planning model of sequencing constraints* similar to the ones in the traveling salesman problem that yield accurate predictions for the changeovers at the planning level. The lower level involves solving the scheduling problem in the reduced space of binary variables and with number of slots according to the information obtained from the upper level model, yielding a lower bound on the profit. The procedure iterates until the difference between the upper and lower bound is less than a specified tolerance. The biggest surprise with the new aggregated planning model is that for small problems (e.g. 5 products, 2 reactors) it usually predicts the exact scheduling solution (i.e. zero gap). For larger problems it predicts much smaller gaps than a relaxed model that simply underestimates the changeover times. A limitation, however, is that the aggregate planning model leads to a larger MILP model, which, however, can be solved with a rolling horizon approach with relatively small compromise of the optimality. The largest problem solved by Muge involved 15 products and 6 reactors, over 48 week period. The problem involved 10,092 0-1 variables, 25,798 continuous variables, and 28,171 constraints and was solved in about 11,000 secs using the rolling horizon approach. As for scheduling problems, the two level procedure works very well thanks to the accuracy of the planning model. For example a problem with 4 reactors, 6 products and 1 week, was solved in 4 major iterations, requiring 460 secs. Several papers emerged from this work.

In the last part of her Ph.D. Muge worked with John Wassick on an important extension, the short-term scheduling of a multi-product batch plant which consists of parallel batch reactors that are connected to continuously operating finishing trains to form work groups. Finishing operations are required to convert the outputs from the reactors to trade products for a diverse set of markets and customers. A complication is that each time a product switch occurs, not only the reactors but also the finishing trains need to be cleaned up and made ready for the next product. Since these clean up operations involve sequence-dependent changeovers, determining the optimal sequence of production is of great importance for improving equipment utilization and reducing the costs. The main challenge of modeling this scheduling

problem arises from the structure of the plant, where the work groups are not fixed, but are flexible in the sense that subsets of work groups can be selected by manipulating valves that interconnect the reactors with the finishing trains. Therefore, in addition to determining the optimal production sequence given the sequence-dependent changeovers with high variance, there is the challenge of regrouping the units periodically when the demand varies from one period to the next one, or when new products are introduced while maximizing profit. In order to address these issues, Muge used as basis long-term planning model and added special constraints to enforce that workgroups have the same assignment and sequence of products. As an example, Muge solved a problem with 10 products, 6 reactors and a horizon of 4 weeks, which gives rise to an MILP with 1992 0-1 variables, 2904 continuous variables and 6183 constraints. The optimal solution was obtained in 644 CPUs. This manuscript has been submitted for publication and is available to CAPD members and is available to CAPD members.

Design and Planning of Responsive Supply Chains

New Development: **Completed and submitted paper on probabilistic inventory for safety stocks
Lagrangian decomposition for joint location-inventory problem**

Students: **Fengqi You [Ph.D. started Jan 2006]**

Fengqi successfully passed his Ph.D. qualifier. The major goal of this project, which is being performed in the context of the Enterprise-wide Optimization initiative, is to develop a comprehensive optimization model that allows the effective integration of long term strategic design decisions for multisite process manufacturing facilities, with short term planning for production and distribution that accounts for the time delays and inventories across the supply chain. The motivation for this project is that virtually all optimization models assume that material can be transferred instantaneously. In that way lead times, or response times once orders are placed, are not taken into account. It is our major goal to establish the trade-offs between economics and deterministic and probabilistic lead times.

As a first step Fengqi addressed the long term design problem of a supply chain, for which he developed a superstructure model that involves both dedicated and multiproduct continuous plants. The time horizons considered are of the order of years, and no inventories are considered, which qualitatively corresponds to the worst case for deterministic lead times. The problem is posed as a bi-criterion optimization problem in which the objectives are to maximize net present value and to minimize lead time. In order to reflect lead times for different choices of topologies Fengqi considered constraints that measure the duration of the longest time path of chemical flows from a supplier to a customer by way of manufacturing sites. The proposed constraints involve transportation times and residence times in processes. For dedicated plants the times these are simply constants, while for multiproduct plants they correspond to cycle time plus residence time minus its processing time. Note that the cycle time results from a scheduling optimization that has to be accounted for. In the case of dedicated plants the problem can be formulated as an MILP, since nonlinear terms defining the lead times can be linearized. For the case of multiproduct plants, the model and leads to a nonconvex MINLP problem. Fengqi obtained results on a production network for polystyrene resins that involves a dedicated process (styrene), and two multiproduct plants (solid polystyrene and expandable polystyrene). Three potential plant sites are located in PA, TX and AL. the PA site can install all the three types of plants, the TX sites can only install plant I, and the AL site can only install plant 2 & 3. Two suppliers of ethylene are located in TX and OH, and two suppliers of benzene are located in TX and VA. Two customers of SPS resins are located in CA and NY, another two customers of EPS resins are located in GA and MN. For 3 time periods (2, 3 and 5 years) the model involved 133 0-1 variables, 2249 continuous variables and 3041 constraints. To develop the trade-off curve the ϵ -constraint method was used. DICOPT required 3.5 minutes while BARON required about 100 hours. The solution involving shortest lead time of 8.85 days had an NPV of \$158 million, while the longest lead time was 14.42 days at a much higher NPV of \$1,261 million. This work was recently published in a book on supply chain optimization by Georgiadis and Papageorgiou.

Fengqi extended the above problem to a probabilistic model for stockout. In this way instead of using a deterministic lead time that only relies on transportation and production times with zero-inventory, Fengqi

proposed an expected lead time that represents the expected value of time delays incurred by transportation and production across the supply chains, and accounting for inventories in the distribution centers. This expected lead time is then used as a quantitative measure of supply chain responsiveness. This probabilistic model for the safety stock levels accounts for demand uncertainty. This probabilistic constraint is reformulated as deterministic constraints for which Fengqi obtained analytical expressions for the triangular and the normal distributions of the demands. The bi-criterion optimization model was also solved with the ϵ -constraint method, but Fengqi has also developed a hierarchical algorithm for the solution of the resulting large-scale MINLP problem that is based on decoupling of the decision-making levels (strategic and operational). Fengqi solved two examples related to the styrene problem described above. The results show that small changes in expected lead time can lead to significant changes in the net present value and the network structure, which in turn suggests the importance of integrating responsiveness into the design and operations of process supply chain network. An interesting by-product of the model is a trade-off curve of safety stock versus expected lead time, which provides very useful information. Fengqi has completed a manuscript on this work which has been submitted for publication and is available to CAPD members.

Fengqi has also been working on a classical operations management problem for joint location-inventory problem inspired by the work of Shen, Coullard, and Daskin (2003). The problem is as follows. The locations of the supplier and the retailers are known and the distances between them are given. The supplier to retailer lead time is assumed to be the same for all the retailers. A number of candidate sites are given to locate the distribution centers (DC), and there are fixed setup costs when distribution centers are established. Each retailer has a normally distributed demand which is independent of other retailers' demands. Each DC can connect to more than one retailer, but each retailer should be only assigned to exactly one distribution center to satisfy the demand. Linear transportation costs are incurred for shipments from supplier to DC and from DC to retailer. Most of the inventory in the network is hold in the DCs, where the inventory is managed with a (Q, r) policy with type I service. Inventory costs are incurred at each DC and are consisting of both working inventory and safety stock. The retailers only maintain a very small amount inventory whose costs are ignored. The objective is to determine how many DCs to locate, where to locate them, which retailers to assign to each DC, how often to reorder at the DC, and what level of safety stock to maintain so as to minimize the total location, transportation, and inventory costs, while ensuring a specified level of service. Fengqi first reformulated the model as a mixed-integer nonlinear programming (MINLP) problem. Using its convex relaxation model for preprocessing, he developed a local optimization heuristic method to obtain near-global optimal solution very quickly. Based on this model, Fengqi also proposed a Lagrangean relaxation and decomposition algorithm for the global optimization of the model. He obtained numerical results for examples ranging from 33 to 150 retailers. Fengqi is in the process of completing a manuscript on this work.

Optimal Scheduling of Crude Oil Operations

New development: **Symmetry breaking constraints for new continuous time MILP model and their integration through CP**

Student: **Sylvain Mouret [Ph.D. started Jan 2007]**

Sylvain, who joined the department in January from Ecole Polytechnique in Paris, is working in a new project funded by Total through the collaboration with Pierre Pestiaux. The specific problem that Sylvain is addressing is the scheduling of crude oil operations in the front-end of a refinery that is composed of crude marine vessels, storage tanks, charging tanks and CDUs. Crude vessels unload crude oil into storage tanks during a time window depending on the arrival time to the refinery. These crude-oils are then mixed into charging tanks before being distilled by CDUs which separate the charged oil into fractions such as gas oil, gasoline, kerosene, heating oil and residues. Given arrival times of marine vessels, capacity limits of tanks, flow rate limitations, initial key components concentrations in vessels and tanks, components concentration ranges as part of distillation specifications, demands for each mixed oil, and a time horizon, the objective is to determine time and volume variables of all crude-oil transfer operations in order to maximize the gross margins of distilled mixed oil.

Sylvain has developed a novel continuous time model for this scheduling problem that relies on the idea of postulating potential assignments of tasks (slots) to operations that include the various transfers in the network. The proposed approach consists then of assigning specific transfer operations to an ordered set of tasks whose cardinality has to be postulated. The objective used is to minimize the cost of crudes that need to be processed to satisfy given demands. By introducing the appropriate 0-1 variables for assignments, as well as disjunctive constraints that enforce non-overlapping operations (e.g. inlet and outlet of a given tank), the problem can be formulated as an MILP provided the constraints on compositions for blending are relaxed. Level variables are used to control tank inventory and composition as well as crude-oil mixture recipe and demand. Volume variables are used to control transfer operation flow rate and composition. Once the solution is obtained for this problem, an NLP subproblem is solved with fixed 0-1 variables to enforce the composition constraints. This yields an upper bound, which surprisingly Sylvain has found to be very tight, and in many instances with zero-gap. Sylvain has applied this solution approach to several problems from the Lee et al. (1996) paper obtaining very encouraging results,

One feature in the proposed MILP model, however, is that it has many degenerate or “symmetric” solutions. Qualitatively the reason is that different assignments of the binary variables may represent the same schedules. This has a negative impact on the branch and bound method as it leads to the enumeration of a larger number of nodes. Sylvain has addressed the symmetry challenge by using concepts from regular languages theory. Sylvain defined a regular language, and its corresponding Deterministic Finite Automaton (DFA), in order to restrict the possible sequences of operations assigned to the set of time-slots. This restriction is then added to the model under the form regular constraints which are linear network flow constraints as explained in Côté et al. (2007). Sylvain tested the model on the 4 instances of a crude-oil operations scheduling problem from Lee et al. (1996). The results show that, even though many new variables and constraint are added to the original model, the search space is drastically reduced (<100 nodes explored) so that the overall computational expenses become much smaller for large instances with many time-slots. The largest problem had 3 vessels, 6 storage tanks, 4 charging tanks and 3 CDUs. Postulating 30 tasks (slots) the MILP was solved in 567 sec (53 nodes) and the NLP in 31 secs leading to a solution with 0% gap. Sylvain has taken this approach one step further by implementing a constraint programming procedure for performing inference on the regular constraints at the branch and bound enumeration of the MILP in order to avoid the large size of the LPs. Sylvain is currently testing several branching rules that rely on time-slot ordering and operations ordering. Preliminary results have provided significant reductions. For instance the time for the largest problem could be reduced from 560 secs to 114 to 146 secs depending on the branching rule that is used. Sylvain will continue exploring the best way of integrating CP into this MILP model.

Planning of Refinery Operations

New development: **Implementation of swing cuts method**

Student: **Abdulahraman Alattas [Ph.D., started Jan 2007]**

Abdul successfully passed his Ph.D. qualifying exam. His project is a joint collaboration with BP through Ignasi Palou-Rivera in the Enterprise-wide Optimization project. The major objective is to develop refinery planning models that incorporate nonlinear process models, in contrast to the common fixed yield models that lead to robust LP models that are prevalent in industrial practice (eg. PIMS).

The specific problem that Abdul is addressing is for complex refinery configurations for processing heavy crudes. The crude is introduced into the crude distillation unit (CDU) that combines the atmospheric distillation column and the vacuum distillation column and produces the first cuts of the crude. These include the overhead fuel gas, straight-run naphtha, straight-run gasoline, straight-run light distillate, straight-run gas oil and the bottom residue. Along with the CDU, the configuration of interest includes the following process units: naphtha reforming unit that produces reformed gasoline and fuel gas from straight-run naphtha, catalytic cracking unit that produces fuel gas, gasoline and fuel oil from straight-run light distillate and straight-run gas oil, residue hydrotreating unit that treats the bottom residue to blending

quality, and product blending units that combine different intermediate product streams to produce the desired final products: premium gasoline, regular gasoline, diesel, fuel oil and residue. The objective of the planning model is to determine the types, quantities and mixing strategies for the different crude oils available for purchase, so that the refinery will meet the objectives of maximizing profits while meeting specific demands over a specified time period. Abdul has implemented both fixed-yield models and swing-cuts models in order to assess their relative benefits. The swing cuts model, which can also be formulated as an LP, can be thought of a model that has the flexibility of transgressing the fixed boundaries of the fixed yields models within certain limits. In a specific instance Abdul found that the swing cuts model can predict a solution with 10% improvement in the profit, largely due to different decisions in the purchase of crudes.

In the last few months Abdul has investigate the feasibility of incorporating the aggregate distillation models proposed by Jose Caballero, as well as related short-cut methods. This requires the CDU to be represented through cascaded columns, including columns with steam stripping. A direct implementation of the aggregate model led to significant convergence problems, often yielding infeasible NLP solutions. In order to improve the convergence, Abdul has derived valid linear inequalities for the flows of vapor and liquid, since these obey complex relations given the thermal integration of streams in the cascaded columns. These inequalities have shown to significantly improve the robustness of the solution of the NLPs. Abdul has been able to solve up to 4 cascaded columns with up to 20 components. He is now in the process of adapting his models to the use of pseudocomponents, and exploring the use of other short-cut methods that make use of fractionation indices.

Planning for Petroleum Allocation

Post-doctoral fellow: Roger Rocha [started September 2007]

This is a new project involving Roger Rocha from Petrobras who is visiting for two years. The main objective of this work is to investigate a mathematical programming approach to solve the Petroleum Allocation Problem at Petrobras. Petroleum allocation must be programmed so that sufficient supplies of required crude oil reach refineries along the planning horizon. This must be done by taking into account strategic planning and operational constraints along the petroleum supply chain as follows. Crude oil can either be locally produced or imported from abroad. Local crude oil comes from production sites, mostly offshore, and is transported either by tankers or pipelines. Imported oil is only transported by tankers. After reaching maritime terminals, crude oils are either exported or shipped to Petrobras refineries. At the refineries, petroleum is processed in crude distillation units (CDUs) on daily scheduled production campaigns. These campaigns are defined by: consumption rates of different petroleum categories, duration, release date, and deadlines to completing them. Roger has proposed an MILP formulation of the problem which relies on a time/space discretization network. The formulation involves some inequalities which are redundant to the mixed integer model but not necessarily to LP relaxation. As the model itself was not able to solve industrial size instance of the problem, not even to find a feasible after 15 days of computation, he implemented a heuristic to find an initial feasible solution by fixing some variables associated with the hardest part of this model and solving the remaining part. In order to improve the solution quality we use a local search method by optimization called local branching, that in most of the case studies could find a solution guaranteed to be no more than 10% of optimality in less than 5 hours.

Recently Roger has developed some valid inequalities which are associated with polytopes that have been extensively studied in the literature. Furthermore, separation routines for strong valid inequalities associated with these polytope are readily available in some commercial solvers. Use of this feature allows a substantial reinforcement of the underlying LP relaxation to be attained. He has tested on some industrial-size instances of the problem involving approximately 40 crude oils, 6 tanker types, 8 maritime terminals involving 12 docks, 11 refineries, and 21 crude distillation units over a time horizon of 72 discretized intervals. Typical instances have 40,000 binary variables, 160,000 continuous variables, and 40,000 constraints. He is currently performing computational tests to compare his results against XPRESS-MP.

Scheduling of Batch Multiproduct Plants

New Development: **Publication of paper for simultaneous batching and scheduling**

Collaboartors: **Pedro Castro (INETI, Portugal)**

Pedro Castro has been collaborating with us for a number of years, especially in light of his postdoctoral stay in 2004-2005. He has mostly devoted his efforts to continuous time RTN models, and to developing effective computational strategies for effectively solving MILP models for large scale scheduling problems.

More recently Pedro with input from Muge has a new mixed integer linear program (MILP) for the optimal short-term scheduling of single stage batch plants with sequence dependent changeovers and optimal selection of the number of batches to produce. It is a continuous-time formulation employing multiple time grids that is based on the resource-task network (RTN) process representation. The main novelty is that aggregated processing and changeover tasks are considered that account for the time required to produce all batches of the product, plus the changeover time to the next product in the sequence. The performance of the new formulation was studied by Pedro through the solution of 12 example problems for the objective of revenue maximization and 4 for the objective of makespan minimization. The same problems were solved by a multiple time grid implicit batching approach, by a continuous-time model with global precedence sequencing variables, by a model with immediate precedence sequencing variables that does not determine the timing of events, and by a constraint programming model. The new formulation emerged overall as the best performer for the scenario of maximum plant flexibility, where different batches of the same product can be produced in different units. The model developed by Muge with immediate precedence sequencing variables was the fastest but it is not a general scheduling model in the sense that it assumes a single cyclic schedule in each unit, which can be broken, but two or more cyclic schedules per unit may result. In such cases, subtour elimination constraints can be added and the problem solved iteratively to find a feasible schedule at the likely expense of removing the global optimal solution from the feasible space. When compared to the implicit batching approach, the computational effort of the new formulation was typically one order of magnitude lower, which in practice indicates that the new formulation can tackle larger problems. When compared to the traditional approach of considering a single processing task per batch, fewer event points are needed, which results in significantly lower computational effort as illustrated through the solution of several example problems. A manuscript describing this work has been published and is provided as a reprint.

Scheduling of Paper Machines with Replacement of Felts

Visiting scholar: **Daniela Drummond [started January 2007]**

Daniela, a visitor from University of Campinas in Brazil, has been developing an MINLP model for the optimization of the production schedule and replacement of felts in the press section of a paper machine. Using multivariate statistics through the methods of PCA (Principal Components Analysis), PCR (Principal Components Regression) and PLS (Partial Least Squares) Daniela has developed empirical models for predicting the useful life of felts for absorbing water in the press section of paper machines. These models have been incorporated into an MINLP that determines the planning of production of paper in order to minimize the cost, which consists of replacement of the felts in the press section, cost of energy to operate the presses, and cost of energy in the drying section. The constraints of the model consider the water removed in the pressing and the drying, as well as the efficiency of the felts in the press section that varies with the running time. Daniela has solved several examples using industrial data. The results show that improvements of 6% are obtained in the water removal for pressing, and 2% in the reduction of total cost.

Simultaneous Scheduling and Control

New development: **Simultaneous design, scheduling and control of a Methyl-Methacrylate Continuous Polymerization Reactor**

Collaborator: **Antonio Flores-Tlahuacac (Professor U. Iberoamericana)**

This is a collaboration with Antonio Flores from the Universidad Iberoamericana (Ignacio's alma mater). Antonio, being a control person, has been motivated by the idea of using dynamic models for control in scheduling problems. Given our interest in the area of scheduling, we decided to join forces in a work in which the basic idea is to combine a scheduling optimization model, with a dynamic model for the optimization and control of the transitions. In this work we addressed the simultaneous scheduling and control problems for a continuous stirred tank reactor (CSTR) that produces multiple products within a cyclic schedule that involves sequence dependent transitions. The scheduling part of the model is similar to our previous work with Jose Pinto. Because, dynamic profiles of both manipulated and controlled variables are also decision variables, the dynamic part of the model is formulated with DAEs. The resulting problem can then be cast as a Mixed-Integer Dynamic Optimization (MIDO) problem in which the objective function includes income from sales of products, inventory costs and transition costs that takes into account through quadratic deviation terms, the amount of off-specification material produced during product transition. To solve the MIDO problem it is transformed into an MINLP using orthogonal collocation, where initial guesses on the duration of the transitions must be provided. These are then revised in a subsequent optimization. Antonio and Sebastian Terrazas-Moreno subsequently developed a new MINLP formulation to simultaneously solve the scheduling and control problems in polymerization reactors during a cyclic manufacturing operation. In contrast to the previous model, the problem need not be solved sequentially by iteratively assuming fixed lengths for the duration of the transitions. Another interesting feature of the model is that by assuming that the performance of each polymer is only dependent of the initial conditions, the problem can be formulated in terms of 0-1 variables for assigning products to slots. The transitions are then simply computed from the difference in the initial conditions. This leads to a much smaller number of 0-1 variables compared to the case when sequence dependent transitions are explicitly modeled.

In another work, Antonio in collaboration with Sebastian Terrazas addressed the simultaneous scheduling and control (SSC) problem using Lagrangean Decomposition as presented by Guignard and Kim. The model was decomposed into scheduling and control subproblems, and solved using a heuristic approach used before by Sarette Van den Heever in oilfield problems. The method was tested using a Methyl Methacrylate (MMA) polymerization system, and the High Impact Polystyrene (HIPS) polymerization system, with one continuous stirred-tank reactor (CSTR), and with complete HIPS polymerization plant composed of a train of seven CSTRs. In these case studies, different polymer grades are produced using the same equipment in a cyclic schedule. The computational times in the first two examples were lower for the decomposition heuristic than for the direct solution in full space, and the optimal solutions found were slightly better. The example related to the full scale HIPS plant, which involved 16 0-1 variables and 22,700 variables, was only solvable using the decomposition heuristic. The manuscript describing this work has been published and is available in the next newsletter.

Finally, in recent work Antonio and Sebastian developed a Mixed-Integer Dynamic Optimization (MIDO) formulation for the simultaneous process design, cyclic scheduling, and optimal control of a Methyl Methacrylate (MMA) continuous stirred-tank reactor (CSTR). Different polymer grades are produced in terms of their molecular weight distributions, so that state variables values during steady states are kept as degrees of freedom. The corresponding mathematical formulation includes the differential equations that describe the dynamic behavior of the system, resulting in a MIDO problem. The differential equations were discretized using the simultaneous approach based on orthogonal collocation on finite elements, rendering a Mixed Integer Non-Linear programming (MINLP) problem where a profit function is to be maximized. The objective function includes product sales, some capital and operational costs, inventory costs, and transition costs. The optimal solution to this problem involves design decisions: flowrates, feeding temperatures and concentrations, equipment sizing, variables values at steady state; scheduling decisions:

grade productions sequence, cycle duration, production quantities, inventory levels; and optimal control results: transition profiles, durations, and transition costs. Antonio and Sebastian formulated and solved the problem in two ways: as a deterministic model and as a two-stage programming problem with hourly product demands as uncertain parameter described by discrete distributions. The paper describing this work has been completed and is available to the CAPD members.

Software for MINLP Optimization in Design and Scheduling

Research Assistants: Rosanna Franco (started July 2006)

Rosanna Franco has been working on the web-interfaces that are available in:
<http://newton.cheme.cmu.edu/interfaces>

She completed the new PC-interface Synheat for handling isothermal streams that was developed by Jose Maria Ortega and Arturo Jimenez. That interface should be available soon on the web. Rosanna is also completing a new interface for integrated water systems, based on the work of Ramkumar Karuppiah.

The current list of programs that we have available, most of them in our website, are the following:
 (description in <http://egon.cheme.cmu.edu>)

Synthesis:

SYNHEAT	MINLP synthesis heat exchanger networks (Yee) Also includes transshipment model for targeting (Papoulias)
STEAM	MINLP Model for the synthesis of combined cycles for utility plants (Bruno) Model includes correlations for steam, efficiencies and cost data
GLOBESEP	Global NLP optimization for synthesis of separation networks and single feed/mixed products (Quesada)
WATER	Global NLP Model for synthesis of wastewater treatment configuration (Galan)
EXTRACTOR	Disjunctive MINLP for synthesis of liquid-liquid extraction systems (Reyes)
GDP-DISTILL	GDP Model for the optimal selection of number of trays and feed tray location in distillation columns using tray-by-tray model (Barttfeld)

Batch design:

BATCHSPC	MINLP and MILP models for multiproduct batch plants single product campaigns (Kocis, Voudouris)
BATCHMPC	MILP model for multiproduct batch plants mixed-product campaigns (Birewar, Voudouris)

Scheduling:

STN	State-Task-Network MILP formulation for scheduling multipurpose batch plants. Both the Kondili, Pantelides and Sargent (1993) model and the Maravelias and Grossmann (2003) are implemented.
PARALLEL	MINLP continuous multiproduct scheduling on parallel lines Features feasibility preanalysis (Sahinidis)
MULTISTAGE	MINLP continuous multiproduct in multistage plants (Pinto)
CYCLE	LP/MILP aggregate flowshop scheduling (cycle time/makespan) Includes loop tracing algorithm (Birewar)
STBS	MILP short term multistage scheduling (Pinto, Bolio)
CRUDEOIL	MILP model for refinery scheduling (Lee, Pinto)
DECAY	MINLP model for scheduling of clean-up of parallel furnaces (Jain)
UTILPLAN	MILP multiperiod model for utility plants (Iyer)
PRODEV	MILP model for scheduling of tests in new product development (Schmidt, Najimas) MILP for resource scheduling in new product development (Jain, Maravelias)

Planning:

PLANNER	MILP multiperiod model for capacity expansion in process networks (conventional and lot sizing model) (Sahinidis, Norton)
MULTISITE	MILP model for planning the selection of processes and capacity expansion in

GREENPLAN	different geographical location and accounting for transportation costs (Turkay) Bi-criterion MILP model for selection of processes that maximize the net present value and minimize toxicity (Drabbant)
NETCHAIN	Multiperiod MILP for supply chain optimization of multisite facilities with flexible processes, intermittent deliveries, changeovers and inventories (Bok/Iyer)

Nick Sahinidis' Group

THRUST 1—OPTIMIZATION ALGORITHMS, THEORY, AND SOFTWARE

Algorithms and Software for Global Optimization of NLPs and MINLPs

Student: Ms. Xiaowei Bao (Ph.D. student in Chemical and Biomolecular Engineering at the University of Illinois)

Collaborator: Mohit Tawarmalani (Associate Professor, Purdue University)

As a result of research that was initiated in 1991, we have developed algorithms and software for global optimization of NLPs and MINLPs. The main results of this work have been:

- A theory of convex extensions that provides a systematic means for developing the functional expression of convex/concave envelopes of nonlinear functions of continuous and integer variables.
- An entirely linear outer-approximation of factorable nonlinear programs. Relying on LP relaxations results in a robust solution approach to global optimization.
- A variety of range-reduction schemes for reducing the search space of integer and nonlinear programs.
- Finite branching schemes for concave minimization and two-stage stochastic integer programs.
- The BARON software for the global optimization of NLPs and MINLPs.

Xiaowei Bao last semester initiated work on the global optimization of quadratically-constrained nonconvex quadratic programs. This is a class of optimization problems with applications in various settings, including facility location, multiperiod refinery scheduling and planning, and circle packing problems. The approach that Xiaowei is currently investigating relies on branch-and-bound. While standard approaches relax each nonconvex term separately, Xiaowei is exploring relaxations of entire quadratic constraints. Based on results from the theory of convex extensions (Tawarmalani and Sahinidis, 2002), these relaxations are polyhedral and have the advantage of capturing interactions between several nonconvex terms simultaneously, thus resulting into tighter bounds and faster convergence.

Nick Sahinidis and **Mohit Tawarmalani** are currently maintaining the BARON software. Over the past year, updates to BARON have included (1) the introduction of a new heuristic for local search and (2) improved range reduction for monomial functions. In addition, an interface to the Xpress LP code and the MOSEL modeling system was completed in summer 2007 and is currently in the final testing stages by a customer of Dash Optimization. Ongoing work on BARON involves the development of tighter relaxations for MINLPs, work that is expected to be implemented and made available by the end of this coming spring semester.

Global Optimization for Nonlinear Least Squares Problems

Student: Mr. Keith Zorn (Ph.D. student at CMU)

Keith Zorn who joined the group in November has begun the study of least squares parameter estimation problems from a global optimization point of view. This problem will serve to train Keith in the area of global optimization, before he embarks on the study of more complex problems. In the long run, Keith is currently planning of investigating applications of global optimization in quantum chemistry. The main questions that Keith will address in the context of nonlinear least squares are: (1) does global optimization

provide better solutions to standard curve fitting problems and (2) can global optimality be reached within reasonable computing times, i.e., a small multiple of the time taken by local optimization algorithms?

Global Optimization for Machine Learning Problems

Student: Mr. Xuan Shi (M.S. student at CMU)

Xuan Shi, who joined the group in November, is working on applications of global optimization in the area of Machine Learning. One of the central questions in neural computing is how to select the neural network parameters (architecture, activation functions, and weights) so as to obtain the best possible neural network model for a given natural or artificial system. This project considers the problem of determining the global minimum of an error function commonly used for training of neural-networks. This is a highly non-linear problem plagued with multiple local optima. Local search algorithms, such as backpropagation, often result in neural networks that are unnecessarily complex and may overfit the data. The project will involve the computational implementation and testing of an algorithm developed earlier by Voudouris and Sahinidis (unpublished), while, in addition, enforcing the first-order optimality conditions. A general-purpose code will be written in GAMS and a generic interface will be developed in order to facilitate testing of the new training algorithm on a variety of test problems from the Machine Learning Repository at <http://mlern.ics.uci.edu/MLRepository.html>. This new training algorithm will be evaluated by comparing it against backpropagation, support vector machines, and other popular learning algorithms.

Algorithms and Software for Black-box Optimization

Student: Mr. Luis Miguel Rios (Ph.D. student in Industrial Engineering at the University of Illinois)

This project is currently aiming at a systematic testing of existing derivative-free algorithms that are capable of optimizing black-box problems. Derivative-free optimization is an area of recent interest and rapid growth, fueled by a growing number of applications, especially in the oil and gas, and chemical process industries. The major challenge is that the objective function in many problems is expensive to evaluate, while no bounds or Lipschitz constants are available, and strategies to directly estimate derivative information are impractical or expensive. The most recent systematic testing of derivative-free algorithms for solving problems of this nature was done 9 years ago. **Luis Miguel Rios** has collected 225 test problems from the globallib and princetonlib collections and solved them under different conditions using 18 different black-box solvers. These codes were: Adaptive Simulated Annealing, ASA version 26.23 by Ingber; Solvers APPS, DIRECT, EA, PATTERN, and SOLIS-WETS under Design Analysis Kit for Optimization and Terascale Applications (DAKOTA), Version 4.0, from Sandia National Laboratories; Derivative Free Optimization (DFO), Version 2.0, by Conn, Gould, Toint, and Scheinberg; Multilevel Coordinated Search (MCS), by Huyer and Neumaier; the Nelder-Mead algorithm implemented in MATLAB; Nonlinear Optimization for Mixed Variables and Derivatives (NOMAD), Cycle 6, by Couture, Audet, Dennis, and Abramson; SID-PSM, Version 0.3, by Custódio and Vicente; and solvers EGO, GLB, GLC, LGO, and RBF under TOMLAB.

The main conclusions from this computational study were that: (a) even obtaining a feasible solution cannot be taken for granted for these problems/solvers, (b) larger problems diminish the chances for obtaining good solutions, (c) LGO and MCS are better, on average, than other solvers, (d) all solvers are 'useful' in the sense that there are at least a few problems for which each solver is best in terms of solution quality.

Current plans in this line of research include the addition of non-smooth problems in the test set collection and the inclusion of additional solvers. Test problems of black-box models from industry are currently sought. In the long run, we plan to develop novel algorithms for this difficult class of optimization problems.

Algorithms and Software for Linear Optimization Problems (LP) and Extensions

Students:

Mr. Deepak Channamariyappa (M.S. student at CMU)

Mr. Joseph Elble (Ph.D. student in Industrial Engineering at the University of Illinois)

Mr. Yiqi Zhu (M.S. student at CMU)

Algorithms for solving LPs represent the workhorse of optimization systems for solving large-scale MILPs, NLPs, and MINLPs. While highly successfully commercial software exist for solving LPs, none is guaranteed to provide reliable results in the sense that they are all subject to the effects of floating point arithmetic and round-off errors. The goal of this project is to develop new simplex algorithms and software in order to provide tools that can reliably solve large-scale linear optimization problems even in the presence of round-off errors. Towards this goal, we plan to develop symbolic algorithms for linear programming preprocessing, matrix scaling, and matrix factorization.

Over the past three years, **Joseph Elble** worked simultaneously on two related M.S. theses, one in Industrial Engineering and one in Computer Science, both of which were completed in August 2007.

In his Industrial Engineering thesis, **Joseph Elble** investigated the current state of direct methods for solving sparse linear systems. The data structures for sparse linear systems were studied, and the thesis explored the four stages of direct methods: analysis, numerical factorization, forward and backward elimination, and iterative refinement. This work was exploratory in nature, aiming at understanding conditions under which current methods for linear systems work well and conditions under which such methods are likely to give incorrect results.

In his Computer Science thesis, **Joseph Elble** studied scaling of LPs. Scaling of linear programs, while poorly understood, is definitely not devoid of techniques. Scaling is the most commonly used preconditioning technique utilized in linear programming solvers. The purpose of scaling is to improve the conditioning of the constraint matrix and hopefully decrease the computational effort for solution. Most importantly, scaling provides a relative point of reference for absolute tolerances. Many techniques for obtaining scaling factors for linear systems were investigated in this thesis. With a focus on the impact of these techniques on the performance of the simplex method, over half a billion LP solves were executed in Nick's Linux cluster at Illinois. Some of the scaling techniques studied are, by construction, computationally more expensive than others. Some are more successful in practice than others. There is no current scaling technique that dominates all others.

The long-term goal of this project is to develop novel algorithms for LPs and make them available via distributed implementations on modern FPGAs.

Deepak Channamariyappa has begun the study of preprocessing techniques for Linear Programming. These techniques currently represent more of an art rather than science in the field of Linear Programming. The main goal of Deepak's project is to implement and compare the computational performance of different preprocessing techniques. It is expected that this study will be beneficial for the development of useful tools for nonlinear optimization problems as well.

Yiqi Zhu has begun the study of successive Linear Programming algorithms for the solution of nonlinear programs. These techniques were developed in the sixties by the chemical process industries but have not resulted in successful general-purpose NLP software. However, they appear to be promising for the solution of very large-scale nonlinear optimization problems with a very large number of nonzeros and degrees of freedom. We envision the development of a trust region technique that is provably convergent and is successful for such problems, when implemented on top of a robust and efficient linear programming solver.

THRUST 2—APPLICATIONS OF OPTIMIZATION IN BIOLOGY, CHEMISTRY, ENGINEERING, AND MEDICINE

Inverse Imaging Problems in X-ray Crystallography

Student: **Mr. Alexander Barton Smith (Ph.D. student in Chemical and Biomolecular Engineering at the University of Illinois)**

A vast majority of 3D structures of compounds are derived using the physics of X-ray diffraction, especially in the case of important biological macromolecules. The computation of a structure from X-ray diffraction data, however, remains a very challenging, nontrivial problem, both experimentally and computationally. A major obstacle, coined, “the phase problem,” represents a dilemma in which phase information, critical to computation of the 3D structure of a crystal, is not directly measurable in a traditional X-ray diffraction experiment.

The phase problem has recently been approached via combinatorial optimization techniques and the resulting sieve method has been demonstrated to be effective for phasing centrosymmetric structures (Smith, Xu and Sahinidis, 2007). The purpose of the current work is to (a) develop a more robust enforcement of atomicity constraints in direct space and (b) extend the approach of Smith et al. from the centric to the most challenging non-centric case, thus making the approach directly applicable to proteins and other chiral molecules.

Alexander Smith has recently developed a mixed-integer linear programming model for phasing based on the minimal principle; one which includes the introduction of specific atomicity constraints. Building on the MILP model presented in Smith et al. (2007), atomicity is constrained through sampling of electron density on a grid. First, a set of points is selected, at random, to sample the unit cell. Electron density is calculated at these points in terms of the integer variables, which describe the phases. Physical constraints in direct space are then enforced using the grid of electron density. Computational results demonstrate that the proposed approach can solve many challenging structures that were not solvable with the widely used crystallographic software package Shake-and-Bake. This past semester, Alex completed a computational implementation of our algorithms for incorporation within the Shake-and-Bake system by our collaborators at the Hauptman-Woodward Medical Research Institute. Our algorithms have replaced the default algorithms in the Shake-and-Bake code for the solution of centrosymmetric crystal structures. This is a significant advance, given that the Shake-and-Bake code represents the state-of-the-art in crystallographic computing and that it had not been changed since the mid 1990s.

Protein Side-chain Conformation Problem

Student: **Wei Xie (Ph.D. May 2007, Chemical and Biomolecular Engineering at the University of Illinois; currently at American Airlines Operations Research and Decision Support Group, Fort Worth, Texas)**

The protein side-chain conformation problem is a central problem in proteomics with wide applications in protein structure prediction and design. The problem calls for determining the side-chain orientations for a protein whose backbone structure is already known. Computational complexity results show that the problem is hard to solve. Yet, instances from realistic applications are large and demand fast and reliable algorithms. **Wei Xie** has developed a new global optimization algorithm which for the first time integrates residue reduction and rotamer reduction techniques previously developed for the protein side-chain conformation problem. We showed that the proposed approach simplifies dramatically the topology of the underlining residue graph. Computations show that our algorithm solves problems using only 1 to 10% of the time required by the mixed-integer linear programming approach available in the literature. In addition, on a set of hard side-chain conformation problems, our algorithm runs 2 to 78 times faster than SCWRL

3.0, which is widely used for solving these problems. Our algorithm currently represents the state-of-the-art algorithm for the protein side-chain problem.

Protein Structural Alignment

Students:

Shweta Shah (PhD. Student at CMU)

Wei Xie (Ph.D. May 2007, Chemical and Biomolecular Engineering at the University of Illinois; currently at American Airlines Operations Research and Decision Support Group, Fort Worth, Texas)

Aligning proteins based on their structural (3D) similarity is a fundamental problem in molecular biology with applications in many settings, including structure classification, database search, function prediction, and assessment of folding prediction methods. Structural alignment can be done via several methods, including contact map overlap (CMO) maximization that aligns proteins in a way that maximizes the number of common residue contacts. **Wei Xie** has developed a reduction-based exact algorithm for the CMO problem. Our approach solves CMO directly rather than after transformation to other combinatorial optimization problems. We exploit the mathematical structure of the problem in order to develop a number of efficient lower bounding, upper bounding, and reduction schemes. Computational experiments demonstrate that our algorithm runs significantly faster than existing exact algorithms and solves some hard CMO instances that were not solved in the past. In addition, the algorithm produces protein clusters that are in excellent agreement with the SCOP classification. An implementation of our algorithm is accessible as an on-line server at <http://eudoxus.scs.uiuc.edu/cmoc/cmoc.html>. This algorithm currently represents the state-of-the-art algorithm for the CMO problem.

Shweta Shah, who joined the group in November, is beginning to work on this problem. Her first task is to produce an improved on-line server implementing the algorithm developed earlier by **Wei Xie**.

Combinatorial Library Design

Student: **Wei Xie (Ph.D. May 2007, Chemical and Biomolecular Engineering at the University of Illinois; currently at American Airlines Operations Research and Decision Support Group, Fort Worth, Texas)**

Designing proteins with novel or improved functions via combinatorial libraries involves first mutating or recombining existing sequences and then exploring the resultant sequences. Compared to the traditional sequence design approaches that require detailed knowledge of protein structure and function, combinatorial library design approaches are more straightforward to implement and prove quite successful in practice. An indispensable component of a successful library design approach is a powerful computational method that reaches balance between diversity and activity. Although quite a few computational studies have been conducted in the past, limited theoretic analysis exists that unveils the inherent difficulty of the proposed models, which would provide important guidelines for the future development of models and algorithms for this problem. **Wei Xie** studied this problem from a computational complexity perspective and reached several interesting results. We showed that existing models for combinatorial library design vary substantially in difficulty: while some models are fairly easy in that they admit low order polynomial-time algorithms, others may demand exponential time to solve. In addition, we propose several new algorithms for combinatorial library design under sequence-independent site-directed chimeragenesis (SISDC) protocol, and the resultant implementation outruns popular RASPP package by a factor of 10 to 100.

Protein-Ligand Docking

Student: Saurabh Awasthi (M.S. student at CMU)
Collaborator: Professor Mike Domach (Chemical Engineering, CMU)

A central problem in drug design is to identify small molecules that can bind on certain active sites of a protein and deactivate the corresponding function of the protein. Key to solving this problem is the ability to identify the site on a given protein where a given molecule is most likely to bind. This is frequently accomplished through the minimization of binding free energy. **Saurabh Awasthi**, who joined the group in November, is beginning to study this problem.

Steady State Optimization with Guaranteed Robust Stability under Parametric Uncertainty

Student: YoungJung Chang (Ph.D. December 2006, Chemical and Biomolecular Engineering at the University of Illinois; currently postdoc with Costas Maranas at Penn State)

Recently, Chang and Sahinidis (2005) proposed a global optimization method to find robustly stable steady-state solutions to biochemical process optimization problems using analytical necessary and sufficient stability conditions as constraints. This approach aimed to address structural and general model uncertainties and could produce conservative solutions for the case of parametric uncertainties alone. In the current work, we show how to extend the approach developed in Chang and Sahinidis (2005) to handle parametric uncertainties in a way that does not lead to conservative designs. The methodology proposed relies on an analytical approach to enforce robust stability in steady-state optimization problems. The resultant formulation is intuitive and exact, and a global optimization algorithm is developed to solve it. In contrast to prior approaches to this problem, our methodology requires no bifurcation analysis of the system while providing a deterministic guarantee for the least conservative stable solution.

DNA Sequencing by Hybridization

Student: YoungJung Chang (Ph.D. December 2006, Chemical and Biomolecular Engineering at the University of Illinois; currently postdoc with Costas Maranas at Penn State)

DNA sequencing by hybridization (SBH) uses DNA chips to reconstruct a DNA sequence from subsequences. SBH is often used in custom re-sequencing and mutation detection. If errors are present in the experimental data, the reconstruction problem is NP-hard and, thus, computationally challenging. This problem has been studied extensively but no current approaches provide global optima, alternative solutions, or a guarantee for the correctness of solutions. With **YoungJung Chang**, we formulated the DNA SBH problem with errors as an integer linear program with an exponential number of constraints. A row generation solution algorithm was developed to solve this model. The proposed approach solved large SBH problems to global optimality efficiently and was able to locate all the alternative optimal solutions. These alternative solutions exhibited a wide range in terms of their ability to reproduce the target DNA sequence.

Protein Binding Site Identification

Student: YoungJung Chang (Ph.D. December 2006, Chemical and Biomolecular Engineering at the University of Illinois; currently postdoc with Costas Maranas at Penn State)

In this work with **YoungJung Chang**, the problem of simultaneous identification of a protein binding pattern and sites has been solved within an integer programming framework. Instead of using a probabilistic model, a difference model was used to make the problem more tractable, and an exact efficient linearization scheme was proposed for the resulting nonlinear model. Depending on the amount of available biological information, we could employ two different formulations. If the pattern shape is known, then we can solve an easier problem, for which the state-of-the-art integer programming solver is

very efficient. Even when there is limited prior knowledge on the pattern such as its maximal and minimal lengths, the proposed formulation can solve the problem within a reasonable amount of time..

A Branch-and-Bound Algorithm for the Continuous Facility Layout Problem

Student: Wei Xie (Ph.D. May 2007, Chemical and Biomolecular Engineering at the University of Illinois; currently at American Airlines Operations Research and Decision Support Group, Fort Worth, Texas)

Finding optimal facility layouts is a classic problem in Process System Engineering as well as Operations Research. As initial models were unsuitable for instances of unequal facility sizes or unknown potential positions, continuous facility layout (CFL) models were introduced to address these limitations by modeling facilities as geometric entities and searching for an optimal 2-dimensional packing. However, solving these new models becomes dramatically harder: finding optimal layouts for these models is beyond reach of current optimization techniques, except for tiny instances. **Wei Xie** obtained several important theoretical results for this problem. First, we proved that it suffices to enumerate finitely many candidate solutions to secure an optimal solution, despite the fact that CFL admits infinitely many feasible layouts. We then developed a specialized branch-and-bound algorithm to further boost search efficiency by exploiting problem structure to prune large portions of the solution space. Comprehensive computational studies show that this new algorithm substantially outperforms three existing approaches. We also discussed extensions of this algorithm to more general layout instances.

Portfolio Optimization for Wealth-dependent Risk Preferences

Student: Mr. Luis Miguel Rios (Ph.D. student in Industrial Engineering at the University of Illinois)

Empirical and theoretical studies of preference structures of investors have long shown that personal and corporate utility is typically multimodal, implying that the same investor can be risk-averse at certain levels of wealth while risk-seeking at others. **Luis Miguel Rios** has studied the problem of optimizing the portfolio of an investor with an indefinite quadratic utility function. The convex and concave segments of this utility reflect the investor's attitude towards risk, which changes based on deviations from a fixed goal. Uncertainty is modeled via a finite set of scenarios for the returns of securities. A global optimization approach based on BARON was developed to solve the proposed nonconvex optimization problem. We performed computations in order to investigate the effect of short sales and demonstrate that the proposed approach systematically produces portfolios with higher values of skewness than the classical expectation-variance approach.

Medical Diagnosis and Prognosis

Collaborator: Professor Hong Ryoo (Associate Professor, Department of Industrial and Information Engineering, Korea University, Seoul Korea)

Over the past two decades, diagnostic techniques have grown out of the desire to replace surgical biopsy by breast cancer diagnosis that is based solely on the use of samples obtained through fine needle aspirates (FNA). Once FNA samples are taken from the breast mass, the material is examined for a number of characteristics of each nuclei, including size, shape, and texture. These attributes can then be used to classify the sample as benign or malignant. The fundamental question is how to perform this last step of differentiating between benign and malignant samples. One approach is to use FNA data from hundreds of patients that were surgically diagnosed and learn from these how to diagnose future patients based on FNA samples alone. The underlying mathematical problem calls for the development of a discriminant function

that separates two sets of points in a high dimensional space. This is a problem that arises in data analysis and pattern recognition problems in many domains, including financial modeling and investment planning, behavioral modeling, and data-driven managerial decision making. Except for certain special cases that are easily solvable, this problem is known to be challenging. With **Hong Ryoo**, we approached the problem based on the hypotheses that it suffices to develop the best possible pair of hyperplanes for separating the experimental data. We then devised a systematic optimization methodology for solving this problem. On a set of approximately 600 clinical FNA samples from the University of Wisconsin Hospital, our methodology yielded 99% correct breast cancer diagnosis. Compared to the 95% accuracy of the best previous techniques, our developments would imply 24 fewer misdiagnosed patients for this small sample alone. This is an important improvement given that millions of patients undergo breast cancer diagnosis every year. A paper on this subject is nearing completion. We plan to extend our methodology to diagnosis of other types of medical conditions as well as prognosis of the long-term behavior of the disease.

Design of Novel, Environmentally Benign Chemicals

Students:

Apurva Samurda (M.S. student at CMU)

Currently seeking postdoctoral researcher (synthetic chemist)

Due to concerns regarding the depletion of the ozone layer by chlorofluorocarbon (CFC) refrigerants, extensive research efforts have been directed towards finding environmentally benign CFC replacements. In the mid 1980s, it was proposed that the search for appropriate Freon and other CFC replacements can be approached computationally by using group contribution techniques to predict properties of potentially new compounds. The main question is then how to efficiently search the astronomically large space of all potential group combinations in order to identify the most promising refrigerants.

In work funded by an NSF/Lucent Technologies Industrial Ecology Fellowship, we developed an optimization model and a systematic solution methodology for the Freon replacement search problem. Methyl chloride and other CFCs used as refrigerants in the past turned out to be some of the solutions of our model. Furthermore, our model was able to identify several *novel* potential alternative refrigerants (compounds that do not appear in chemical databases as well as some that appear but have never been used as refrigerants.) This is the first theoretical approach to propose novel potential refrigerants after finding the *complete* set of solutions to a Freon replacement problem that was open for the last 20 years.

Recently, we extended our approach to the problem of identifying replacements for refrigerants currently used in retail food refrigeration. With funding from the Environmental Protection Agency and the National Science Foundation, we identified over 3000 potential new secondary refrigerants. A paper on this subject is nearing completion. We currently plan to seek to hire a synthetic chemist as a postdoc in order to synthesize and test a carefully select subset of 12 of these compounds.

In the long term, we plan to develop similar computational and experimental techniques for the design of fire suppressants, industrial solvents, polymers, drilling fluids, and drugs with an emphasis on minimizing the environmental impact over the entire life cycle of the new compounds.

Apurva Samurda, who joined the group this past November, has begun to study the problem of designing secondary refrigerants. The plan is to use a model earlier developed by Nanda and Sahinidis in order to identify potential new secondary refrigerants.

Erik Ydstie's Group

Real Time Optimization and Distributed Adaptive Control

Eduardo J. Dozal-Mejorada (Ph.D.)

Wayo (Eduardo) has developed new stability results for adaptive control algorithms, The results are being prepared for publication in Automatica and IEEE Transaction on Control. The method can be applied for an online optimization technique for unconstrained optimization based on extremum seeking control. The optimization algorithm behaves as a stand-alone module using existing process models to gather information and adaptively drives the process to its extremum without altering the simulation code. The optimizer uses a second order approximation to a high order system and takes a set of modified equality constraints constructed so that they have the same geometric characteristics of the true steady-state behavior of the process. The module employs a dynamic model to obtain a steady-state model of the system. The steady-state model is then used to find the optimal input to the process. The optimum point obtained by the optimizer is tracked throughout process operation. The advantages of this adaptive algorithm over other traditional optimization approaches include: 1) it uses current process models, 2) it can handle multiple objective functions, 3) it only applies a subset of all process states, 4) it self-learning, and, 5) it is appropriate for steady-state or dynamic models. We have developed theoretical convergence and stability results for the exact-matching case and have started working on the robustness problem. Successful implementation has been applied to two case studies: 1) optimization of biomass production in a bioreactor, and, 2) optimization of silicon production in a reactor.

Passivity Based Control of Multi-Phase Reactor Systems

Student: Yuan Xu (Ph.D.)

Yuan has developed results to demonstrate uniqueness of the multi-component-multi-phase flash. He is in the process of extending these results to the reactive flash. He also develops stability theory. The focus of this research is on passivity-based control of multiphase systems with reaction transport (diffusion and convection). To study this type of system, a metallurgic process has been chosen involving the carbothermic reduction of aluminum. This process takes carbon and aluminum oxide feed to produce aluminum. Work thus far has included developing a model using MATLAB to describe a variation of this process, as is being created by ALCOA, the world's largest aluminum producer. Proportional control is currently being used to explore the controllability, stability, and sensitivity of the system. Future work will be directed toward interfacing the FACT thermodynamics program with MATLAB via a MEX Interface. This will allow for the model to handle ranges of temperatures and compositions without extra work needed to determine equilibrium constants, activity coefficients, and extent of reactions. Another future improvement will be to incorporate computational blocks using the SIMULINK tools in MATLAB. Not only will this give the model a more "user-friendly" interface, but it will also allow for the exploration of modular simulation, where numerous monolithic simulation blocks can be combined and individually interchanged or altered without major recoding needed to modify the overall model.

Modeling the Vapor Recovery Reactor in Carbothermic Aluminum Production using Multi-Scale Modeling Methods

Student: Mohit Aggarwal (co-supervised with Prof Lee White) (Ph.D.)

Mohit has worked on modeling chemical process systems using invariant groups, He has made good progress in the development of models for the gasifier in the IGCC process and the vapor recovery section of the carbothermic aluminum process under development by ALCOA. In this process aluminum is

produced in a high temperature (2000C) two stage process. A significant amount of aluminum vaporizes from the aluminum reactor and this aluminum must be recovered in a vapor recovery column in order to maximize the production yield and save energy. We will model the primary reaction in the column, the gas flows as well as solid transport. The focus is to develop efficient methods to represent the thermodynamics and the multiphase behavior of the VRR. This work is a continuation of Vianey's modeling work (PhD/PostDoc 2004). Mohit will develop a three phase (liquid/solid/gas) multi-scale, simulation model which will be used for system design, scale-up, economic evaluation and control studies.

Process Networks

Student: Michael Wartman (PhD)

Michael developed a process analog of the Tellegen theorem of electrical network theory. He introduced a novel modeling framework for studying dynamics, distributed control, and optimization of complex chemical process networks. We are interested in developing distributed control systems that self-organize so that stability and optimality follows as a consequence of how the networks are put together and how they are connected with boundary conditions or other networks. By considering only the topology of the system, basic conservation principles and the second law of thermodynamics we have developed a multi-component analog to Tellegen's Theorem of electrical circuit theory. This result has been combined with the passivity theory of nonlinear control. It is shown that under certain convexity conditions the network converges to a stationary solution. Under similar conditions, it is shown that the network is self-optimizing in that the entropy production is minimized.

Multi-Scale Modeling of Particulate Processes with Fluid Flow

Student: Juan Du

This research addresses modeling and control of yield and size distribution in particulate processes. We use a novel method for solar-grade silicon production, the thermal decomposition of silane gas in a fluidized bed reactor, as a benchmark. We have developed a model and control scheme for the behavior of particles in a fluidized bed reactor based on silicon production. The model simulates growth of silicon particles with ordinary differential and algebraic equations that track particle movement through discrete size intervals. The model solves quickly and is easily tuned to fit experimental data. The passivity-based inventory controller maintains a constant mass of a specified size of silicon particles. The results from this research will be published in Powder Technology.

Last year we have matched the model to pilot-scale process data obtained from Solar Grade Silicon LLC in Moses Lake, Washington. We developed an observer for on-line estimation of process states and parameters. The method allows for estimation of particle size distribution in real time. We have combined the observer based modeling scheme with inventory control for total mass and seed-mass control for on-line stabilization of the particle size distribution. The model and control scheme is being used for scale-up and control system design for production system under current development at SGS. It is expected that when the new process is in operation that it will provide several competitive advantages relative to the current Siemens process, including higher throughput per process unit, lower capital cost and significantly lower energy usage.

PUBLICATIONS:

B-08-01

V. M. Zavala, C. D. Laird and L. T. Biegler, "A Fast Computational Framework for Large-scale Moving Horizon Estimation," submitted for publication (2007). <http://dynopt.cheme.cmu.edu/papers/papers.html>

B-08-02

Sager, S., Y. Kawajiri and L. T. Biegler, "On the Optimality of Superstructures for Simulated Moving Beds: Is One Pump Sufficient for Each Stream," submitted for publication (2007). <http://dynopt.cheme.cmu.edu/papers/papers.html>

B-08-03

Agarwal, A., L. T. Biegler and S. E. Zitney, "Simulation and Optimization of Pressure Swing Adsorption Systems using Reduced-Order Modeling," submitted for publication (2007). <http://dynopt.cheme.cmu.edu/papers/papers.html>

B-08-04

Zavala, V., M., C. D. Laird and L. T. Biegler, "A Fast Computational Framework for Large-Scale Moving Horizon Estimation," submitted for publication (2007). <http://dynopt.cheme.cmu.edu/papers/papers.html>

B-08-05

Jain, P., L. T. Biegler, and M. S. Jhon, "Parametric Study and Estimation in CFD-based PEM Fuel Cell Models," submitted for publication (2007). <http://dynopt.cheme.cmu.edu/papers/papers.html>

B-08-06

U. Kaplan, M. Turkay, B. Karasozen, L. T. Biegler, "Optimization of Supply Chain Systems with Price Elasticity of Demand," submitted for publication (2007). <http://dynopt.cheme.cmu.edu/papers/papers.html>

B-08-07

Flores Tlacuahuac, A. and L. T. Biegler, "Integrated Control and Process Design during Optimal Polymer Grade Transition Operations," *Computers and Chemical Engineering*, accepted for publication (2007). <http://dynopt.cheme.cmu.edu/papers/papers.html>

B-08-08

Flores Tlacuahuac, A., S. Terrazas Moreno, and L. T. Biegler, "On Global Optimization of Highly Nonlinear Dynamic Systems," *I & EC Research*, accepted for publication (2007). <http://dynopt.cheme.cmu.edu/papers/papers.html>

G-08-01

Caballero, J.A. and I.E. Grossmann, "An Algorithm for the Use of Surrogate Models in Modular Flowsheet Optimization," submitted for publication (2007). http://egon.cheme.cmu.edu/Papers/paper_caballero_grossmannkirking.pdf

G-08-02

Erdirik-Dogan, M., I.E. Grossmann and J. Wassick, "Short-Term Scheduling of Batch Plants with Parallel Reactors Forming Mobile Work Groups," submitted for publication (2007). <http://egon.cheme.cmu.edu/Papers/Workgroupsfinishinscheduling.pdf>

G-08-03

Karuppiyah, R. K.C. Furman and I.E. Grossmann, "Global Optimization for Scheduling Refinery Crude Oil Operations," to appear in *Computers & Chemical Engineering* (2007). <http://egon.cheme.cmu.edu/Papers/CrudeSchedRamkumarKevin.pdf>

G-08-04

Terrazas-Moreno, S., A. Flores-Tlacuahuac and I.E. Grossmann, "Simultaneous Design, Scheduling, and Optimal Control of a Methyl-Methacrylate Continuous Polymerization Reactor," submitted for publication. <http://egon.cheme.cmu.edu/Papers/scdc-v2.pdf>

G-08-05

You, F. and I.E. Grossmann, "Design of Responsive Process Supply Chains under Demand Uncertainty," submitted for publication (2007). http://egon.cheme.cmu.edu/Papers/Responsive_Supply_Chain_Uncertainty_You_Grossmann.pdf

S-08-01

Bao, X. and N. V. Sahinidis, "Finite Algorithms for Global Minimization of Separable Concave Programs," in T. Coleman, P. Pardalos and S. Vavasis (eds.), *Global Optimization*, Fields Institute Communications, American Mathematical Society, submitted.

REPRINTS

Biegler, L. T., "Large-scale Nonlinear Programming: An Integrating Framework for Enterprise-Wide Dynamic Optimization," *Computer Aided Chemical Engineering*, 24, V. Plesu and P. S. Agachi (eds.), pp. 575-582, Elsevier, Amsterdam (2007).

Biegler, L. T., "Optimization" pp. 3-60 – 3-70 in Perry's Chemical Engineer's Handbook, 8th edition, D. W. Green (ed.), McGraw-Hill, New York (2008).

Biegler, L. T., "A Survey of Simultaneous Dynamic Optimization Strategies," *Chemical Engineering and Processing*, 46, pp. 1043-1053 (2007).

Chang, Y. and N. V. Sahinidis, Global Optimization in Stabilizing Controller Design, *Journal of Global Optimization*, 38(4), 509-526, (2007).

Caballero, J.A., A. Odjo and I. E. Grossmann, "Flowsheet Optimization with Implicit Models and Complex Cost and Size Functions using Chemical Process Simulators," *AIChE J.*, 53, 2351-2366 (2007).

Castro, P.M., M. Erdirik-Dogan and I.E. Grossmann, "Simultaneous Batching and Scheduling of Single Stage Batch Plants with Parallel Units," *AIChE J.*, 54, 183-193, (2008).

Erdirik-Dogan, M. and I.E. Grossmann, "Planning Models for Parallel Batch Reactors with Sequence-dependent Changeovers," *AIChE J.*, 53, 2284-2300 (2007).

Jussi Hakanen, Yoshiaki Kawajiri, Kaisa Miettinen and Lorenz T. Biegler, "Interactive Multi-Objective Optimization for Simulated Moving Bed Processes," *Control and Cybernetics*, 36(2), pp. 283-302 (2007).

Karuppiah, R. and I.E. Grossmann, "Global Optimization of Multiscenario Mixed Integer Nonlinear Programming Models arising in the Synthesis of Integrated Water Networks under Uncertainty," *Computers and Chemical Engineering*, 32, 145-160 (2008).

Naraharisetti, P. K., B. Y. S. Ong, J. W. Xie, T. K. Y. Lee, C.-H. Wang, and N. V. Sahinidis, In Vivo Performance of Implantable Biodegradable Preparations Delivering Paclitaxel and Etanidazole for the Treatment of Glioma, *Biomaterials*, 28(5), 886-894, (2007).

Terrazas-Moreno, S., A. Flores-Tlacuahuac, I.E. Grossmann, "Lagrangean Heuristic for the Scheduling and Control of Polymerization Reactors," *AIChE J.*, 54, 63-182, (2008).

Yoshiaki Kawajiri and Lorenz T. Biegler, "Large Scale Optimization Strategies for Zone Configuration of Simulated Moving Beds," *Computers and Chemical Engineering*, 32, pp. 135-144 (2008).

You, F. and I.E. Grossmann, "Chapter 3: Optimal Design and Operational Planning of Responsive Process Supply Chains," In *Supply Chain Optimization*, (Eds. M. Georgiadis and L. Papageorgiou), pp. 107-134. Volume 4: Part II, Wiley-VCH Verlag. (2007).