Implementation of Model Predictive Control for Glucose Regulation on a General Purpose Microprocessor

Leonidas G. Bleris and Mayuresh V. Kothare

Abstract— In order to effectively control nonlinear and multivariable models, and to incorporate constraints on system states, inputs and outputs (bounds, rate of change), a suitable (sometimes necessary) controller is Model Predictive Control (MPC), also known as receding horizon control and moving horizon control. For processes with slow dynamics and low sampling rates, MPC is typically implemented on a dedicated workstation. In this work we present the implementation of MPC on a general-purpose processor, providing a low-cost, low-power consumption and small in size implementation. In order to test the performance of the embedded MPC we examine a glucose regulation problem. Additionally, profiling results of the performance of the processor are provided.

I. INTRODUCTION

Model Predictive Control originated in the chemical process industries mainly due to the need for a controller capable of handling constrained and multivariable nonlinear processes. Because of the computational requirements of the optimizations associated with MPC, it has primarily been applied to processes with slow dynamics. Existing implementations of MPC typically perform numerical calculations using workstations (currently) in 64-bit Floating Point (FP) arithmetic, which is too expensive, power demanding and large in size. For systems with fast dynamics, where the size and the application precludes the use of a dedicated workstation, alternative solutions have to be investigated. Thus, there is an increasing need for reducing MPC on a chip. Some approaches to this direction have been reported in literature [1], [2], [3].

We have proposed in [4], [5] an efficient implementation of MPC by reducing the precision of the microprocessor to the minimum, while maintaining optimal control performance. By reducing the precision we increase the optimization speed, and at the same time we minimize the power consumption and the overall chip area (thereby the cost). This reduced precision application-specific processor can achieve sampling speeds as low as 0.032 seconds for relatively large problems [5]. To quantify the advantage of reducing the precision, estimations for both 64-bit FP and 16-bit LNS circuits show that for an arithmetic unit that computes addition, subtraction, multiplication and division, the size required is about 17 times smaller than a 64-bit floating point processor. The advantage of formulating the on-line MPC problem in this fashion is that it directly extends to other objective functions, as opposed to the MP approach which applies to QPs [2], [3].

In this paper we provide the results of the implementation of MPC on a general-purpose processor that provides a low-power, small in size and relatively inexpensive solution. We examine the performance of MPC using processor in the loop co-simulations; the optimizations associated with MPC are carried out on the general processor and control is applied to a model that runs on a host PC. A case study of glucose regulation for diabetic patients is examined. Additionally the performance of MPC running on this board is profiled and the results are presented.

The paper is organized as follows: Section II contains the main theoretical aspects of MPC and information on the optimization associated with the MPC operations. In Section III we provide the details of the hardware implementation of MPC, information about the profiling and the co-simulations. The glucose regulation case study is given in Section IV. We conclude the paper with remarks on our research results and an analysis of issues that are currently under investigation.

II. THEORETICAL BACKGROUND

Controllers belonging to the MPC family are generally characterized by the following steps [6] (The schematic of the overall system is given in Figure 1):

1. Initially the future outputs are calculated at each sample interval over a predetermined horizon $N$, the prediction horizon, using the process model. These outputs $y(t + k|t)$ for $k = 1 \ldots N$ depend up to the time $t$ on the past inputs and on the future signals $u(t + k|t)$, $k = 0 \ldots N - 1$ which are those to be sent to the system.

2. The next step is to calculate the set of future control moves by optimizing a determined criterion in order to keep the process as close as possible to a predefined reference trajectory. This criterion is usually a quadratic function of the difference between the predicted output signal and

Fig. 1. Block diagram of MPC.

$y_{ref}$

Model Predictive Control

$u(t+k|t)$

System

$y(t)$

$u(t)$
the reference trajectory. In some cases the control moves 
\( u(t + k|t) \) are included in the objective function in order
to minimize the control effort:

\[
J_P(k) = \sum_{k=0}^{P} \left\{ |y(t + k|t) - y_{ref}|^2 + Ru(t + k|t)^2 \right\} \quad (1)
\]

\[|u(t + k|t)| \leq b, \quad k \geq 0 \quad (2)\]

where \( y(t + k|t) \) are the predicted outputs, \( y_{ref} \) is the desired
set reference output, \( u(t + k|t) \) the control sequence and \( R \) is
the weighting on the control moves, a design parameter. This
system is subject to input constraints given by the vector \( b \).

3. Finally, the first control move \( u(t|t) \) is sent to the
system while the rest are rejected. This is because at the
next sampling instant the output \( y(t + 1) \) is measured by
the system and the procedure is repeated with the new values so
that we get an updated control sequence.

A. Optimization

The optimization is a fundamental part of MPC since it
results in optimal control inputs for the process. The
computational effort required in MPC derives almost entirely
from the optimization algorithm. Logically the choice of the
optimization technique is decisive to the performance of the
controller [7]. Typically, quadratic functions are used as cost
functions in MPC. Equation 1 can be simplified and rewritten as:

\[
\begin{align*}
\text{minimize} & \quad q(x) = \frac{1}{2} x^T G x + g^T x \\
\text{subject to} & \quad a_i^T x = b_i, \quad i \in E \quad (3) \\
& \quad a_i^T x > b_i, \quad i \in I \quad (4) \\
& \quad |x(t + k|t)| \leq b, \quad k \geq 0 \quad (5) 
\end{align*}
\]

where \( x \) is the vector of variables (corresponds to the
actuation \( u \)), and the constraints have been generalized and
divided into two sets, \( E \) or equality constraints and \( I \) or
inequality constraints. Any constant term (not depending on \( x \))
appearing after the simplification of equation 1 can be ig-
nored since it will not affect the solution of the minimization
problem.

We use Newton’s algorithm to solve problem by com-
bining the constraints into the cost function, using barrier
functions for the inequality constraints, and penalty functions
for equality constraints, defined as:

\[
\begin{align*}
d_i(x) &= \mu_i (a_i^T x - b_i)^2, \quad i \in E \quad (6) \\
d_i(x) &= \mu_i \log(a_i^T x - b_i), \quad i \in I \quad (7) 
\end{align*}
\]

resulting in the unconstrained non-linear problem:

\[
\begin{align*}
\text{minimize} & \quad f(x) = \frac{1}{2} x^T G x + g^T x + \sum_i d_i(x) \\
\end{align*}
\]

The problem of equation 8 can be solved numerically approx-
imating \( f(x) \) by a quadratic function around \( x \), obtaining the
gradient \( \nabla f(x) \) and Hessian \( H(x) \), and iterating:

\[
x^{(k+1)} = x^{(k)} - H^{-1}(x^{(k)}) \cdot \nabla f(x^{(k)}) \quad (9) 
\]

The complexity of this algorithm is dominated by the com-
putation of \( H^{-1} \), which requires \( O(n^3) \) operations, where \( n \)
is the number of variables to optimize.

III. HARDWARE IMPLEMENTATION OF MPC

For the real-time implementation of model predictive
control we use the high-performance single board computer
phyCORE-MPC555, illustrated in Figure 2. This board packs
the power of Motorola’s embedded 32-bit MPC555 micro-
controller within a miniature footprint. The MPC555 is a
high-speed 32-bit Central Processing Unit that contains a 64-
bit floating point unit designed to accelerate the advanced
algorithms necessary to support complex applications. All
signals and ports of the MPC555 extend to two Molex high
density (0.635 mm pitch) 160 pin header connectors. These
high density pins allow it to be plugged like a “big chip”
into user target hardware.

Fig. 2. PhyCORE-MPC555 board and MPC555 processor (2cm×2cm).

In order to implement the optimization required by MPC
on this target we use a combination of software tools: Code-
Warrior Integrated Development Environment (IDE) [8],
MATLAB, Real-Time Workshop and SIMULINK. By com-
bining state-of-the-art MATLAB/SIMULINK environment
with the robust CodeWarrior Development Studio, we were
able to compile a C++ implementation of optimization and
download the operations on the MPC555 processor. The size
of the code downloaded was on average 30KBytes. The
phyCORE-MPC555 can be populated with a maximum of 4
MB flow-through synchronous external BURST-SRAM for
DATA storage and a maximum of 4 MB Flash-ROM memory
at 0 Wait-States.

5163
A. Code profiling

A profiler can analyze the amount of time a program spends performing various tasks of the optimization. A profiler can help detect bottlenecks (time consuming routines that data passes through) and routines that are inordinately slow. Clearly a good profile of the runtime performance of code requires more information than a raw count. Most of the profilers perform statistical sampling of the runtime environment; these profilers are called passive or sampling profilers. A passive profiler divides the program being profiled into evenly-sized “buckets” in memory. It then samples the processor's program counter at regular intervals to determine which bucket the counter is in. The main advantage of a passive profiler is that it requires no modification to the program under observation. Also, passive profilers distribute the overhead that they incur evenly over time, allowing the post-processing steps to ignore it. On the other hand, they cannot sample too frequently or the sampling interrupt will overwhelm the program being sampled.

B. Co-simulation

In order to test the performance of MPC on the embedded target we use Processor-In-the-Loop (PIL) co-simulations. The board is connected to a host workstation and in PIL mode, a plant model runs in non-real-time on the host workstation in Simulink. Meanwhile, generated code running on the MPC555 exchanges signals via RS232 serial communication with the Simulink simulation running on the workstation. At each sample interval, Simulink performs model updates and sends output signal data via RS232 to the MPC555. During PIL testing, one can include monitoring probes in the Simulink plant model. These probes may assist the code validation by displaying signal data. The signal data retrieved from the microcontroller is used for the next simulation step of the plant model.

IV. ACTIVE DRUG DELIVERY DEVICES

The development of novel medical sensors and drug delivery methods have introduced multiple benefits to medical devices. These include safety, efficacy, robustness, and patient compliance. For example, advances over the past few years in microfabrication technologies have allowed researchers to create microneedles [9], [10] that painlessly cross the uppermost layer of the skin to effectively, efficiently, and painlessly deliver drugs to depths as shallow as 1mm. Their sizes range from one millimeter to one micron in diameter, being too minute to stimulate nerve endings and cause pain. In parallel several other drug delivery approaches are under research and development [11], including non-invasive needle-free injection devices. The key to needle-free injection is to release the drug in liquid form at the proper (extremely high) velocity so that it diffuses through the skin. These drug delivery devices have the potential to maximize patient compliance, especially for diseases that require frequent drug doses.

Looking at the future, the end target is to develop small and compact drug delivery devices, that require minimal power and can maximize the therapeutic results of a drug. Noteworthy [12], for the former, integrated batteries can be used in order to provide enough power to an implant for five to eight years of life, without causing side effects. The latter can be accomplished by minimizing selected cost functions in a control procedure (the actual cost function of the optimization will depend upon the particular application and control objectives). Classic control algorithms (i.e. ON/OFF, PID, gain scheduling etc) can be used for drug-delivering medical devices. These algorithms can be implemented easily on chip, are robust, and have relatively low energy consumption. Nevertheless, the human organism is one of the most complex dynamical systems. Because of the inherent existence of nonlinearities, constraints, patient variability and delays the need for advanced control schemes becomes essential. The increased understanding of physiological, pharmacokinetic and pharmacodynamic models gives us the ability to apply advanced model based control schemes. In order to effectively control nonlinear and multivariable models, and to incorporate constraints in the manipulated variables (bounds, rate of change), a suitable controller is model predictive control.

A. Blood Glucose Control in Diabetic Patients

Diabetes mellitus is a chronic health condition where the human body is unable to produce insulin and properly breakdown sugar (glucose) in the blood. The insufficient insulin production or lack of responsiveness to insulin, results to hyperglycemia (high blood glucose levels, 120mg/dL). There are two primary types of diabetes mellitus, type I (insulin-dependent or juvenile-onset), which may be caused by an autoimmune response, and type II (non-insulin-dependent or adult-onset). Symptoms include hunger, thirst, excessive urination, dehydration and weight loss. Complications can include heart disease, stroke, neuropathy, poor circulation leading to loss of limbs, hearing impairment, vision problems and death.

The treatment of diabetes requires regular insulin injections, proper nutrition and exercise in order to maintain normoglycemia, defined as blood glucose 70–100mg/dL. Insulin and glucagon are the hormones responsible for glucose regulation. Both insulin and glucagon are secreted from the pancreas, and thus are referred to as pancreatic endocrine hormones. Most of the long-term complications associated with diabetes result from sustained hyperglycemia, but hyperglycemia can result in very acute symptoms such as coma and death.

A significant effort has been devoted to the development of closed-loop controllers for blood glucose control. In the following sections we provide the results of co-simulations, where we apply MPC to maintain normoglycemia. For approaches to applying control to diabetic subjects the reader is referred to [13], [14], [15].

B. Minimal Glucose Model

Minimal models of glucose and insulin plasma levels have been developed [16], [17], [18] for humans using frequently-
sampled intravenous glucose tolerance (FSIGT) tests. During a FSIGT test, a single intravenous injection of glucose is given to a fasting subject and blood samples are collected at regular timed intervals. The blood samples are then analyzed for glucose and insulin concentration. Figure 3 shows a typical response from a normal subject.

As illustrated in Figure 3, the glucose level in plasma starts at a peak due to the injection, drops to a minimum which is below the basal glucose level, and then gradually returns to the basal level (dashed line). The insulin level in plasma rapidly rises to a peak immediately after the injection, drops to a lower level which is still above the basal insulin level, rises again to a lesser peak, and then gradually drops to it's basal level. Depending on the state of the subject, there can be wide variations from this response that can determine the condition of the patient [17].

The glucose minimal model involves two physiologic compartments: an interstitial tissue compartment and a plasma compartment. Insulin leaves or enters the interstitial tissue compartment at a rate proportional to the difference between the plasma insulin level, I(t), and the basal level I_b (equation 11). If the plasma insulin level falls below the basal level, insulin leaves the interstitial tissue compartment, and if the plasma insulin level rises above the basal level, insulin enters the interstitial tissue compartment. Insulin also disappears from the interstitial tissue compartment via a second pathway at a rate proportional to the amount of insulin in the interstitial tissue. The differential equations corresponding to the above analysis are:

\[
\frac{dG(t)}{dt} = k_1(G_b - G(t)) - X(t)G(t) \quad (10)
\]

\[
\frac{dX(t)}{dt} = k_2(I(t) - I_b) - k_3X(t) \quad (11)
\]

where t is time, G(t) is the plasma glucose concentration at time t, I(t) is the plasma insulin concentration at time t, and X(t) is the interstitial insulin at time t, with G(t0)=G_0 and X(t0)=0. G_b is the basal plasma glucose concentration and I_b is the basal plasma insulin concentration. The insulin sensitivity is defined as \( S_f = k_2/k_3 \) and the glucose effectiveness is defined as \( S_G = k_1 \). Basal plasma concentrations of glucose and insulin are typically measured either before, or 180 minutes after, administration of glucose. There are four unknown parameters in this model: \( k_1, k_2, k_3, \) and \( G_0 \) that depend on the particular subject and can be estimated experimentally. We use the following parameters adopted from [18], [19]:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( G_b )</td>
<td>92 (mg/dL)</td>
</tr>
<tr>
<td>( I_b )</td>
<td>11 (µU/mL)</td>
</tr>
<tr>
<td>( G_0 )</td>
<td>92 (mg/dL)</td>
</tr>
<tr>
<td>( k_1 )</td>
<td>0.038291 (1/min)</td>
</tr>
<tr>
<td>( k_2 )</td>
<td>3.5364 × 10^{-9} (1/min)</td>
</tr>
<tr>
<td>( k_3 )</td>
<td>0.0015002 (1/min)</td>
</tr>
</tbody>
</table>

C. Multi-model MPC

In the minimal glucose model, glucose is utilized at the constant rate \( k_1 \), having the negative feedback effects of the interstitial insulin, represented by the term \(-X(t)G(t)\). An additional amount of plasma insulin will cause the amount of interstitial insulin \( X(t) \) to change, which will cause the rate of glucose utilization to change. In order to capture efficiently this effect we use a multi-model MPC approach. That is, we capture the glucose levels responses to pulse inputs of insulin for different initial conditions for the interstitial insulin. Under the assumption that the interstitial insulin is estimatable [20], [21] at each sampling interval, the predictive controller utilizes the proper model that best describes the dynamics of glucose-insulin in the human body.

As illustrated in Figure 4 we obtain the glucose level responses to a pulse input of 100(µU/ml) of insulin, for different initial conditions of interstitial insulin (from top starting at 0 with intervals of 0.005). For all these cases the glucose level returns to the basal level, slowly after reaching the minimum at 30min.

D. Co-simulation results

In order to test the performance of this multi-model MPC in real time we proceed with the processor-in-the-loop co-simulations. The setup is illustrated in Figure 5, where we have the MPC chip in closed loop with the monitored patient (the minimal model on the host workstation). The sampling time for the system is set at one second.
The simulation results for control horizon of 5, a prediction horizon of 31 and a fixed number of optimizations at 40, are given in Figure 6. The first from the top subplot represents the concentration of glucose in the blood, the second is the manipulated variable (insulin injection), and the third subplot is the glucose intake (due to two small meals). Due to the insulin injection we manage to keep the glucose concentration between the desired limits, avoiding hyperglycemia that would result in the absence of control and hypoglycemia that would result for overdosing insulin.

The profiling results, given in Figure 7, illustrate that the time required for the computation of the optimal insulin dosage from the predictive controller remains under one second for all examined control horizon cases, and prediction horizons of 21 and 31. Therefore, there is sufficient time for increasing the number of optimizations and adding fail-safe approaches that will guarantee an optimal, safe and consistent behavior by this embedded model predictive control implementation. In Figure 8 we provide the profiling results for variable number of iterations in the optimization, control horizon 7, and prediction horizons of 21 and 31. As expected the computational time grows linearly.

V. CONCLUDING REMARKS

A real-time implementation of model predictive control using a general purpose processor has been provided in this paper. Profiling results illustrate that the optimizations require less than one second for most cases. An implementation of real-time MPC on-a-chip for diabetes control, would additionally require the incorporation of a glucose estimator, some changes in the optimization code (so as to successfully deal with disturbances and uncertainty), the addition of fail-safe and scheduling routines, and finally the integration of the sensors and actuators. Therefore we should expect an increase in the computational time. Nevertheless, for the particular case of glucose regulation the sampling time is not a restriction, since most of the glucose sensors today operate with a sampling rate of 5 minutes.

Novel microfabrication technologies have allowed researchers to create minute, hollow needles that painlessly cross the uppermost layer of the skin to deliver drugs to the dermal layer. Some products under development can effectively, efficiently, and painlessly deliver drugs by bolus injection or infusion to depths as shallow as 1 mm. Coupled with additional advances in microelectronics some companies are developing disposable pumps for use with chronically administered drugs. The end target is to integrate a dedicated microprocessor for the application of real-time feedback control. There are numerous potential applications in this area: control of physiological processes, glucose control, muscle control, respiration control, drug infusion control, cardiac pacemakers and defibrillators, heart rate control, blood flow and pressure control, HIV control, and neurological implants. Drug-delivering closed-loop controllers can improve the effectiveness and safety of drugs, but also the quality of life of patients. A sensor can monitor the effect of the drug and through a control algorithm adjust the dosage to the optimum levels for effectiveness, reducing thereby side effects and potentially minimizing the costs. On the other hand, several small-scale industrial and consumer
intend to provide a practical co-design framework for the embedded controller for real-time control applications. We created with MPC. Thus, this is a mixed software-hardware (FPGA) that will accelerate the matrix operations associated by a specially customized Field Programmable Gate Array (FPGA) that will accelerate the matrix operations associated with MPC. Invited plenary lecture at the 7th IFAC Symposium on Dynamics and Control of Process Systems (DYCOPS-7), Boston, MA, July 2004.


