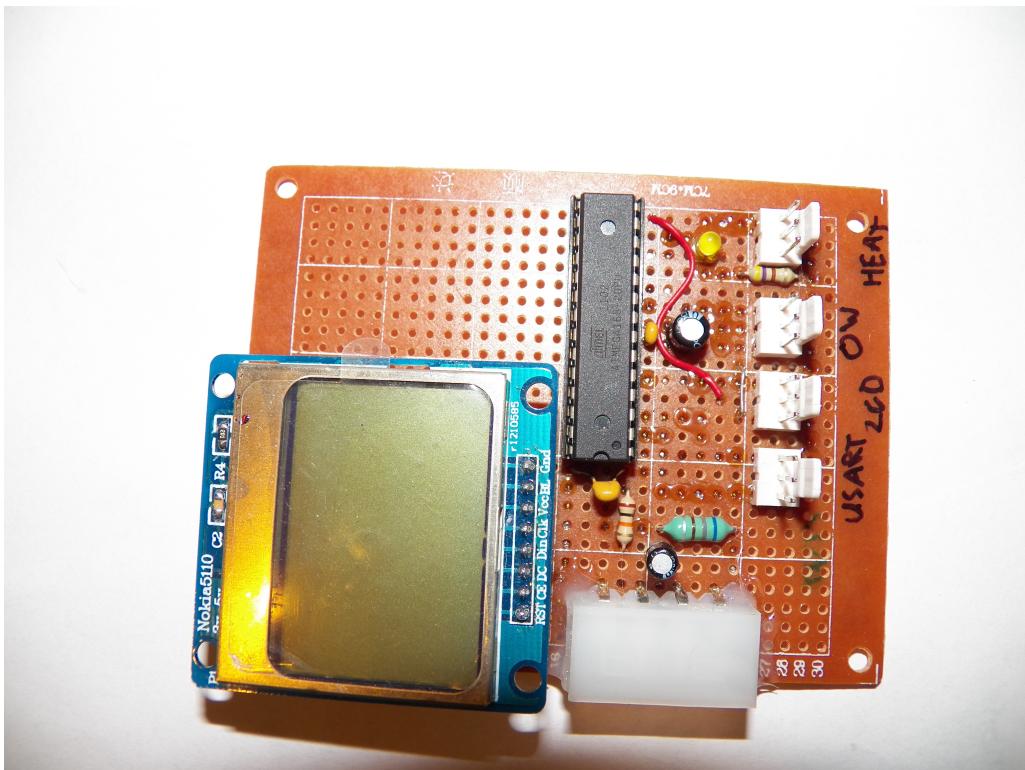


Multitasking Scriptable Autotuning PID Platform



January 21, 2016

Contents

1 Hardware	1
1.1 Theory	1
1.1.1 Bipolar transistor biasing	1
1.1.2 Phototransistor	1
1.2 Definitons	1
1.3 Goals	1
1.4 Layout	2
1.5 ZCD board	3
1.5.1 Schematic	3
1.5.2 Calculation	3
1.5.3 Measurements	4
1.6 Thermometer	6
1.6.1 Heater	6
1.6.2 Temperature sensor	6
1.7 SCR board	7
1.7.1 SCchematic	7
1.7.2 Calculation	7
1.7.3 Measurements	7
1.8 Main board	7
1.8.1 Microcontroller	7
1.8.2 Power filtering	8

1.8.3	Connectors	8
1.8.4	Extendability	8
2	Software	9
2.1	Theory	9
2.1.1	(9
2.1.2	Cooperative multitasking	9
2.2	Goals	10
2.3	Language selection	10
2.4	Bootloader	10
2.4.1	Motivation	10
2.4.2	Previous work	11
2.4.3	Implementation	11
2.4.4	Communication protocol	12
2.4.5	Error checking	12
2.4.6	Use	12
2.5	External components	13
2.5.1	Onewire library	13
2.5.2	PID implementation	13
2.6	Architecture	13
2.6.1	Theory	13
2.6.2	Implementation	14
2.7	Modules	14
2.7.1	Configuration	14

2.7.2	Clock	14
2.7.3	Serial communication	15
2.7.4	Commands	15
2.7.5	Zero-cross detector	15
2.7.6	Triac control	16
2.7.7	Temperature measurement	16
2.8	Plotting tool	16
3	Algorithms	17
3.1	Theory	17
3.1.1	Parallel PID controller	17
3.1.2	Anti-windup	17
3.1.3	Plant models	18
3.1.4	PID tuning algorithms	18
3.1.5	Digital PID controller	19
3.2	PID implementation	19
3.3	Units of measure	20
3.4	Nonlinearity	20
3.5	Ziegler–Nichols sustained oscillations	20
3.6	Astrom–Hagglund sustained oscillations	21
3.7	Sampling rate correction	21

1 Hardware

1.1 Theory

1.1.1 Bipolar transistor biasing

Saturation occurs when both PN junctions are forward-biased. In this mode the transistor is fully open, corresponding to logical "on" state. Cutting current to the base renders the device fully closed, or in logical "off" state.

1.1.2 Phototransistor

The phototransistor behaves similarly to an ordinary transistor with the exception that its base region is sensitive to light. When no light is falling onto the base, the emitter current is:

$$I_T = I_{CBO} + \beta I_{CBO} = (1 + \beta)I_{CBO} \quad (1)$$

If then light is shone onto the base region, the concentration of electrons and holes increases. A similar photo current, as with photodiodes, flows. The difference is that the current gets amplified $1 + \beta$ times compared to a photodiode.

Devices only two LEDs are available. This mode of operation is called 'floating base'. It is thermally unstable.

The alternative is to connect the (photosensitive) base to a suitable DC bias. In this configuration negative feedback is realized. As a result, the gain β decreases. The negative feedback provides further benefits beyond thermal stability.

1.2 Definitons

device – the entirety of all physical components, resulting from this project.

CTR – current transfer ratio, optocoupler characteristic.

1.3 Goals

The device is intended as a learning project for the student, but also as a open software, open-hardware project, which anyone can create and use. Thus, the following hardware design priorities have been identified, in order of decreasing importance:

- safety – the device shall not pose a fire or electric shock hazard to the end user
- reconstructability – the device shall be composed **only** of worldwide accessible components
- longevity – the device shall remain operational for 5 years of uninterrupted service with 95% confidence
- price – the BOM for the complete device shall not exceed 100BGN
- extendability – the number of input sensors and the number of output controllers, shall be trivially configurable
- ease of assembly – it shall be possible for a person with zero hardware experience to manufacture the device
- simplicity – each component shall fulfill a specific purpose, and the number of components shall be the lowest possible

1.4 Layout

Due to the requirement of extendability, the device shall consist of a number of printed circuit boards, in contrast to a single monolithic PCB. Each PCB shall fulfill a sole purpose, and any number of different modules shall be able to mate together. The following distinct roles have been identified:

- high-voltage input stage – called zero-cross detector or ZCD board from now on
- low-voltage input stage – called temperature sensor or thermometer from now on
- computational stage – called main board from now on
- high-voltage output stage – called software controlled rectifier board or SCR board from now on

The resulting design exhibits the following characteristics.

Only a single ZCD board is required, because mains waveform is invariant across the device in its entirety. Only a single main board is required, as the selected microcontroller, although inexpensive, provides plenty of resources for numerous control loops.

In order to satisfy the requirement for simplicity, the main board is configured for a single SCR output board. However, soldering additional connectors to the main PCB is trivial, thus achieving extensibility. The SCR output board is long-life and supports loads of up to 1kW.

The most flexible part of the system is the thermometer configuration. Due to the selected temperature sensing IC, virtually unlimited (technically up to 2^{56}) devices are supported **without any hardware changes**.

1.5 ZCD board

The ZCD board is a sensory input to the microcontroller.

Because the voltage of mains power is alternating, it is impossible to output precise amounts of power without knowing the phase of the waveform. The implementation of the ZCD board is straightforward and extremely simplified. In fact, an extensive internet search has demonstrated no other PCB has ever been designed with such a level of simplicity. In other words, **the designed PCB contains fewer elements than any known PCB for the same purpose!** This produces problems, which have deterred other designers. However, all artifacts have been dealt with in software.

1.5.1 Schematic

Please refer to appendix A1 for the schematic and layout of the board.

1.5.2 Calculation

In order to protect the main board (and thus the user) from dangerous voltages, galvanic isolation is required. The standard means to this end are transformers and optocouplers. Optocouplers possess numerous advantages over transformers for our application:

- compactness
- low price
- negligible phase shift

Furthermore, among optocouplers, the variation is considerable. We select a component with anti-parallel input LEDs, specifically designed for zero crossing – SFH620A-3. This is the most sensitive version of the IC (highest CTR), as input power is our greatest concern.

Striving for minimal component count and price, the standard solution with a 10W input power resistor is dismissed. Thus, we need to work with 1/4W, E24 resistors. Due to the optocoupler's acceptable CTR, and extensive signal conditioning in software, this solution will prove to be viable!

It is worthy to note that the resistor rated voltage is of utmost importance. Because our resistors are rated to 200V peak, it would be a dangerous mistake to use a single resistor. Therefore, the $V_{AC} = 230V$, $V_{peak} = V_{AC} * \sqrt{2} = 325V$ is safely spread onto two identical resistors.

Let's suppose the line voltage varies from $V_{min} = 200VAC$ to $V_{max} = 250VAC$ rms.

$$P_{inputresistors} = \frac{V_{max}^2}{R_1 + R_2}$$

$$R_1 + R_2 \geq \frac{V_{max}^2}{P_{inputresistors,max}} = \frac{250^2}{0.25 + 0.25} = 125\text{k}\Omega$$

We select $R_1 = R_2 = 68\text{k}\Omega$.

$$i_{in,min} = \frac{V_{min} - 1.65}{2 * R_1 * 1.05} = \frac{198.35V}{142.8\text{k}\Omega} = 1.39\text{mA}$$

The output stage:

$$i_C \geq i_{in,min} * CTR_{min} \approx 1.39\text{mA} * 0.34 \approx 0.47\text{mA}$$

$$i_{leakage} \leq 1\mu\text{A}$$

$$V_{IL} = 0.3Vcc = 0.3 * 5\text{V} = 1.5\text{V}$$

$$V_{IH} = 0.6Vcc = 0.3 * 5\text{V} = 3\text{V}$$

If we strive to be below 1V for logic zero:

$$i_C * R_{output} = 1V$$

$$R_{output} = 1\text{V}/0.47\text{mA} = 2.13\text{k}\Omega$$

We select $R_3 = 2.4\text{k}\Omega$.

1.5.3 Measurements

Firstly, a temperature measurement is performed. The device is allowed to run for 10 minutes. Subsequently, each component is measured for overheating. Because component temperature measurement instrumentation is both expensive and difficult to apply, the following rule of thumb is used: *if a silicone component is too hot to keep your finger on it, it is too hot*. Although the stated method is vastly imprecise, it works well, because the skin pain temperature (about 60°C) is far lower than silicone semiconductor Absolute Maximum Temperature (often 150°C).



We observe that:

1. The pulse is very wide – about 3.2ms \equiv 32% of the half-period.
2. The pulse is centered. This is great, because we can estimate the true zero crossing

in software.

1.6 Thermometer

It is impossible to examine the temperature sensing element in isolation to the heater. Therefore, in this chapter, the complete plant, or in other words the combination of heater, thermal mass and thermometer, will be examined.

As this is an educational project, the quickest possible system response is desired. Therefore, the thermometer is directly glued to the surface of the heater. Unfortunately, even in this setup, the maximum possible heating rate is about $6^{\circ}\text{C}/\text{min}$ and cooling is even slower.

1.6.1 Heater

Initially, a 60W incandescent light bulb was selected. As European wall power exhibits frequency of 50Hz, the highest switching frequency is 100Hz by half-periods. Astonishing to the experimenter:

- The filament is not inert enough to integrate consecutive pulses, even if every odd half-wave is enabled.
- The human eye is unable to integrate the resulting 50Hz flicker.

As a result, looking in the controlled bulb is extremely annoying.

Consequently, a fish tank heater with nominal (maximum) power of 50W was selected. The observed temperature curves are equivalent sans the maddening light flicker. **The thermometer is glued to the surface to the heater for fastest response possible.**

1.6.2 Temperature sensor

The Dallas Semiconductor DS18S20 has been selected due to a variety of reasons:

- low price
- ease of interfacing to digital components
- ease of wiring
- extreme flexibility of integration

This device is incredible. It can operate solely over two wires – including the ground wire. It performs digital temperature conversions, removing the need of an ADC (although our selected microcontroller features such). It can coexist on the same bus with as many as 2^{56} = infinite number of other onewire sensors.

1.7 SCR board

1.7.1 SChematic

1.7.2 Calculation

Because we want to operate with more precision than an on-off controller, the following requirements are layed out:

- Life must exceed 3153600000 switches.
- Switching times must be in the microsecond range.
- Galvanic isolation.

In the light of this sepcification, it immediately becomes obvious that an electro-mechanical relay is not suitable. Again an optocoupler has been selected. The MOC3023 is canonical for power controll applications. It is a triac output optocoupler, specifically designed to drive a power triac. The selected power triac is BT136 600E, providing control over applicances of up to $4A * 230V = 920W$.

1.7.3 Measurements

1.8 Main board

The main board houses the microcontroller, provides stable power to it and contains connector blocks to the other boards.

1.8.1 Microcontroller

The selected microcontroller is atmega168 from Atmel. It provides plenty of computing power and sufficient 16KB flash program memory at extremely modest cost. It can be programmed in assembler, C or C++ with widely available and free of charge tools. The

datasheet is well written and a plethora of application notes are available from Atmel. Not to mention the extensive worldwide community, readily providing help to anyone new to the subject.

1.8.2 Power filtering

Three groups of components are responsible for providing clean and steady power to the microcontroller.

Firstly, an LC filter at the power connector of the board provides filtering of the external power. This protects against insufficiently filtered power supplies and allows powering the board from a low cost wall adapter "brick". Furthermore, any EMI picked over a long power supply cord, is eliminated. Moreover, the LC filter isolates the board from the power supply, allowing the PSU to power other devices in the same time, free of digital switching noise.

Secondly, a combination of two capacitors, in close physical proximity to the microcontroller power leads, further conditions the power supply rail. As the microcontroller draws significant current for intervals far shorter than one microsecond, the PCB traces gain significant impedance. Consequently, this second group of capacitors needs to be located as close as possible to the IC. What's more, commodity electrolytic capacitors cannot operate at such high frequencies. Consequently, a ceramic capacitor is added, to handle the high frequency current pulses.

Lastly, the RESET pin of the controller is extremely sensitive to even short power line pulses (of the length of one clock cycle). Therefore, the exact RC circuit, recommended by Atmel, is used.

1.8.3 Connectors

The selected connector is NX5080-03SMR. This is a 3-pin, fixed orientation, low-current connector. The pin number is ideal for providing both power and a communication bus to connected boards. Because no terminal blocks are used, and all four connectors are wired in consistent fashion, incorrect wiring of the board is impossible.

1.8.4 Extendability

It has been paid attention to provide free PCB real estate, as well as conveniently located free processor pins. Consequently, it is trivial to add more connectors, or even a display, to the main board at a later moment of time.

2 Software

2.1 Theory

2.1.1 (

Radix tree) The radix tree is a data structure, suitable for storing and searching sets of strings. Each node, that is an only child, is merged with its parent. The operations, that can be performed on a radix tree, are as follows:

- insertion
- deletion
- searching
- find all strings with common prefix
- find predecessor
- find successor

All those operations exhibit complexity of $O(k)$, where k is the length of the string.

On the other hand, radix trees can only be applied to strings, or to sets which are directly mappable to strings. They also need serializable types to work on.

2.1.2 Cooperative multitasking

Cooperative multitasking is a strategy for systems, which need to run separate tasks in parallel. This is different from cooperative multitasking, where the scheduler periodically interrupts the running process and initializes context switch.

Using this strategy forces the programmer to design all the tasks in a specific way. Each task must return control to the environment after doing a small amount of work. This can become extremely difficult if there is no obvious main loop, or for long-running functions. If a process does not return timely, all other processes become CPU starved and freeze.

One of the advantages of cooperative multitasking is that context switches are controlled by the programmer. This way shared data can be protected and fewer atomic sections need to be entered. Another consequence of this rather simple design is low overhead.

2.2 Goals

The device is a typical realtime embedded system. As such, mainframe programming techniques are not applicable and embedded systems approach is required. Consequently, the following requirements are layed out to the program:

- Convenient programming – during development, reprogramming the device shall be quick and easy, preferably consisting of a single action.
- Convenient communication – both debug data and process information shall be easily accessible during device operation.
- Error tolerance – the device shall not freeze up upon an error, but should instead make best effort to keep the control loop active.
- Determinism – the programmer needs to have full control over what is executed when.
- Scriptability – the device shall be able to get reconfigured without a human at the other end id est a PC program shall be provided, which reconfigures the device in arbitrary ways.

In the oppinion of the author, all of these requirements have been fully fulfilled with the current version of the software. The written software is free and open-source in it's entirety - it is protected under the MIT license.

2.3 Language selection

The selected processor can be programmed in avr-assembler, C or C++. Additionally, compilers exist for other less widespread languages, such as ADA. The author has selected the C language for it's simplicity, widespread use and compiler support for all of it's features.

2.4 Bootloader

2.4.1 Motivation

Several approaches were evaluated to reprogram the chip.

As the chip resides in a socket, instead of being soldered to the Main board, it could be removed from it's socket and inserted into a programmer. This approach is useful, but was discarded as extremely time-consuming.

The next possible approach is In-system programming. This constitutes soldering a header onto the Main board, and connecting it to the SPI bus of the microcontroller. This approach was also rejected for the following reasons. The author was reluctant to pay both the price in real estate on the PCB and the additional complexity of wiring the header to the microcontroller. Not to mention using up some pins. Last but not least, this approach would significantly complicate the RESET pin noise protection circuit.

The most complex, but in the same time best performing, approach was selected. Utilizing the fact that serial communication to a PC is needed for other reasons, the author decided to transfer new programs over the same bus. This way the device can be reprogrammed purely by software, possibly even from a remote location.

2.4.2 Previous work

Many bootloader programs exist for the selected chip. Unfortunately, all of them are either:

- bloated with unnecessary functionality, thus large in size,
- do not work out of the box and thus require manipulation of the source code to work or
- are licensed under an incompatible license than the very permissive MIT license.

2.4.3 Implementation

The bootloader was implemented in standard-conforming C language, and written by the author in its entirety. It is nearly impossible to create smaller in size bootloader for this device. The program code is simple and straightforward, lacking complicated preprocessor directives. Consequently, it is in the opinion of the author, easy to read and modify by anyone. The code is of course publically visible and licensed in a way, that any person can freely and legally modify and redistribute it.

The bootloader instructions reside at the high end of the flash memory, while the application resides in the low end. Consequently, the application never knows that a bootloader program was installed.

The finished product was called 'megaboot' - a bootloader for atmega devices. The source code in its entirety is provided as appendix A1.

2.4.4 Communication protocol

The XMODEM protocol has been selected. This protocol is not used in modern high-bandwidth, high-complexity networking. To the contrary, this protocol was popular when computers possessed resources comparable to the selected microcontroller. It is a simple protocol with very little overhead (channel efficiency is 97%). Furthermore, it is easy to implement in few program instruction, using up small amount of flash memory.

2.4.5 Error checking

Error checking is provided throughout the program:

- CRC sums onto each received XMODEM packet
- "magic character" starts each packet
- addressing wrong (inexistent) flash pages is protected against
- receiving packets out of order is also checked agains

All the error checking code is conditionally included via the only preprocessor definition. This serves twin purpose. Not only does it clearly indicate error checking code apart from the actual program logic, but also provides the option to remove all error checking, should a person attempt to reduce the program size to the next smaller option.

2.4.6 Use

The build system passes the symbol `BOOTLOAD` to the application program. This symbol holds the address of the first instruction of the bootloader. The application is then free to jump to that address whenever required. In an assembler program, this would have been performed via a `RJMP` instruction. On the other hand, higher level languages cast the pointer to a function call address (optionally with the `_Noreturn` attribute) and call it.

After the bootloader has been invoked, it expects program data over the communication channel. Fortunately, many applications are available, which support the XMODEM protocol. One example is the popular and free serial terminal '`minicom`'.

2.5 External components

2.5.1 Onewire library

A C library, written by Peter Dannegger, Martin Thomas and others, is being used for communication with the thermometer.

2.5.2 PID implementation

A professional implementation of a parallel pid with integral saturation, implemented by Atmel employees, is used.

2.6 Architecture

2.6.1 Theory

Numerous program architectures have been published throughout embedded programming books. However, they all fall in one of the following three groups:

- Continuous execution – suitable for very simple programs or devices, which do exactly one thing. The software performs its tasks one after the other, waiting as long as needed for the tasks to complete.
- Cooperative multitasking – used for complex embedded applications in resource-constrained environments. Timer tick is defined and upon each tick, the software performs all scheduled tasks. The differences with the above are:
 1. The tasks are performed once per specified time period, and not continuously. This is important for the PID algorithm.
 2. As a consequence, tasks must be written with multitasking in mind. Even a single process, which blocks for longer than the timer tick, would violate the real-time operation of the system.
- True preemptive multitasking – used for complex systems, such as personal computer operating systems. This approach is far superior to all others, because programs are easily written, and it is the operating system's responsibility to schedule them efficiently. Unfortunately, this is a complex task, utilizing expensive (*id est* slow and memory consuming) algorithms. Nevertheless, there exist lightweight multitasking operating systems for atmega devices.

The author has selected the second approach.

2.6.2 Implementation

The file 'src/main.c' performs scheduling of the services, provided by the remaining modules. The author has put effort into making all other modules independent of each other. This is expected to provide the following benefits:

1. The code is simple and easy to understand.
2. The implementation is not prone to error, due to changing an unrelated portion of the code.
3. Software modules are easy to remove from the current project and insert into another, unrelated, project.

2.7 Modules

2.7.1 Configuration

Following best programming practices, the author has exported global project setting to the file 'src/config.h'. From there it is easy to control:

- all serial communication options
- hardware bindings i.e. which pins fulfill which tasks
- thermometer library settings
- if error checking to be performed

2.7.2 Clock

This module utilizes a separate hardware timer to provide timekeeping and event management services. The module keeps track of relative time, that is seconds elapsed from turn-on. The clock wrap period is 136 years.

Event management is achieved through the function `clock_sleep_until_next_second()`. This is the essence of the multitasking system.

2.7.3 Serial communication

The built-in 'Universal synchronous and asynchronous receive and transmit unit' or USART module is utilized in an elegant way. By providing implementation to functions `put_char()` and `get_char()`, the `uart` software modules actually binds '`stdin`', '`stderr`' and '`stdout`' streams to the USART unit. This facilitates usage of standard stream operations, such as '`printf()`' and '`scanf()`'. An added advantage is that the serial communication implementation can be changed - e.g. to flow over USB or Bluetooth, and zero client code will need to be modified.

2.7.4 Commands

Utilizing the multitasking architecture of the software, a 'commands' module is implemented, which accepts user instructions at all possible time points, irrespective of currently executed tasks. An added advantage of the commands system is that the device is fully scriptable. Elaborating, a 'bash' script can issue commands to the system and read status back. Such a script can, for example, log temperature data and perform control actions by the device. As a matter of fact, such a script is provided, together with a Qt4 implementation of a temperature data visualization tool. For more details, see the section Plotting tool

2.7.5 Zero-cross detector

An elaborate algorithm for mains voltage control was implemented. The following options were considered:

- Perform pulse width modulation on the control signal to the heater. The sinusoidal nature of mains voltage introduces an absolute timing error of up to $2 * 10\text{ms}$. In the author's opinion, with this simple scheme, only on-off control is possible.
- Perform classical phase control. This approach has proven itself over the years. Furthermore, extensive literature is available, concerning this well-understood phenomenon.
- Measure AC phase and turn on only specific integer number of half-waves.

As a PID control algorithm is desired, the first approach is immediately discarded. The author moved away from the second approach for the following reasons. The device in its entirety is connected to the mains in a single point. Phase control not only produces significant electro-magnetic interference, but also requires a zero-cross detector. Such a detector relies on the clean waveform of the measured signal. Consequently, in such a

scheme, the triac would be injecting noise into the zero-cross sensor, necessitating extensive filtering, with all of the associated complexity and latency.

Consequently, the author has selected the final approach. By paying the price of less resolution than actual analog phase control, the following benefits have been attained:

- Significantly reduced EMI - the selective activation of separate half-periods introduces a harmonic of 100Hz frequency. However, its power is believed to be significantly lower than the combined power of the infinite spectrum of a phase-controlled system.
- Simplified control - phase control would have necessitated floating point numbers. Linking in the floating point math library is expensive (both execution speed and storage space wise).
- Indeterminate - the IEEE standard for floating point implementation in computers highlights numerous cases, where the result of a computation is implementation-dependent. In contrast, an integer number implementation is always fully deterministic.

But wait! The author has identified an algorithm to increase the precision to theoretically infinite value. By utilizing a variation of Bresenham's line algorithm, cumulative error is eliminated given enough time. Furthermore, the algorithm supports safe modification of the setpoint, without relying on the fact, that the system tick is of known duration. The implementation of the algorithm is quite simple and can be observed in file `src/zcd.c`, being contained in the function `should_turn_on()`;

2.7.6 Triac control

2.7.7 Temperature measurement

2.8 Plotting tool

Plotting a signal in real time can be achieved by numerous approaches. Among them are MATLAB, Scilab, Mathematica, Qt etc. However, those that are not proprietary, are still badly suited to real-time operation. Thus were selected the 'python' and 'matplotlib' technologies for their versatility.

The written plotting tool is called 'rtplot' - RealTime Plotting utility. It operates on a vector of signals, plotting them against different time periods. The longer time periods obtain their datapoints via algebraic averaging of the next shortest plot. The program output covers the whole display and presents axis labels and units of measure. Individual datapoints are also displayed, in addition to the interpolating line between them.

3 Algorithms

3.1 Theory

3.1.1 Parallel PID controller

The analog PID controller is an enhancement of the PI controller. An additional component is added to the PI sum. It depends on the speed of change of the error but is expected to be zero at steady-state. Here is how PID control looks like in the time domain:

$$u(t) = K_p e(t) + K_p \frac{1}{T_i s} \int_0^{t_1} e(t) dt + K_p T_d \frac{de(t)}{dt} = u_p(t) + u_i(t) + u_d(t) \quad (2)$$

where T_d is called derivative time constant.

In state space the PID control law looks like this:

$$u(s) = K_p e(s) + K_p \frac{1}{T_i s} + K_p T_d s e(s) = u_p(s) + u_i(s) + u_d(s) \quad (3)$$

And for the transfer function (this is a parallel structure PID) we have:

$$W_{pid}(s) = \frac{u(s)}{e(s)} = K_p + K_p \frac{1}{T_i s} + K_p T_d s = K_p \left[\frac{T_i T_d s^2 + T_i s + 1}{T_i s} \right] \quad (4)$$

The advantages of the PID controller include swift transient processes and reduced overshoot. The main disadvantages is that when either the setpoint $u(t)$ or the Load Disturbance changes, the effect is opposite – the derivative term strives for overshoot. This is called 'set point kick' and is malicious.

3.1.2 Anti-windup

Among numerous practical applications, the permissible range of values for the controller output is limited. For example, a mechanical valve cannot be opened wider than its physical construction allows. A triac-controlled heater cannot supply negative thermal energy. When the control signal $u(t)$ exceeds the maximum possible control, the integral factor becomes ineffective.

The most popular algorithms for alleviating this problem are:

- Limiting the magnitude of the integral sum in $[u^{max}, u^{min}]$.
- Summing only on valid computed control (not clipped).

3.1.3 Plant models

Some algorithms for tuning controllers rely on a model of the plant. Others rely on observing a step response or stable oscillations state. Thus, firstly the most commonly used models are examined.

A two parameter model accounts for pure delay and first order plant dynamics. It is used for integrating plants.

$$W(p) = \frac{a}{Lp} e^{-Lp} \quad (5)$$

,where:

L - pure time delay,

a - coefficient.

A three parameter model instead of aperiodic type.

$$W(p) = \frac{K_0}{1 + Tp} e^{-Lp} \quad (6)$$

where:

L - pure time delay,

K_0 - gain,

T - dominant time constant.

3.1.4 PID tuning algorithms

The **tangent method** is a graphical or computational algorithm. A tangent line is constructed through the inflection point of the plant's step response. By measuring the coordinates of the intersection of this tangent with y or x axis, the parameters of 2 or 3 parameter model are estimated.

The **area-based method** for estimating coefficients of a 3 coefficient plant is a computational method. Measured is the area between the setpoint and step response. Also, measured is the area between the step response and x axis. Solely by using those two measurements, the area-based method estimates 3 parameter models.

Optimization-based methods can fit models of unspecified order.

The **first method of Ziegler-Nichols** is significant both for its simplicity and for its history value. The algorithm defines mapping from coefficients in (5) to recommend p, i and d gains for the closed system. Even among the plethora of modern methods, Ziegler-Nichols I remains a useful starting point for more complex methods, such as optimizing or self-tuning.

The **second method of Ziegler-Nichols** is based on steering the plant into sustained oscillations. The period and magnitude of the resultant oscillations are measured. Together

with the loop time and the amplitude of the relay, a second table provides suggestions on K_p , T_i , T_d . This method is considered an improvement on the Ziegler-Nichols I method.

Astrom-Hagglund also utilize the self-oscillating mode of target plant. However, this method advises a different way of controlling the system into oscillations. This relay-based method is believed to be more effective than ZN for stable or inert plants.

Of course there exist numerous more complicated tuning algorithms, many of which proprietary. Just to name a few: optimization of value function, neural networks, robust parameters for controlling stochastic plant.

3.1.5 Digital PID controller

One method relies on the following assumptions:

- $u_i(t)$ is estimated using second rectangle method
- $u_b(t)$ is estimated via first backwards finite difference.

This is the non-recursive description of the controller:

$$u(k) = K_p \left[e(k) + \frac{T_0}{T_i} \sum_{i=1}^k e(i) + \frac{T_d}{T_0} (e(k) - e(k-1)) \right] \quad (7)$$

If the control law is to be computed by a digital device, a more efficient approach would be as follows. Let us assume

$$\begin{aligned} K_i &= \frac{K_p T_0}{T_i} \\ K_d &= \frac{K_p T_d}{T_0} \end{aligned}$$

Then the PID controller can be written in the following recursive format:

$$u(k) = K_p e(k) + I(k) + K_d (e(k) - e(k-1)) \quad (8)$$

$$I(k) = I(k-1) + K_i e(k), I(0) = 0 \quad (9)$$

3.2 PID implementation

Used is a discrete pid algorithm with parallel structure and integral windup. The computer implementation expects proportional integral and derivative gains, as opposed to time constants. Consequently, the responsibility for translating from T_i, T_d to K_i, K_d falls to the client software, written by the author.

3.3 Units of measure

Because various physical and computational resources are performed in the system, unit conversions need to be carefully considered.

We are constrained by the PID algorithm software to use 'int16_t' values for input, output and coefficients. Furthermore, coefficients are represented in 9s6 format i.e. 128 == 1.0.

Secondly, the temperature sensor interface library returns 'decicelsius_t'. This is an integral value of measured degrees celsius, multiplied by 10. The accuracy of the thermometer is 0.5°C

Last constraint is the input to the triac control. In an attempt to increase the resolution of the triac control algorithm as far as possible, the author has used the whole range of 'uint16_t' values. In other words, 0 corresponds to 0% or 0V at the heater, and $2^{16} = 65536$ corresponds to 100% or 230V at the heater.

Apparently, the PID algorithm needs to read 'decicelsius_t' and output 'uint16_t', which is impossible. One of the two units needs to be selected for the algorithm to work with, and the other - converted to/from. The author has selected 'decicelsius_t' as the more human-readable format. Furthermore, no additional precision is lost by using this shriked type.

3.4 Nonlinearity

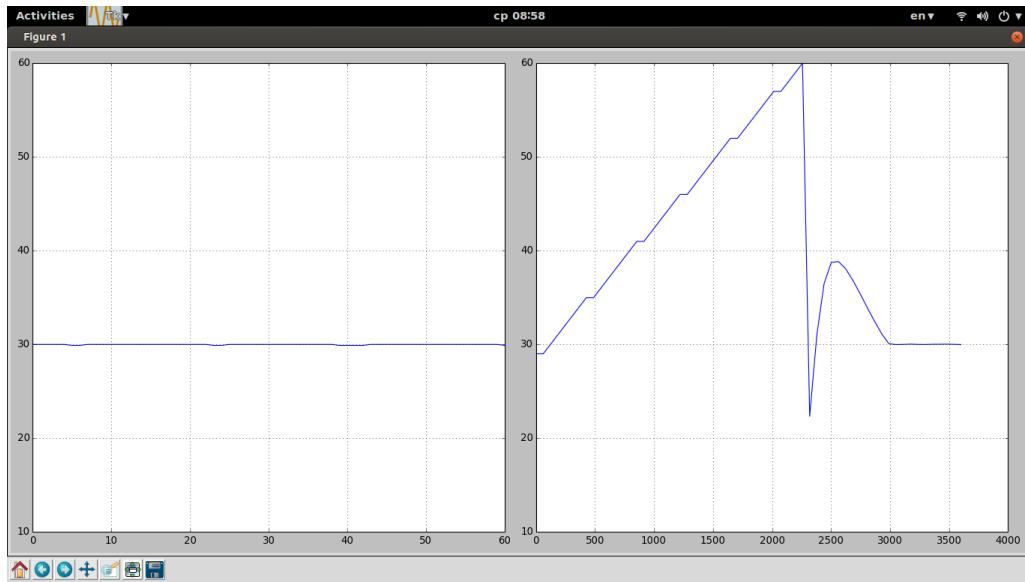
The heating circuit is expected to exhibit good linearity. However, the cooling process is both not under the control of the regulator, and extremely non-linear. The author expects cooling to conform to the Stefan-Bolzman law of radiated thermal energy in free space:

$$L = \frac{\sigma}{\pi} T^4$$

If this is true, estimating a transfer function of the whole system is pointless. The resulting linear system will be close to reality only for a very limited range of setpoints.

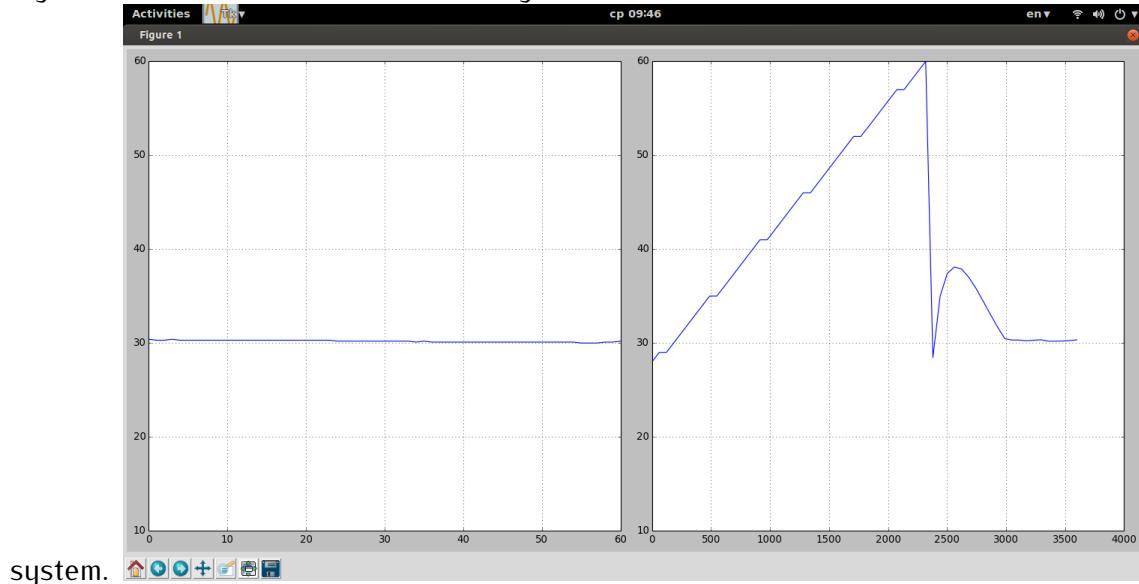
3.5 Ziegler-Nichols sustained oscillations

Initially, the Ziegler-Nichols algorithm for obtaining sustained oscillations and thus critical gain K_u and critical period T_u was attempted. It was observed that if any oscillations are observed at all, their amplitude is lower than the absolute error in the system.



3.6 Astrom-Hagglund sustained oscillations

Secondly, replacing the PID controller with a relay was attempted. The relay acts as a sign function. It was observed that again no sustained oscillations are exhibited by the



system.

3.7 Sampling rate correction

Following advice in the literature, the author increased the control algorithm period to roughly 1/10'th of the dominant open loop system time constant, namely to 60 seconds. Finally sustained oscillations were observed.

