# Using MicroPython to create a CBUS module

## Introduction

It’s nearly five years since I first released a set of code libraries to help people create MERG CBUS modules using the Arduino ecosystem. Before then, the non-expert member would have had to wait for one of ‘the experts’ to get around to designing the CBUS module they needed. Additionally, many early MERG modules were coded in PIC assembly language, making derivatives and modifications difficult for the non-specialist.

My intention – a philosophy, if you will – was to provide everything required to create a CBUS FLiM module that would talk to FCU and JMRI, learn events and node variables, send CBUS messages, etc. This provides a platform for the designer to add the specific functionality – or personality – of their module. I’m pleased that a number of members have found these libraries useful for creating their own layout control modules and I am constantly amazed by their creativity and ingenuity.

However, technology moves on apace and we now have access to many more powerful and low-cost microcontrollers and boards developed around them. So, whilst it has been possible for some time to design CBUS modules using the Arduino ecosystem, this has still meant learning the C and C++ languages.

One significant development has been in the area of languages, specifically those that are easier for beginners to get to grips with, because the larger and faster processors no longer absolutely need the efficiency of lower-level languages or assembler.

## The Python programming language

The Python programming language has been around for 20+ years and frequently appears in global Top 3 lists of the most popular languages. It is often seen as an easier environment for beginners, and for a number of reasons. It is an interpreted language and provides an interactive command line – think of a 1980s home computer on steroids - which promotes exploration and experimentation without the usual edit/compile/link/upload/crash/debug cycle that we traditionally associate with embedded systems development. It also has a rich set of data structures ‘out of the box’, such as classes, lists of various kinds, queues, events, locks, etc. For the more experienced programmer, it is a very productive language, enabling the creation of new functionality very quickly. Rarely, if ever, does it crash or hang the hardware; it just produces a (mostly) useful error message.

MicroPython is an implementation of Python for microcontrollers and supports a growing range of familiar hobbyist platforms, including the Raspberry Pi Pico, ESP8266 and ESP32, as well as some other Arm-based processors such as STM32 and Atmel SAMD. It includes the vast majority of ‘big’ Python’s functionality (reportedly around 90% of the most-used facilities), omitting only those elements not relevant or achievable on a microcontroller, but adding functionality for interfacing with the outside world. (You may also come across CircuitPython, which is a fork of MicroPython with an emphasis on use in education settings. Like Python, MicroPython and its derivatives are open-source projects).

Being an interpreted language, it runs slower and requires more memory than compiled languages, maybe 10x in some cases, but it is, in my opinion at least, 10x more productive. Rather like the Arduino ecosystem, there is in-built and 3rd party support for the common things we would like to communicate with, including displays, servos, sensors, etc and anything you can attach using the I2C and SPI bus protocols, including CAN controller chips! WiFi is supported too, on boards with the necessary hardware.

The are many, many learning resources available for Python (and increasingly, MicroPython too), including books, tutorials, YouTube videos, blogs, etc, etc.

Performance is more than adequate for all but the most time-sensitive of applications, for which C, C++ and assembler still exist as options.

(It’s worth pointing out that MicroPython will never be available for smaller microcontrollers such as 8-bit PICs and AVR-based Arduino modules. They are simply not powerful enough.)

## MicroPython for CBUS

I have recently created a set of MicroPython modules to achieve a similar aim. Initially just a direct port of my Arduino code, the productivity of the Python environment has enabled me to add some new things that members may find interesting and perhaps useful. (Rather confusingly, Python uses the term ’module’ to describe a separate file or files of code that can be imported to add to its built-in functionality).

At the basic level, the modules provide similar functionality to my Arduino libraries, and I have provided examples that represent a ‘blank’ module to which the user can add their specific functionality. Depending on your hardware, you may just need to change the pin numbers for the CAN transceiver chip and the CBUS switch and LEDs, if fitted. (I also have a couple of CBUS shield designs available; see later).

Using these code modules is even easier than Arduino, as all that is required is to copy the source (.py) files to the board of choice. Whilst there are a number of free IDEs available to make life easier, everything can be done with a text editor, a terminal app, and a couple of command line programs.

For those used to developing with my Arduino libraries, the next steps will be familiar. You can register functions to be called when CBUS messages and previously taught events are received, and then process those messages according to your program’s logic. However, there are other ways to skin this particular cat, as I shall explain later.

## Choosing your hardware

As I mentioned earlier, MicroPython supports a growing range of microcontroller platforms, including some that members may already be familiar with. Code written for one platform will be mostly portable to others, and virtually 100% so with a few minor changes. There is even a version for Unix and Windows if you wish to learn or test things out on your PC.

I have chosen the Raspberry Pi Pico board for my initial development work. It is cheap, fast, has a fairly large memory, and is readily available in quantity from many UK suppliers. (Also, we have a Raspberry Pi retail store here in Cambridge, so I can easily pick one up when we’re out shopping). It is now also available with WiFi capability (as the Pico W). It does require an external CAN bus device for communicating with CBUS, but these are readily available, or you can easily construct your own for a few pounds. I currently support the ubiquitous MCP2515 CAN controller chip and I have published PCBs for a couple of matching ‘shield’ designs. (\*\*\* insert photos \*\*\*)

As it stands, there is no currently-supported or available processor with a built-in CAN controller peripheral. The ESP32 has such a peripheral but this is not (yet …) supported by MicroPython. The leader of the MicroPython project sells a board based on an STM32 microcontroller which does have CAN support (the pyboard), but this is comparatively expensive and is currently out of stock (as of January 2023) due to the global semiconductor shortage.

For more ambitious programs, the Pico has a larger available memory than the ESP32 or ESP8266. I haven’t yet tried running my code on the ESPs, so I can’t confirm their suitability yet.

(Note that there is also traditional Arduino support for the Pico, and my C/C++ CBUS libraries run well on it).

## Multitasking

MicroPython has a simple multi-tasking capability that enables multiple threads of execution (aka ‘tasks’) to appear to run concurrently. This enables us to move away from the traditional ‘superloop’ approach traditionally used for embedded programming (e.g., the loop() function present in every Arduino sketch), which often leads to pages of spaghetti code. By composing code in smaller, concurrent blocks, we can make programs easier to reason about, read, write, and debug.

Multitasking in MicroPython is co-operative rather than pre-emptive, meaning that tasks must explicitly yield back to the scheduler, although this tends to happen naturally as they wait for things of interest to happen (interrupts, messages, events, timers, etc). A task that fails to co-operate will just block everyone else out (just like early Windows programming).

The Pico’s RP2040 Arm processor has two cores, and this immediately invites speculation on how they might both be used. Python has historically not been able to use multiple processor cores (qv. Python GIL), even on large systems. Suffice it to say that we can use both cores of the Pico in MicroPython … but it’s complicated.

That said, for simple functionality, you don’t need to jump into the complexity of multi-tasking as you can follow the same familiar application structure as an Arduino sketch, with the equivalent of its straightforward setup and loop functions. The example programs provide a good starting point.

One of the more powerful capabilities is the provision of an interactive command-line (aka ‘REPL’) that can be used whilst CBUS message processing happens ‘in the background’. This enables one to query variables, call library functions, send CBUS messages, and even write simple Python functions on the fly. This brings a new dimension to debugging and promotes learning and experimentation without the overhead of continual recompilation.

(The Arduino C/C++ support for the Pico incorporates FreeRTOS for taking advantage of the second core and is probably the better option if you really need the additional performance – or just want to use what you paid for!).

## Getting started

There are many ways to get started depending on your existing programming skills and what you hope to achieve.

For beginners, and even experienced programmers who are new to Python, I would suggest you first get comfortable with the Python language itself, its structure and basic syntax. You can do this using the ‘big’ Python interpreter that is almost certainly already installed on your PC. If it isn’t, you can download it from <https://www.python.org/downloads/>. Make sure you get version 3 as version 2 is now deprecated.

The next step is to get yourself a Pico (or another supported board) and follow one of the many tutorials on how to get started with it. This will almost certainly involve downloading a simple IDE called Thonny and using this to install MicroPython on your Pico. Then get comfortable with this environment, editing programs, uploading them to the Pico and running them.

Rather than fill this article with reams of code, I have created a separate ‘getting started’ document which is included with the Python files in my GitHub repo (see links below). This document assumes you have the necessary hardware to connect to your CBUS.

## Beyond a simple Arduino module

One of my personal interests is layout automation, which can be blamed on a thriller novel I read as a teenager in the 1970s that featured an extensive model railway that ran itself. That idea has stayed with me all these years!

I wanted to be able to implement a degree of automation that didn’t depend on the complexity and cost of a PC (or single-board computer such as a Raspberry Pi) running JMRI or expensive commercial software like TrainController.

I have taken advantage of the productivity of Python, and larger memory, to add functionality that didn’t appear in my Arduino libraries.

There are a couple of ‘abstract’ classes that provide the foundation for other, more familiar, layout objects:

The first class implements the ‘publish and subscribe’ pattern (aka pubsub). This allows a task to say: “I’m only interested in these one or two CBUS messages. I’m going to sleep now but please wake me up when one of them arrives, at which point I’ll look at it and do something”. The task consumes no processing time whilst it waits.

The next is CBUS history, which extends a powerful concept introduced by Ian Hogg’s CANCOMPUTE module. A task can maintain a list of recently received CBUS messages which is added to as new messages arrive from the bus. This ‘history’ has a finite lifetime, so only (say) the last 10 seconds’ worth of messages is kept. The task then waits until a message of interest is added to its history, at which point it wakes up and can execute queries against the current list. For example, it might ask: “has this group of three messages been received, in this order, within the last 10 seconds, and within a timespan of 5 seconds? If so, I’ll do something interesting”. This is useful where we wish to apply application logic to sequences of more than one message without resorting to complex state machines in code. Any number of tasks can maintain their own individual history, with different lifetimes, message filters and queries, subject only to memory and processing time constraints. The MERG wiki contains some interesting use-cases for CANCOMPUTE (e.g., interlocking), any of which could be also satisfied by this approach.

A ‘sensor’ is a simple object (cribbed from JMRI) that maintains its state by listening for two CBUS events (using a pubsub object) which are produced by whatever hardware it is monitoring (maybe spot or block detection, a turnout microswitch connected to an input module, or servo position events from a CANMIO). A sensor may also be created with a further CBUS message which it uses to poll the feedback device, in order to establish its initial state. Once created, we can query the sensor’s state at any time or block and wait for it to change. This is a binary sensor (with two distinct states), but we can also create ‘value’ sensors, whose state is a value that depends on data events produced by, for example, an RFID or RailCom detector. A ‘multi-state’ sensor can represent the current position of a turntable with multiple exits.

We can now turn to classes that represent various ‘real things’ on a layout that we may wish to operate, interrogate, and automate, including turnouts and signals, train detection, routes, clocks, and DCC locos. A couple of us are also experimenting with RailCom for additional feedback and reporting.

A turnout (or semaphore signal) is simply an object created with the two CBUS events needed to operate its real-life counterpart. At any point in time, it has a state (e.g., thrown or closed), even if that is initially unknown. We can then associate a binary sensor object with this turnout to provide positive feedback. Thus, we can operate the turnout and either wait for its new state to be confirmed by the sensor object or query its state at any time later.

Groups of turnout and signal objects can be composed into ‘routes’, enabling us to operate them as a group in a prototypically correct sequence. It is further possible to ‘lock’ a route and its component objects so that it can’t be changed whilst a train movement is in progress. The route is then released or allowed to time out. An individual object may appear in multiple different routes (e.g., a turnout in a complex station throat) but only one of those routes may be acquired and set at a time.

By adding a couple of pushbutton switches (using a history object) we can create a simple NX (eNtry/eXit) route. Thus, operating the two switches in order (as detected by a query to its history object) will cause the route to be acquired, locked, and set. The NX route object itself can produce its own CBUS feedback events, so we can control the mimic panel lighting correctly according to the prototype, illuminating the switches and route as it is selected. Of course, it should not be possible to set a route if any of its track sections is currently occupied by a train, and this requirement is also catered for.

Classes also exist for multiple-aspect colour-light signals, as well as groups of signals which will cascade in prototype fashion when the first signal’s aspect is set.

There are also classes to control DCC locos, either using CBUS messages to a MERG CANCMD, or a serial or network connection to a DCC++ command station running on an Arduino board.

Bringing all these pieces (or rather, objects) together, we eventually get most of the way to automated (or at least sequenced) train movements. A train can traverse the layout between a predefined origin and destination, using detection sensors as waypoints to determine where it is. As it approaches a signal, it can interrogate that signal’s state and either set the turnout or route in front of itself, or wait for the human signaller to do so. Multiple movements can be in progress at the same – even human operated ones, so long as they obey the signals!

I have also created some clock classes, so automation sequences can be synchronised with the time of day, be that wall-clock or fast-clock time. A scheduled movement could wait until a specific time of day before departing, and then pause at en-route station stops for the desired period, according to the timetable.

The productivity and expressiveness of the Python language and the ability to compose complex objects from simpler ones, means that a user’s program could achieve all of this in a relatively few lines of code, as I hope the example code below shows.

\*\*\* insert object diagram here \*\*\*

We are used to sending single CBUS messages around the bus to other modules and having them act upon those messages. I have carried across my implementation of Dave McCabe’s CBUS long message protocol (RFC0005) from my Arduino libraries to this code base. This protocol enables modules to send and receive messages that span more than a single CBUS message, for example: to send a text message to be displayed by a module that is connected to a large LED display. As Python is an interpreted language, it is possible to pass around fragments of Python code which can then be executed by the recipient module. I have a number of future use-cases for this functionality. (That’s a lie; I have no idea how to use this but perhaps you do).

## Interfacing hardware

Much of the foregoing has assumed a CBUS layout where all operations happen via the sending and receiving of CBUS messages and events. Being a microcontroller, it is certainly possibly to connect things directly to the Pico (or other board). This covers the usual LCB requirements such as switches, LEDs, servos, displays and other types of sensors, and all of these can be integrated into a MicroPython program.

For example, I have created a servo class for directly-attached servos. It can be operated directly in code, or it can listen for CBUS events and operate itself, as well as emit servo movement events (i.e., start, midpoint, complete).

In common with most current microcontrollers, the Pico (and the ESPs) is a 3.3V device, meaning that some kind of logic level shifting will be required to interface with 5V devices, including most existing MERG modules. I am currently working on a development of my Pico CBUS shield that includes built-in level-shifting on some inputs. External level-shifting modules of various types are easily available from the usual sources.

## Other possibilities

I mentioned earlier the Pico W, which has a WiFi interface. I have used this to create a wireless CBUS interface, rather like a CANUSB4 or CANETHER but supporting multiple concurrent wireless clients. This only required about 20 lines of code over and above the existing library facilities.

With Internet access, it can also provide a precise time source for the layout, using CBUS FCLK messages. This would also be possible using an ESP32 or ESP8266, although these are constrained by having less memory than the Pico, and thus the size of programs that they run.

## Current status

As at the time of writing (January 2023), the basic CBUS module code is complete and has been tested using my PCB designs. It is now being tentatively released and evaluated by a few MERG members. I will be giving a presentation to the Raspberry Pi SIG on 17th January. My thanks go to John Wilson for testing and bug-finding.

By the time you read this, I hope there will be some real-world examples of other members’ work available to talk about. I will also have created a wiki page on the MERG Knowledgebase.

## Code examples

\*\*\* insert examples here \*\*\*