Real-Time Systems - Assignment 02 Group 9

Zhengtao Huang (5833469, zhengtaohuang)

Barry Tee Wei Cong (5662834, btee)

1 Introduction

In this lab assignment, a fully-written code of a synthesiser has been provided, that has bugs due to inefficient task execution in an infinite loop. This report will seek to apply the knowledge learnt from CESE4025 Real-Time Systems and answer the embedded system questions listed in the lab manual, including the use of the built-in Zephyr Real-Time Operating System (RTOS) scheduler to implement a fully functional synthesiser, which will be documented in each section. Our report will focus on the main topics, which include characterising the tasks, separating them into threads, handling overload, and using an interrupt-driven peripherals interface. The overall goal of the functional synthesiser is to produce clean sounds with high responsivity.

2 First Encounter with the Synthesiser

As shown below, the team would refer to the assignment instructions for the four main tasks.

Task 1: Check if the state of any peripheral (switch or rotary encoder) has been changed and update the appropriate synthesizer data structures.

Task 2: Check and process the input keys from the keyboard.

Task 3: Synthesise/Generate the sound to be played.

Task 4: Send the generated sound to the audio amplifier.

2.1 Qn 1 - Task 3 and 4

The team checked the hardware and started to switch SW1 to the up position with one note playing. The team then used the Logic Analyser to record the Worst Case Execution Time(WCET) and Average Case Execution Time(ACET) as shown in Table 1.

Qn1	Task 3 (ms)		Task 4 (µs)	
Values	31.911	31.913	2.500	2.500
	31.900	31.895	2.750	2.500
	31.903	31.909	2.500	2.500
	31.908	31.903	2.500	2.500
	31.895	31.900	2.500	2.500
WCET	31.913 ms		2.750 μs	
ACET	31.904 ms		2.525 μs	

Table 1: Tasks 3 & 4 - ACET and WCET

2.2 Qn 1 - Task 2 - Use Cases

The team defined 2 Use Cases that will lead to ACET and WCET. We believed both Use Cases stated below are two extremes that are general and representative of measuring ACET and WCET as shown in Table 2.

Case 1 - Only One Note is Playing Case 2 - Multiple Notes are Playing

Task 2	Case 1 (µs)		Case 2 (µs)	
Values	6.500	1.750	10.250	10.000
	2.000	1.750	1.750	6.250
	1.750	1.750	1.750	9.750
	2.000	1.750	1.750	9.750
	1.750	2.000	1.750	19.750
WCET	6.500 μs		19.750 µs	
ACET	2.300 μs		$5.889 \; \mu s$	

Table 2: Task 2 - ACET and WCET

2.3 Qn 1 - Task 1 - Use Cases

The team defined 2 Use Cases that will lead to ACET and WCET. We believed both Use Cases stated below are two extremes that are general and representative of measuring ACET and WCET as shown in Table 3.

Case 1 - No Peripherals Are Tuned

Case 2 - Frequency Shifter Knob Turned

Task 1	Case 1 (µs)		Case 2 (µs)	
Values	817.75	817.75	879.75	817.00
	817.75	817.75	879.75	880.75
	817.75	817.75	879.75	880.75
	817.75	817.75	879.75	879.25
	817.75	817.75	879.75	818.25
WCET	817.75 μs		880.7	50 μs
ACET	817.75 μs		867.475 μs	

Table 3: Task 1 - ACET and WCET

2.4 % of Processor Utilisation (PU)

In Table 4, the WCET and the respective PU of the tasks are shown. The period for all tasks in the superloop is set to 50 milliseconds.

Period - 50ms	Time - WCET	PU (%)
Task 1	880.75 µs	1.762
Task 2	19.750 µs	0.040
Task 3	$31.913 \; \text{ms}$	63.826
Task 4	2.750 μs	0.006
Total	32.740 ms	65.634

Table 4: Processor Utilisation - Task 1 to 4 (WCET)

The team assigned 64.00% to Task 3 inside the superloop to accommodate Task 3 not to easily lapse into overload situations, while setting a boundary for Task 3's task to have excessive occupation within one full period.

3 Synthesiser Tasks Overhaul

3.1 Qn 2 - Lower and Upper Bound - Chosen Period

The team first explored the possible range available to choose for Task 2, 3, 4. The lowest period that the team considered plausible was 30 ms. Any value lower than 30ms was not ideal as the audio output of the synthesiser would sound distorted. The highest period was 70ms as any higher value would cause the audio output to be slow and unresponsive.

The team chose 50ms as the period for Task 2, 3, 4 for two main reasons. The first reason is due to having sufficient duration for all four tasks to be feasible to schedule, given that the processor is sufficiently utilised with a low probability of overloading. The second reason is that 50ms is fast enough for quick music beats to be played. This enables high system responsivity, which is a requirement in music genres like Electronic Dance Music(EDM).

Although the team would like to reduce the period to 40ms, the period would be set at 50ms to ensure that minimal issues crop up during the development stages for the nine questions to be completed in the assignment.

The team first explored the possible range available to choose for Task 1. The lowest period that the team considered plausible was 2ms. Any value lower than 2ms was not ideal as the audio output of the synthesiser would sound distorted. The highest period was 4ms as any higher value would cause the audio output to be slow and unresponsive.

The team chose 3ms as the period for Task 1 for it to not overload and yet be responsive.

3.2 Qn 3 - Set Priority for Each Thread

Before determining the priority for each thread, the team read the thread priorities offered by Zephyr RTOS. As shown in Figure 1, Zephyr RTOS provides two options, cooperative threads and preemptible threads.

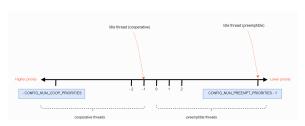


Figure 1: Cooperative Thread vs Preemptible Thread

Since cooperative threads will continue to run even if higher priority threads are ready to run, it is not the team's goal for cooperative threads to ensure a task complete without interruption. In contrast, the team having been taught of preemptive task scheduling chose preemptible threads that can be interrupted by the scheduler to allow another thread of higher priority to run. This process also known as preemption is selected for the music synthesiser to function based on the requirement set in the assignment.

Another rationale for the selection of preemptible threads for priority allocation is also based on the logical process of how the music synthesiser is being operated. The case of the rotary switch turning or the key being pressed is an example of user's actions having higher priority in Task 1 and 2 respectively. This in turn cause Task 3 and 4 to generate the audio signal and have the signal amplified based on the concerted higher priority tasks that user implemented to output from the synthesiser.

This means if a preemptible thread, i.e. Task 3 and 4 is currently running and a higher priority thread, i.e. Task 1 and 2 becomes ready to run, the scheduler will interrupt the current thread and switch context to the higher priority thread.

Zephyr documentation reveals that "the kernel supports a virtually unlimited number of thread priority levels". Hence, to reduce the complexity of the assignment, four threads have been configured for the four assigned tasks. The allocation is shown below.

```
[1] for Task 1
```

- [2] for Task 2
- [3] for Task 3
- [4] for Task 4

3.3 Qn 4 - Separate Superloop into Threads

Based on the priority set, the code block was written to implement the preemptible threads for the respective tasks 1 to 4. In the following code block, Task 1 to 4 were coded with their respective function names priority numbering, as shown in bold text.

```
k_tid_t my_tid_0
= k_thread_create(&my_thread_data_0,
                stack0.
                K_THREAD_STACK_SIZEOF(stack0),
                task_update_peripherals,
                NULL, NULL, NULL,
                1, 0, K_NO_WAIT);
k_tid_t my_tid_1
= k_thread_create(&my_thread_data_1,
                stack1,
                K_THREAD_STACK_SIZEOF(stack1),
                task_check_keyboard,
                NULL, NULL, NULL,
                2, 0, K_NO_WAIT);
k_tid_t my_tid_2 =
k_thread_create(&my_thread_data_2,
                stack2,
                K_THREAD_STACK_SIZEOF(stack2),
                task_make_audio,
                NULL, NULL, mem_block,
                3, 0, K_NO_WAIT);
k_{tid_t my_{tid_3} =
k_thread_create(&my_thread_data_3,
                stack3,
                K_THREAD_STACK_SIZEOF(stack3),
                task_write_audio,
                NULL, NULL, mem_block,
```

4, 0, K_NO_WAIT);

As shown in Table 5, when no peripherals are tuned, the execution time for Task 1 takes longer, as ACET shows $821.100\mu s$ as compared to the previous $817.75\mu s$ in Question 1. When multiple threads are used, the system has to save the state of one thread and load the state of another every time a context switch occurs. This explains why Task 1 took longer when threads were used.

However, when the frequency shifter knob is turned, the execution time for Task 1 takes shorter as shown in Table 5, as WCET shows $840.750\mu s$ as compared to the previous $880.750\mu s$ in Question 1. When threads are created for separate tasks, Zephyr schedules these threads to utilise the CPU cores efficiently. If only one core is available, Zephyr will interleave the execution of threads, switching between them rapidly to give the illusion of parallel execution, which is particularly beneficial for I/O-bound tasks, which in this case the frequency shifter knob task benefited with a shorter execution time. This shorter execution time is a significant benefit that outweighed the slight increase in switching-time overhead during idle state.

Task 1	Case 1 (µs)		Case 2 (µs)	
Values	821.25	821.00	819.00	840.00
	821.00	821.00	840.75	820.75
	821.25	821.25	840.75	840.00
	821.00	821.00	840.75	820.75
	821.00	821.25	840.00	820.75
WCET	821.250 µs		840.750 µs	
ACET	821.100 μs		832.350 μs	

Table 5: Task 1 - New ACET and WCET

3.4 Qn 5 - Identify Race Conditions

The team reviewed the source code to find race conditions that would potentially occur. For each race condition listed, the team listed the threads that are involved. The team also described the operation to explain how it qualifies as a data race. Consequences have been listed to analyse their severity. Relevant actions were implemented in the code to resolve the intolerable data race condition.

• Race Condition between threads task_make_audio and task_write_audio

Type of Operation - Both threads are accessing and updating the shared memory block without any mutual exclusion.

Consequences - This race condition might lead to data corruption in the shared memory block, resulting in incorrect sound generation.

Action Taken - As instructed in the lab manual, the team implemented a double buffer with semaphores as the synchronisation method. The detailed solution is documented in question 6 of the lab report. It is ensured that only one of the two threads has access to one block of the double buffer at any given time.

• Data Race on keys array in task_check_keyboard thread and task_make_audio

Type of Operation - The keyboard checking thread might update the keys array while the audio synthesis thread is accessing it.

Consequences - This data race might result in unpredictable behaviour related to key states, hold times, and release times.

```
k_mutex_lock(&mutex_keys, K_FOREVER);
for (int i = 0; i < MAX_KEYS; i++)
    if (key == keys[i].key
        && keys[i].state != IDLE)
        keys[i].state = PRESSED;
        keys[i].hold_time = sys_timepoint_calc
                             (K_MSEC(500));
        keys[i].release_time = sys_timepoint_calc
                                 (K_MSEC(500));
        key_pressed = true;
// The second loop is necessary to avoid
// selecting an IDLE key when a
// PRESSED or RELEASED key is located
// further away on the array
    if (!key_pressed)
        for (int i = 0; i < MAX_KEYS; i++)
            if (keys[i].state == IDLE)
                keys[i].key = key;
                keys[i].state = PRESSED;
                keys[i].hold_time =
                    sys_timepoint_calc
                    (K_MSEC(500));
                keys[i].release_time =
                    sys_timepoint_calc
                    (K_MSEC(500));
                keys[i].phase1 = 0;
                keys[i].phase2 = 0;
                break;
```

k_mutex_unlock(&mutex_keys);

Action Taken - The provided code implements a thread-safe mechanism for handling key state changes within a data structure that represents keys, utilizing a mutex to ensure mutual exclusion. When a thread wants to update the state of keys, it first locks the mutex, which prevents other threads from entering the critical section where the shared keys array is being modified. The first loop searches for a key that matches a specified condition (non-IDLE state and same key value) and updates its state to PRESSED, along with setting its hold and release times. If this key is not found, indicating that it was not previously pressed, a second loop searches for an IDLE key to update with the new state and timing information.

Upon completion of the updates, the mutex is unlocked, allowing other threads to access and modify the keys array. This locking and unlocking pattern around the critical section where the keys array is accessed ensures that only one thread can make changes at a time, thus preventing data race.

• Data Race in Synthesizer::makesynth() function (synth.cpp) and the peripherals_update() function (peripherals.cpp). The two functions are called from the audio synthesising thread and the peripherals checking thread, respectively. The code suggests potential concurrent access to shared resources, specifically the switches[] and encoders[] arrays, without adequate synchronisation, which could result in race conditions.

Type of Operation - Reading and updating the shared state of switches[] and encoders[] arrays across different threads, namely task_update_peripherals and the task_make_audio thread.

Consequences - If these race conditions are not managed, they may lead to an inconsistent representation of peripheral states, causing unpredictable user interface behaviour for the synthesiser. Moreover, there might be missed updates or incorrect readings of switch states and encoder values.

```
k_mutex_lock(&mutex_peripherals, K_FOREVER);else
for (int j = 0; j < MAX_KEYS; j++)
    if (keys[j].state == PRESSED &&
        !sys_timepoint_expired
        (kevs[i].hold_time))
        sample += get_sound_sample(keys[j]);
     else if (keys[j].state == PRESSED &&
            sys_timepoint_expired(
                        keys[j].hold_time))
        keys[j].state = IDLE;
// Apply LPF
if (synth._lpf._cutoff_freq > 0.)
    sample = synth._lpf.filter(sample);
k_mutex_unlock(&mutex_peripherals);
if (k_sem_take(
            &sem_peripherals,
            K_FOREVER) == 0
    set_led(&debug_led0);
    k_mutex_lock(
            &mutex_peripherals,
            K_FOREVER);
    peripherals_update();
    k_mutex_unlock(&mutex_peripherals);
    reset_led(&debug_led0);
```

Action Taken - As shown in the code blocks above, void Synthesizer makesynth function in synth.cpp contains a critical section within k_mutex_lock(&mutex_peripherals, K_FOREVER); and unlock portion. The peripherals_update() function in the task_update_peripherals thread is ensured mutual exclusion by the same mutex. The action taken by the code in synth.cpp with the mutex is to generate an audio sample corresponding to keys that are currently pressed and process it through a low-pass filter if active, all within a thread-safe manner by using a mutex to prevent concurrent access to shared resources.

3.5 Qn 6 - Double Buffering

if (is_mem0)

The team implemented double buffering for the audio buffer first by multiplying the BLOCK_SIZE by 2 in the macro K_MEM_SLAB_DEFINE_STATIC found in line 12 of the audio.cpp file to obtain twice the amount of contiguous memory space. void *mem_block = allocBlock(); in line 106 of the main.cpp is a statement that calls the function allocBlock() and assigns its return value to a pointer variable named mem_block. allocBlock() is a function that allocates a memory block from a memory slab and then starts a transmission for an I2S device.

As part of the void task_make_audio function located in line 167 of the main.cpp file, the code block shown above states that if boolean is_mem0 equates to true, the data are written to the first memory space - mem_0. Otherwise if boolean is false, the data is written to the second memory space - mem_1. The boolean variable is_mem0 is inverted after a piece of data is written in the corresponding memory space. In this way, the memory block that the function writes to alternates with each new operation.

Within the same function, semaphores are used. It is a synchronisation mechanism that can be used to manage access to a shared resource by multiple threads or tasks in concurrent programming. It is often implemented as a counter that tracks the number of units of available resources.

In the code written, **k_sem_give** is a function that increments a semaphore's count. If any threads are waiting on the semaphore , one of them will be unlocked and allowed to proceed. **k_sem_give(&sem_mem0)** signifies that the task has finished using the memory block associated with **sem_mem0**, likewise for **sem_mem1**, and it is notifying other tasks that may be waiting for this memory block

that it is now available. After the task is done populating this memory with audio data, it signals that the memory is ready for use by giving the semaphore to the next function.

As part of the task_write_audio function located in line 196 of the main.cpp file, the code block shown above states a continuous loop represents a task that runs indefinitely to handle the audio data output. This task uses semaphore synchronisation to coordinate access to shared memory blocks that contain audio data. k_timer_start() helps to maintain a consistent period for audio data processing by measuring how long each iteration of the loop takes to execute.

The k_sem_take function attempts to take a semaphore without waiting (K_NO_WAIT). This function returns 0 if the semaphore is taken successfully, which means that the associated memory block is available for use.

If k_sem_take(&sem_mem0, K_NO_WAIT) succeeds, it indicates that the first memory block (mem_block) is ready, and writeBlock(mem_block) is called to output the audio data from this block.

Likewise, if k_sem_take(&sem_mem1, K_NO_WAIT) succeeds, it indicates that the second memory block (mem_block + BLOCK_SIZE) is ready, and writeBlock(((uint8_t *) mem_block)+BLOCK_SIZE) is called to output the audio data from this second block. The reset_led(&debug_led3) call would then turn off the orange LED found on the STM32F4 Discovery Board, indicating that the task has completed its current iteration of processing. The orange LED label is marked with an orange box as indicated in Figure 2 shown below.

Finally, the task is put to sleep whenever the timer expires. This ensures that the task writes audio data blocks at a consistent rate defined by BLOCK_GEN_PERIOD_MS. The duration of sleep is adjusted based on the time it took to process the audio data to maintain the timing regardless of execution delays. This loop continues indefinitely, writing audio data to an I2S device as long as the system is running and producing audio data.

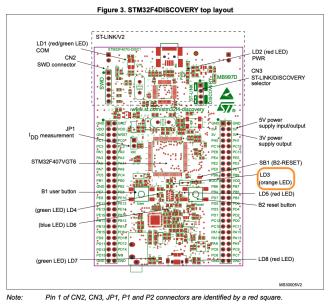


Figure 2: STM32F4 Discovery Board LEDs

4 Overload

```
4.1 Qn 7 - Handling Methods
```

As shown in the code block above, makesynth function in line 480 of synth.cpp, generates audio samples and writes them into a buffer block, called *frame*. The code written <code>set_led(&status_led3)</code>; function call turns on the red LED which serves as an indicator to the user that the synthesiser is overloaded. Likewise after successfully filling the buffer without overload, the code <code>reset_led(&status_led3)</code> at the end of the function turns off the red LED, indicating that processing has been reset and returned to normal.

As for the code logic to detect cases of overload, the code starts by setting a timer that expires just before the deadline of the current buffer generation period, which is (BLOCK_GEN_PERIOD_MS - 5) milliseconds. There is a check for k_timer_status_get. If the system time has reached the k_timer time, it means that the processor is in danger of overloading and missing the deadline.

When an overload is detected, the code enters an overload handling routine. The program stops synthesising the audio for the current frame, and the buffer is zeroed out with a nested for-loop, ensuring that any partially filled buffer does not contain garbage values that could produce noise.

5 Interrupts

5.1 Qn 8 - Switches

```
static struct gpio_callback sw0_cb;
static struct gpio_callback sw1_cb;
static struct gpio_callback sw2_cb;
static struct gpio_callback sw3_cb;
static struct gpio_callback sw4_cb;
static struct gpio_callback sw5_cb;
static struct gpio_callback sw6_cb;
static struct gpio_callback sw7_cb;
```

To move the code for updating the state of the 4 switches (8 pins) from the peripheral thread into separate interrupt handlers, the team first created the static struct <code>gpio_callback variables</code> for <code>(sw0_cb to sw7_cb)</code> in main.cpp as shown in the code block above. They are used to store the callback information for each switch's interrupt.

K_SEM_DEFINE(sem_peripherals, 0, 1);

As shown in the code block above, K_SEM_DEFINE(sem_peripherals, 0, 1) in main.cpp defines a binary semaphore that will be used for synchronization between the interrupt service routine (ISR) and the thread that will process the peripheral updates. The semaphore is initialized to "0", indicating that there's no initial permit for the semaphore. The processing thread will wait until an interrupt occurs.

```
void attach_interrupt_switch(void)
int ret0 = gpio_pin_interrupt_configure_dt
            (&sw_osc_dn, GPIO_INT_EDGE_BOTH);
int ret1 = gpio_pin_interrupt_configure_dt
            (&sw_osc_up, GPIO_INT_EDGE_BOTH);
int ret2 = gpio_pin_interrupt_configure_dt
            (&sw1_dn, GPIO_INT_EDGE_BOTH);
int ret3 = gpio_pin_interrupt_configure_dt
            (&sw1_up, GPIO_INT_EDGE_BOTH);
int ret4 = gpio_pin_interrupt_configure_dt
            (&sw2_dn, GPIO_INT_EDGE_BOTH);
int ret5 = gpio_pin_interrupt_configure_dt
            (&sw2_up, GPIO_INT_EDGE_BOTH);
int ret6 = gpio_pin_interrupt_configure_dt
            (&sw3_dn, GPIO_INT_EDGE_BOTH);
int ret7 = gpio_pin_interrupt_configure_dt
            (&sw3_up, GPIO_INT_EDGE_BOTH);
```

As shown in the code block above, the attach_interrupt_switch function configures GPIO pins to generate interrupts on both rising and falling edges (GPIO_INT_EDGE_BOTH) for each switch. This ensures that any change in the switch state

(on-to-off or off-to-on) will trigger an interrupt. The <code>gpio_pin_interrupt_configure_dt</code> function calls return codes (ret0 to ret7), which should be checked to ensure the configuration was successful.

```
k_sem_give(&sem_peripherals);
return;
```

As shown in the code block above, the switch_pressed function is the ISR that will be called whenever any of the switch's GPIO pins detect an edge. This routine does not process the switch state change; it simply signals the semaphore with k_sem_give(&sem_peripherals) which indicates to the peripheral thread that a peripheral update is required.

By using this design, the system minimises the amount of work done in interrupt context, which is important because ISRs should be as short and fast as possible to avoid missing other interrupts. Offloading the work to a separate thread allows for more complex processing without impacting interrupt latency.

5.2 Qn 9 - Rotary Encoders static struct gpio_callback rot_cb;

To move the code for updating the port expander that controls all the rotary encoders from the peripheral thread into separate interrupt handlers, the team first created the static struct <code>gpio_callback variables</code> for <code>(rot_cb)</code> in main.cpp as shown in the code block above. They are used to store the callback information for the interrupt generated by the port expander.

K_SEM_DEFINE(sem_peripherals, 0, 1);

As shown in the code block above, $K_SEM_DEFINE(sem_peripherals, 0, 1)$ in main.cpp defines a binary semaphore is the same as the one for the switches.

shown in the code block above. the attach_interrupt_switch function configures the GPIO pin connected to the port expander to generate interrupts on the falling edge (GPIO_INT_EDGE_FALLING). This configuration indicates that the interrupt should trigger when the signal is shifted from high to low. gpio_pin_interrupt_configure_dt function calls the return code (ret8), which should be checked to ensure that the configuration was successful.

k_sem_give(&sem_peripherals);
return;

As shown in the code block above, the switch_pressed function is the ISR that will be called when the GPIO pin connected to the port expander detects a falling edge. Similar to the previous example, this routine does not process the state change; it simply signals the semaphore with k_sem_give(&sem_peripherals), indicating that a peripheral update is required.

6 Conclusion

Based on the analysis and implementation detailed in this report, the team aims that the assignment finding has apprised the reader with a comprehensive understanding of real-time systems.

The thorough exploration of thread priorities, period selection, task separation, synchronisation mechanisms, and overload handling has provided a key overview into the intricacies of designing a functional music synthesiser with efficient task execution.

The use of Zephyr RTOS thread scheduler has proven to be pivotal in facilitating the prioritisation and scheduling of tasks, ensuring that the synthesiser functions smoothly and responsively. The implementation of double buffering, semaphore usage, and overload handling mechanisms has further enhanced the robustness and reliability of the synthesiser.

To be exact, the thorough investigation and identification of potential race conditions with the correct measures taken to address them have contributed significantly to the overall stability and performance of the synthesiser.

In conclusion, this assignment has been an enriching experience that has deepened the team's learning and understanding of real-time systems. The chance to work with embedded system design for a real-world application in music synthesis provides valuable insights and practical skills that are essential for a career in embedded systems.

References

[1] Project, Zephyr.

www.zephyrproject.org/common — multithreading — problems — and — their — fixes/ "Common Multithreading Problems and Their Fixes." Zephyr Project, 6 July 2023 (accessed 2 Jan. 2024)

[2] Project, Zephyr.

https — //docs.zephyrproject.org/latest/kernel/services/threads/index.html#thread — priorities "Threads — Zephyr Project Documentation." Docs.zephyrproject.org, 5 Apr. 2022 (accessed 2 Jan. 2024)

[3] TU Delft

https://cese.pages.ewi.tudelft.nl/real - time - systems/docs/Synthesizer_v1_3_schematic.pdf "Synthesizer Print Schematic," CESE, Aug. 29, 2023. (accessed 2 Jan. 2024).

[4] STMicroelectronics

 $https://www.st.com/resource/en/user_manual/\\um1472-discovery-kit-with-stm32f407vg-\\mcu-stmicroelectronics.pdf$ "Discovery Kit with STM32F407VG MCU - User Manual," STMicroelectronics, Oct. 01, 2020. (accessed 2 Jan. 2024).