Computer Architecture

LAB3: Virtual Machine & Linux Device Driver

Simulation Scenario

- We have created an environment that booting Linux kernel on ARM Realview EB by QEMU.
- This LAB will design a virtual hardware running in ARM Realview EB and interacting with Linux device driver and application to achieve some specific work.
- We implement a simple calculator as example.
- Application write two operands to calculator and let it go.
- When calculator complete, it will notify CPU through interrupt. CPU then execute ISR and wake up application to receive the result.

Define Virtual Hardware

- Register file (address and effect value)
  - We have to define the register space and memory space.
  - The hardware will execute the particular behavior according to the related registers' value.
  - Local memory space holds the necessary data.
- 3 basic functions
  - Virtual hardware simulated by QEMU
  - Initial Function
    - When virtual hardware is booting.
  - Read Function
    - When we wish to get some data from the virtual hardware
  - Write Function
    - When we want to write some data to virtual hardware's register file or memory

Design A Virtual Hardware

- Define calculator's register and memory
  - Register
    - control
      - Address: 0x00
      - Effect value: 1 start, 2 stop, 3 clear interrupt
    - operator
      - Address: 0x4
      - Effect value: 0 plus, 1 minus, 2 multiply, 3 divide
  - Memory
    - Address 0x8~0X14
      - 2 operands,1 result and 1 status
- The addresses above are offsets.

LAB3 總說明(1/3)

- 延續lab2的模擬環境，差別在於新增一個虛擬硬體calculator給QEMU進行模擬。
- 教材包檔案(皆為修改好的檔案，可直接使用)如下：

教材包 (Lab3/qemu)	檔案放置位置	說明
caslab_calculator.*	/qemu-3.0.0/hw/char/	QEMU所模擬的虛擬硬體都會放在 hw 資料夾之下，而我們硬體架構採用簡單的 character device，所以我們將設計好的虛擬硬體 – caslab_calculator 放到 hw/char 資料夾下。
Makefile.target	/qemu-3.0.0/hw/char/	讓QEMU知道必須編譯caslab_calculator此虛擬硬體
realview.c	/qemu-3.0.0/hw/arm/	新增了一行程式碼 sysbus_create_simple("caslab_calc", 0x80000000, pic[30]); 意思是指定虛擬硬體的base為0x80000000，且使用IRQ阜號30

- 執行步驟：

1) 在qemu新增完虛擬硬體calculator按照LAB1方法重新建構qemu
   - qemu-3.0.0/build]# make clean
   - qemu-3.0.0/build]# ../configure --target-list=arm-softmmu,arm-linux-user --prefix=${HOME}/workspace/qemu-bin-3.0.0
   - qemu-3.0.0/build]# make && make install
2) 重新對kernel做設定與編譯(教材包內的.config其實已經設定好了)
   - linux-4.14.85]# make ARCH=arm menuconfig
   - 設定**Enable loadable module support**開啟
   - linux-4.14.85]# make ARCH=arm CROSS_COMPILE=arm-none-linux-gnueabi-
3) 於driver資料夾內下make指令編譯device driver
4) 將編譯完成的calc_drv.ko複製到/busybox-1.29.3/_install/home/

LAB3 總說明(2/3)

- 執行步驟：

5) 修改Linux Kernel Device tree

教材包 (Lab3/kernel/ device_tree)	檔案放置位置	說明
arm-realview-eb.dts	linux-4.14.85/arch/arm/boot/dts	ARM Realview 平台架構描述檔
arm-realview-eb.dtsi	linux-4.14.85/arch/arm/boot/dts	細部裝置參數設定檔（多平台共用）

6) 增加我們額外加入的硬體資訊(教材包內的Device Tree描述檔已經設定好了)

- 編輯arm-realview-eb.dts，於文件底部增加

```
&cascalc {
  interrupt-parent = <&intc>;
  interrupts = <0 30 IRQ_TYPE_LEVEL_HIGH>;
};
```

表示該平台(arm-realview-eb)有使用一裝置cascalc，IRQ Number 為30

- 編輯arm-realview-eb.dtsi，於文件內增加裝置相關資訊(位址為0x80000000, 大小為0x1000)

```
/ {
        compatible = "arm,realview-eb";
        //....
        cascalc: cascalc@80000000 {
                compatible = "caslab,calc";
                reg = <0x80000000 0x1000>;
                clocks = <&pclk>;
                clock-names = "apb_pclk";
        };
};
```

7) 重新編譯Device Tree

- linux-4.14.85]# make ARCH=arm CROSS_COMPILE=arm-none-linux-gnueabi- dtbs

LAB3 總說明(3/3)

- 執行步驟：

8) 編譯應用程式
   - program]# arm-none-linux-gnueabi-gcc calc.c -o calculate
   - program]# cp calculate ~/workspace/busybox-1.29.3/_install/home/

9) 重新建立initrd
   - busybox-1.29.3/_install]# find . | cpio -H newc -o > ../initrd
   - busybox-1.29.3/_install]# gzip -f ../initrd

10) 利用qemu-system-arm進行全系統模擬
   - qemu-bin-3.0.0/bin]# ./qemu-system-arm -M realview-eb -kernel zImage -initrd initrd.gz -cpu arm1136 -dtb arm-realview-eb.dtb

qemu-test	檔案位置
qemu-system-arm (步驟1)	qemu-bin-3.0.0/bin/qemu-system-arm
zImage (步驟2)	linux-4.14.85/arch/arm/boot/zImage
arm-realview-eb.dtb (步驟7)	linux-4.14.85/arch/arm/boot/dts/arm-realview-eb.dtb
initrd.gz (步驟8~9)	busybox-1.29.3/initrd.gz

11) QEMU 模擬器開啟(ctrl+alt+3 or GUI menu->view->serial0)
   - home]# insmod cal_drv.ko

   察看Linux kernel設定我們driver的掛載編號
   - home]# cat /proc/devices

   替此device宣告一個node，應用程式要讀寫的時候是對此node做讀寫
   - home]# mknod /dev/calculator c 252 0

   執行應用程式並觀察回傳值是否正確
   - home]# ./calculate 115 5

- LAB3總說明為概略性描述，接下來的投影片為詳細說明，務必仔細閱讀

Design A Virtual Hardware (1/5)

- Virtual hardware state
  - 下圖中，第一個structure為virtual hardware內部自己使用的資料結構，第二個structure則是向QEMU註冊virtual hardware所需要使用的資料結構。
  - 之後我們將會採用這兩組資料結構來進行虛擬硬體的初始化函式(Initial Function)。

```
typedef struct {
    SysBusDevice parent_obj;
    MemoryRegion iomem;

    uint32_t Reg[CASLAB_CALC_RegCount];
    CharBackend chr;
    qemu_irq irq;
    const unsigned char *id;
} CASLabCalcState;

static const VMStateDescription vmstate_caslab_calc = {
    .name               = TYPE_CASLAB_CALC,
    .version_id         = 1,
    .minimum_version_id = 1,
    .fields = (VMStateField[]) {
        VMSTATE_UINT32_ARRAY(Reg, CASLabCalcState,
                             CASLAB_CALC_RegCount),
        VMSTATE_END_OF_LIST()
    }
};
```

Design A Virtual Hardware (2/5)

- Initial Function
  - 向QEMU所模擬的系統BUS註冊虛擬硬體
    - SYS_BUS_DEVICE
  - 告知CPU，若要對虛擬硬體存取必須透過那些function
    - memory_region_init_io
    - 其中caslab_calc_ops內包含了存取該裝置時會使用到的參數、函數，資料結構中包含了read、write時使用的函數與該裝置支援的位元順序與bus寬度，由於我們裝置存取皆以word為單位，因此min/max Size固定給4bytes
  - 註冊虛擬硬體的記憶體空間
    - sysbus_init_mmio
    - 註冊0x1000。假設我們將此虛擬硬體掛載在0x80000000，那CPU存取 0x80000000 ~ 0x80000FFF就會使用上面定義的RW函式。
  - 註冊虛擬硬體的IRQ阜號
    - sysbus_init_irq

```
static const MemoryRegionOps caslab_calc_ops = {
    .read       = caslab_calc_read,
    .write      = caslab_calc_write,
    .endianness = DEVICE_NATIVE_ENDIAN,
    .valid.min_access_size = 4,
    .valid.max_access_size = 4
};

static void caslab_calc_init(Object *obj)
{
    SysBusDevice *sbd = SYS_BUS_DEVICE(obj);
    CASLabCalcState *s = CASLAB_CALC(obj);

    memory_region_init_io(&s->iomem, OBJECT(s), &caslab_calc_ops, s,
                          TYPE_CASLAB_CALC, 0x100);
    sysbus_init_mmio(sbd, &s->iomem);
    sysbus_init_irq(sbd, &s->irq);
    DB_PRINT("INIT\n");
}

static const VMStateDescription vmstate_caslab_calc = {
    .name               = TYPE_CASLAB_CALC,
    .version_id         = 1,
    .minimum_version_id = 1,
    .fields = (VMStateField[]) {
        VMSTATE_UINT32_ARRAY(Reg, CASLabCalcState,
                             CASLAB_CALC_RegCount),
        VMSTATE_END_OF_LIST()
    }
};

static Property caslab_calc_properties[] = {
    DEFINE_PROP_CHR("chardev", CASLabCalcState, chr),
    DEFINE_PROP_END_OF_LIST(),
};

static void caslab_calc_class_init(ObjectClass *oc, void *data)
{
    DeviceClass *dc = DEVICE_CLASS(oc);

    dc->realize = caslab_calc_realize;
    dc->vmsd    = &vmstate_caslab_calc;
    dc->reset   = caslab_calc_reset;
    dc->props   = caslab_calc_properties;
}

static const TypeInfo caslab_calc_info = {
    .name          = TYPE_CASLAB_CALC,
    .parent        = TYPE_SYS_BUS_DEVICE,
    .instance_size = sizeof(CASLabCalcState),
    .instance_init = caslab_calc_init,
    .class_init    = caslab_calc_class_init,
};

static void caslab_calc_register_types(void)
{
    type_register_static(&caslab_calc_info);
}

type_init(caslab_calc_register_types)
```

Design A Virtual Hardware (3/5)

- 初始函式實作完成之後，我們必須要讓QEMU知道此硬體的存在，讓其啟動時可以將此虛擬硬體初始化。
- 上頁圖中，structure caslab_calc_info即是向QEMU宣告此虛擬硬體的相關起始函式與名稱。
- 實作QEMU所要求的增設虛擬硬體必要function
  - caslab_calc_register_types
    - type_register_static(& caslab_calc_info)
  - caslab_calc_class_init
    - 將硬體初始函式指向我們設計的初始函式，並待QEMU於執行週期時使用
  - caslab_calc_init
    - 整個虛擬硬體初始化的源頭
- 由於我們預計是在realview EB中增加，所以必須在hw/arm/realview.c中新增我們的虛擬硬體
  - sysbus_create_simple(" caslab_calc", 0x80000000, pic[30]);
  - 指定虛擬硬體的base為0x80000000，且使用IRQ皁號30

Design A Virtual Hardware (4/5)

- Read Function
  - 前面有提過，CPU只要讀取base~base+0x1000之間的位址就會來執行此函式。
  - 我們採用switch來做一個簡單的decoder。為了實作方便，這先於開頭將offset shift 2bites，將byte index轉為word index
  - 由於我們是以ARM為開發環境，所以ARM Linux在註冊device driver時會先確認硬體編號，此硬體編碼設定在0XFE0~0XFFC。
  - 我們定義在calculator_id_arm[8]中，並且當CPU要求這段位址時，我們將id送回去。ARM device的ID末四位元組一樣。
  - 在我們設計的calculator中，我們設定成應用程式只會讀取最後的結果，故除了ID、result與operand的address之外，皆是不可讀取的位址。

```
static const unsigned char caslab_calc_id_arm[8] =
    {0x11, 0x10, 0x88, 0x08, 0x0d, 0xf0, 0x05, 0xb1};

static uint64_t caslab_calc_read(void *opaque, hwaddr offset, unsigned size)
{
    CASLabCalcState *s = opaque;
    uint64_t ret = 0;
    offset >>= 2;
    switch(offset) {
        case 0x3F8:
            ret = caslab_calc_id_arm[0];
            break;
        case 0x3F9:
            ret = caslab_calc_id_arm[1];
            break;
        case 0x3FA:
            ret = caslab_calc_id_arm[2];
            break;
        case 0x3FB:
            ret = caslab_calc_id_arm[3];
            break;
        case 0x3FC:
            ret = caslab_calc_id_arm[4];
            break;
        case 0x3FD:
            ret = caslab_calc_id_arm[5];
            break;
        case 0x3FE:
            ret = caslab_calc_id_arm[6];
            break;
        case 0x3FF:
            ret = caslab_calc_id_arm[7];
            break;

        case CALC_REG_OPERAND1:
        case CALC_REG_OPERAND2:
        case CALC_REG_RESULT:
        case CALC_REG_STATUS:
            ret = s->Reg[offset];
            break;

        default:
            break;
    };
    DB_PRINT("Read value %lu @ 0x%02lx\n", ret, offset<<2);
    return ret;
}
```

Design A Virtual Hardware (5/5)

- Write Function
  - 與read function類似，但是input除了offset外還多了要寫入的value。
  - 我們一樣採用switch設計decoder，透過offset的判斷，我們依照一開始所訂下的register和memory位址來寫入value。
  - 這裡還多處理了interrupt的部分，我們可看到當CTRL被設定為GO時，write function會去觸發calculate並依據operator的定義完成對應的計算。計算完成之後將interrupt拉起。
  - CPU接收到interrupt之後會觸發ISR，我們設定ISR處理完成後會寫CLRI指令到CTRL的register，此時write function再將IRQ拉下以完成中斷處理。

```
static void caslab_calc_write(void *opaque, hwaddr offset, uint64_t value, unsigned size)
{
    CASLabCalcState *s = opaque;
    DB_PRINT("Write value %lu @ 0x%02lx\n", value, offset);
    offset >>= 2;
    if(offset >= CASLAB_CALC_RegCount)
        return;
    else
    {
        switch(offset)
        {
            case CALC_REG_CONTROL:
                if(CALC_STATUS_IDLE == s->Reg[CALC_REG_STATUS]) {
                    if( value == CALC_CTRL_GO ) {
                        DB_PRINT("CASlab Calculator commit\n");
                        caslab_calc_op(s);
                    } else if ( value == CALC_CTRL_CLRI ) {
                        qemu_set_irq(s->irq, 0);
                    }
                } else if(CALC_STATUS_BUSY == s->Reg[CALC_REG_STATUS]) {
                    if ( value == CALC_CTRL_ABORT ) {
                        //Cancel operations
                        s->Reg[CALC_REG_STATUS] = CALC_STATUS_IDLE;
                        qemu_set_irq(s->irq, 1);
                    }
                }
                break;
            case CALC_REG_OPERATOR://Writable register
            case CALC_REG_OPERAND1:
            case CALC_REG_OPERAND2:
                if(s->Reg[CALC_REG_STATUS] == CALC_STATUS_IDLE)
                    s->Reg[offset] = value;
                break;
            default: //Read only register
                break;
        }
    }
}
static void caslab_calc_op(CASLabCalcState *s)
{
    uint32_t op;
    uint32_t src1, src2;
    uint32_t result = 0;
    if(s == NULL)
        return;
    s->Reg[CALC_REG_STATUS] = CALC_STATUS_BUSY;

    op      = s->Reg[CALC_REG_OPERATOR];
    src1    = s->Reg[CALC_REG_OPERAND1];
    src2    = s->Reg[CALC_REG_OPERAND2];

    DB_PRINT("Doing OP %u with value %u, %u\n", op, src1, src2);
    switch(op) {
    case CALC_OP_ADD:
        result = src1 + src2;
        break;
    case CALC_OP_SUB:
        result = src1 - src2;
        break;
    case CALC_OP_MUL:
        result = src1 * src2;
        break;
    case CALC_OP_DIV:
        if(src2 == 0)
            fprintf(stderr, "Calculator: Invalid operand_2 !!\n");
        else
            result = src1 / src2;
        break;
    default:
        result = 0;
        break;
    }
    s->Reg[CALC_REG_STATUS] = CALC_STATUS_IDLE;
    s->Reg[CALC_REG_RESULT] = result;
    qemu_irq_raise(s->irq);
}
```

Compile The Virtual Hardware

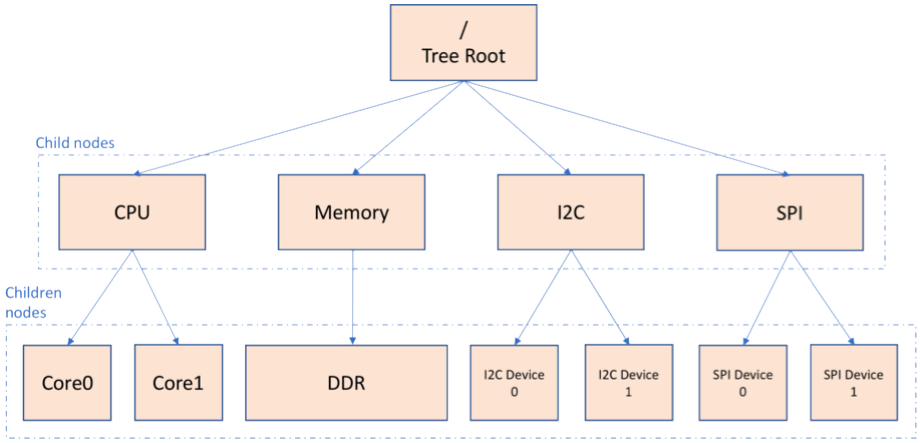
- QEMU所模擬的虛擬硬體都會放在hw資料夾之下，所以我們將設計好的虛擬硬體─calculator.c放到hw資料夾下。
- 為了讓QEMU知道必須編譯此虛擬硬體，在qemu-3.0.0/hw/char資料夾下有個Makefile.target檔案，加入”　obj-$(CONFIG_REALVIEW) += caslab_calculator.o”　，因為我們是針對arm realview系統平台撰寫此虛擬硬體。
- 依照LAB1的方法重新建構qemu，所得到的qemu-system-arm執行檔就有模擬calculator虛擬硬體的功能
- 之後我們將會利用此執行檔來進行純QEMU的全系統模擬。

Linux Device Tree – Overview (1/2)

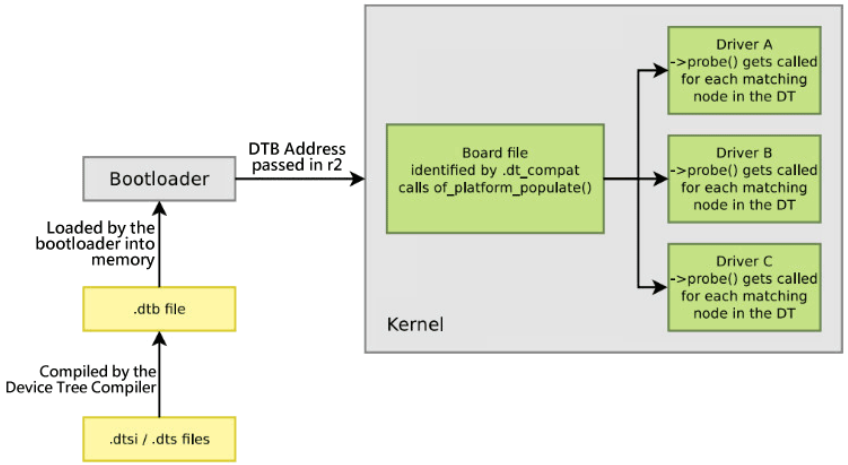
- 有別於x86系統硬體架構統一的特性，於Embedded System上有諸如ARM等硬體架構、針對不同應用又有更多不同種的SoC設計，而各家Vender也為其自家平台提交board specific kernel patch以支援該平台運行Linux的能力；而這造成了過多重複、冗餘的程式碼，導致Linux Kernel變得難以維護。
- 為了解決上述問題，Linux Kernel將各Vender / SoC所共有的硬體抽象化、並引入Device Tree以取代過去由各Vender所自訂的board specific code，以樹狀架構描述該平台的系統架構，交由Bootloader讀取、展開。

Linux Device Tree – Overview (2/2)

- Device Tree描述檔存於arch/<ISA>/boot/dts下，副檔名為.dts/.dtsi即為Device Tree source code。
- Source code內是以純文字表示，每個平台的描述檔結構皆為單一跟節點的樹狀結構，提供了硬體架構、硬體描述與硬體相依性等平台資訊。
- 需使用Device Tree Compiler將Source code轉為.dtb檔(device tree blob)提供Bootloader載入。



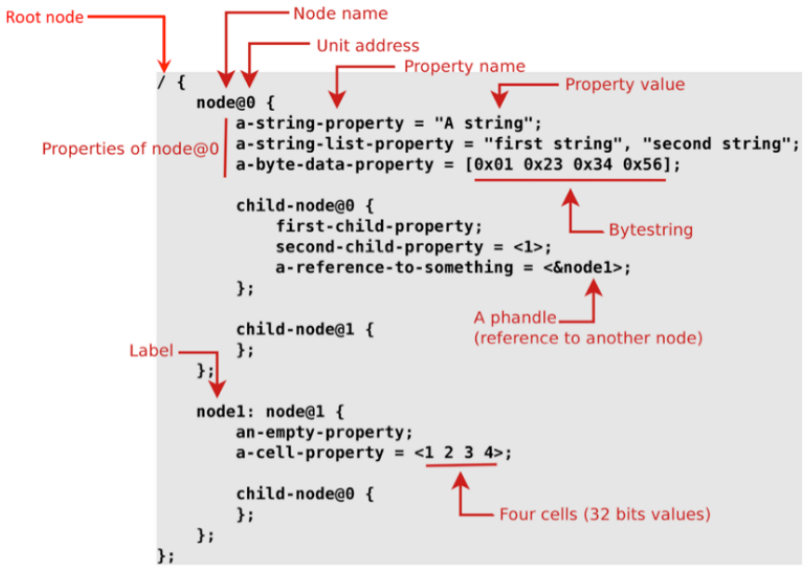
Simple Device Tree



Device Tree Processing Flow

Linux Device Tree – Basic Syntax

- Device Tree內的組成單位為一節點，一個節點內包含了若干個property & value來表示該節點的特性，並由若干節點依據硬體平台組織成一單一根節點的樹狀結構。
- 每個節點的命名規則為 <Name>[@unit-Address]，unit-Address如有提供即表示該硬體的主要位址（相對於父節點）
- 如有提供unit-Address，節點內會須提供property “Reg”來描述記憶體基底與空間大小。
- 以**compatible** property提供Driver裝置對應。



Linux Device Tree – Calculator

- 我們的實驗平台採用ARM的realview-eb的開發版，其Device Tree Source Code為 <Linux Kernel>/arch/arm/boot/dts/arm-realview-eb.dts*
- 我們所設計的Virtual Hardware是直接接上Bus的獨立硬體，其節點放在根節點下即可。
- 對應calculator的硬體特性，Device Tree描述為：
  - 硬體名稱(<manufacturer>,<model>)：
    - caslab,calc
  - 裝置地址、大小：
    - 0x80000000,0x1000bytes
  - 裝置使用clock：
    - 周邊clock(pclk)
  - Interrupt(於.dts內)：
    - 0 - shared processor interrupts
    - 30 - interrupt number
    - Active High

```
168 &cascalc {
169     interrupt-parent = <&intc>;
170     interrupts = <0 30 IRQ_TYPE_LEVEL_HIGH>;
171 };
```

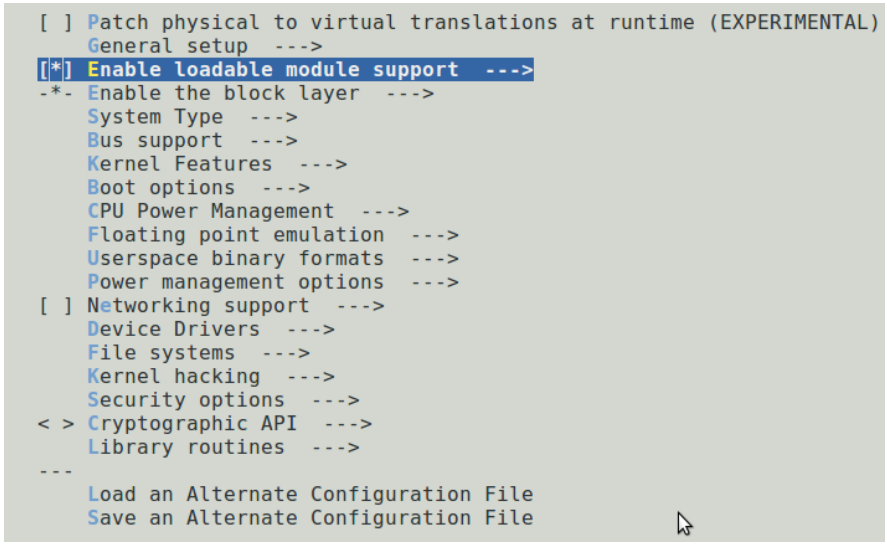
arm-realview-eb.dts

```
506     cascalc: cascalc@80000000 {
507         compatible = "caslab,calc";
508         reg = <0x80000000 0x1000>;
509         clocks = <&pclk>;
510         clock-names = "apb_pclk";
511     };
```

arm-realview-eb.dtsi

Linux Device Driver – Kernel Setting

- Linux device driver和Linux kernel以及所使用的cross compiler具有強烈的相關性，為了正確且順利的開發驅動程式，我們將採用Linux-4.14.85的kernel搭配arm-none-linux-gnueabi-gcc 4.8.3。
- 由於我們所設計的driver為module類型，因此我們重新對kernel做設定與編譯。
- 設定Enable loadable module support開啟後重編kernel。



Linux Device Driver - Overview

- 驅動程式必須依照其所對應的硬體做設計，因此我們配合先前所設計的calculator，採取一樣的register位址以及effect value。
- 我們採用char device來設計calculator的驅動程式。和設計虛擬硬體時一樣，device driver也必須實現basic function。
- Linux在處理硬體存取時，採用類似檔案存取的方式來完成，因此我我們就必須完成讀寫檔案的basic function，定義如下圖。
- 裡面較特殊的是unlocked_ioctl，其被用來對特定register位址寫入特定值，因此從R/W function中獨立出來。
- 除了這五個針對file的處理函示之外，還必須實作open與release兩個function，來處理driver的啟動與結束

```
static struct file_operations cascalc_fops ={
    .owner              = THIS_MODULE,
    .read               = calc_read,
    .write              = calc_write,
    .unlocked_ioctl     = calc_unlocked_ioctl,
    .open               = calc_open,
    .release            = calc_release
};
```

Linux Device Driver - Initial Function

- Initial function
  - 此函式處理所有driver初始化的工作，我們的driver是採用char device driver完成
  - 此函示定義了driver掛載上Linux後的識別碼包含Major與Minor number
  - 我們實作了allocate_memory_region來向kernel要求暫存所需記憶體空間
  - ioremap會將實體空間映射至虛擬記憶體空間
  - 最後是註冊IRQ的function: request_irq ，由於裝置硬體資訊為Device Tree進行維護，我們額外時做了查詢Device Tree內裝置IRQ的function : get_dut_irq，取得Kernel所映射的IRQ Number後才能註冊我們所設計的ISR。

```
static int __init cascalc_init(void) {
    dev_t dev = 0;
    int result;
    if(CASCALC_MAJOR) {
        dev = MKDEV(cascalc_major, cascalc_minor);
        result = register_chrdev_region( dev, 1, DRIVER_NAME);
    } else {
        result = alloc_chrdev_region( &dev, cascalc_minor, 1, DRIVER_NAME);
        cascalc_major = MAJOR(dev);
    }
    if(result < 0) {
        printk(KERN_ERR DRIVER_NAME "Can't allocate major number %d for %s\n", cascalc_major, DEVICE_NAME);
        return -EAGAIN;
    }
    cdev_init( &cascalc_cdev, &cascalc_fops);
    result = cdev_add( &cascalc_cdev, dev, 1);
    if(result < 0) {
        printk(KERN_ERR DRIVER_NAME "Can't add device %d\n", dev);
        return -EAGAIN;
    }

    /* allocate memory */
    base = ioremap_nocache(CASCALC_BASE , CASCALC_MEMREGION);
    if(base == NULL)
        return -ENXIO;

    irqnum = get_dut_irq(DEVICE_COMP);
    if(irqnum == -1) {
        printk(KERN_ERR DRIVER_NAME " Can't request IRQ, Device not found\n");
        return -ENXIO;
    }

    result = request_irq(irqnum, (void *)cascalc_interrupt, 0, DEVICE_NAME, NULL);
    if(result) {
        printk(KERN_ERR DRIVER_NAME " Can't request IRQ %d (return %d)\n", irqnum, result);
        return -EFAULT;
    }

    return 0;
}
```

```
static int get_dut_irq(const char* dev_compatible_name)
{
    struct device_node* dev_node;
    int irq = -1;
    dev_node = of_find_compatible_node(NULL, NULL, dev_compatible_name);
    if(!dev_node)
        return -1;
    irq = irq_of_parse_and_map(dev_node, 0);
    of_node_put(dev_node);
    return irq;
}

static DevicePrivate* allocate_memory_region(void)
{
    DevicePrivate* _prv;
    _prv = (DevicePrivate*)vmalloc(sizeof(DevicePrivate));
    return _prv;

}

static void free_allocated_memory(DevicePrivate *_prv)
{
    vfree(_prv);
    _prv = NULL;
}
```

Linux Device Driver - ISR

- 註冊IRQ的function中，除了通知Kernel皐號之外，還必須告知Interrupt Service Routine(ISR)的函式，我們的例子為cascalc_interrupt 。
- 顧名思義，當CPU接受到對應IRQ Number的時候，就會去執行由我們所定義的ISR，在此函式中，我們先將CLRI寫入Control Register，正好就是我們先前虛擬硬體定義的將IRQ放掉所必須做的對應動作，之後我們用wake_up_interrupt將process從waiting queue釋放出來。這部份我們將會在後面與把process擺入waiting queue中一起來看，這裡我們只要懂ISR做了甚麼即可。
- 當接收到IRQ時，就知道硬體已經完成工作或是需要軟體執行相對應的行為讓硬體可以繼續作業下去。

```
static irqreturn_t cascalc_interrupt(int irq, void *dev_id, struct pt_regs *regs)
{
    if (irq == irqnum) {
        printk("received interrupt from %s\n", DRIVER_NAME);
        writel(CALC_CTRL_CLRI, base + CALC_REG_CONTROL);
        printk("Release interrupt \n");
        flag = 1;
        wake_up_interruptible(&wq);
    }
    return IRQ_HANDLED;
}
```

Linux Device Driver – Read/ Write

- 先前我們提到，kernel將device視為檔案，因此我們必須實作讀檔寫檔等基本操作，那麼應用程式就能用read, write去操作device。
- 這裡最重要的兩個function為：copy_to_user及copy_from_user。
- 因為關係到將data從user space轉換到kernel space，因此我們需要這兩隻函式幫我們把一整串data轉移到driver，driver在針對這整筆資料做相對應的動作。
- 從下兩張圖看到，device_read透過readl將值讀回，並透過copy_to_user送回應用程式(注意不是用return)，而device_write則以copy_from_user得到應用程式想寫入device的值，再以writel寫進device。

```
static ssize_t calc_read(struct file *filp, char *buff, size_t len, loff_t *off)
{
    int i;
    DevicePrivate *prv_data = (DevicePrivate*)filp->private_data;

    /* Only result is readable memory address */
    if(len > 4)
        len = 4;

    for( i = 0; i < len/4; i++ )
        prv_data->buffer[i] = readl(base + CALC_REG_RESULT + i * 4);

    i = copy_to_user(buff, prv_data->buffer, len);

    return len;

}
```

```
static ssize_t calc_write(struct file *filp, const char *buff, size_t len, loff_t *off)
{
    int i;
    DevicePrivate *prv_data = (DevicePrivate*)filp->private_data;

    /* Two operands are writeable memory address */
    if ( len > 8 )
        len = 8;

    i = copy_from_user(prv_data->buffer, buff, len);
    for(i = 0; i < len/4; i++)
        writel(prv_data->buffer[i], base + CALC_REG_OPERAND1 + i*4);

    return len;
}
```

Linux Device Driver – IOCTL(1/2)

- 驅動程式設計，除了上述的對記憶體進行read/write之外，最重要的就是存取register file讓硬體可以做對應的工作，ioctl即是負責此部分的function。
- IOCTL必須提供應用程式一連串的commands,並且當應用程式使用這些command的時候可以做出相對應的動作。
- 讓我們回憶設計calculator時，我們設定operator的offset位址在0x04，而控制的offset位址則在0x00
- 而我們設定提供給應用程式的command有兩個，一個是設定operator的“OP_SET”定義為0x03，另一個則是啟動calculator的“CTRL”定義為0x04，搭配參數(arg)0x01發送”GO”訊號給硬體。

Linux Device Driver – IOCTL(2/2)

- Ioctl除了file指標之外，還會有兩個input，分別就是command與一項value，通常此value是要寫入的值，端看如何設計。
- 這邊我們還結合了waiting queue的概念，因為當應用程式使用“GO”指令要求calculator進行運算時，我們預設應用程式應該是處在等待運算結果回來的狀態，因此將此程序塞入waiting queue，讓出CPU給其他程序得以進行。
- 而OP_SET則是單純進行operator設定，因此利用writel將value寫入operator就完成了。

```
static long calc_unlocked_ioctl(struct file *filp, unsigned int cmd, unsigned long arg)
{
    long ret = 0;
    unsigned int tmp;
    int i;
    printk("IOCTL CMD = %u, value = %lu\n", cmd, arg);
    switch(cmd)
    {
        case CMD_STATUS:
            tmp = readl(base + CALC_REG_STATUS);
            i = copy_to_user((void*)arg, &tmp, sizeof(unsigned int));
        case CMD_OP_SET:
            if(arg >= CALC_OP_ADD && arg <= CALC_OP_DIV)
                writel(arg, base + CALC_REG_OPERATOR);
            else
                printk("Invaild OP %lu\n", arg);
            break;
        case CMD_CTRL:
            tmp = readl(base + CALC_REG_STATUS);
            if(tmp == CALC_STATUS_IDLE) {
                writel(arg, base + CALC_REG_CONTROL);
                wait_event_interruptible(wq, flag != 0);
                flag = 0;
            } else {
                if(arg == CALC_CTRL_ABORT || arg == CALC_CTRL_RESET) {
                    writel(arg, base + CALC_REG_CONTROL);
                    flag = 0;
                }
            }
            break;
        default:
            printk("Invaild CMD %u\n", cmd);
            ret = -ENOTTY;
            break;
    }
    return ret;
}
```

Linux Device Driver – Waiting Queue

- 在ISR與IOCTL中我們都提到了waiting queue的概念，也就是當應用程式必須等待硬體回應時，我們希望他先進入等待佇列而不是採用busy waiting。
- Linux Kernel本身就有提供此功能,標頭檔為 <linux/wait.h>
- 首先必須先宣告等待佇列以及喚醒條件旗標
  - static DECLARE_WAIT_QUEUE_HEAD(wq);
  - static int flag = 0;
- 當我們需要應用程式進去等待佇列時，採用
  - wait_event_interruptible(wq, flag!=0);
  - 則應用程式會進入wq等待被喚醒，且必要條件為flag!=0時。
- 而要喚醒程式時，
  - 會先設定喚醒條件旗標: flag = 1;
  - 在將程序從wq喚醒: wake_up_interruptible(&wq);

Compile Device Driver

- 我們先新增一個資料夾calculator-driver-app，並將設計完成的cal_drv.c放入此資料夾。
- Makefile
  - 由於device driver與kernel相關，因此必須設定kernel資料夾位址
    - KERNELDIR ?= ~/workspace/linux-4.14.85/
    - PWD := $(shell pwd)
  - 設定編譯環境
    - ARCH=arm
    - CROSS_COMPILE=arm-none-linux-gnueabi-
  - Compile
    - $(MAKE) ARCH=$(ARCH) CROSS_COMPILE=$(CROSS_COMPILE) -C $(KERNELDIR) M=$(PWD) modules
- 之後下make指令就可以成功編譯我們的device driver

Application

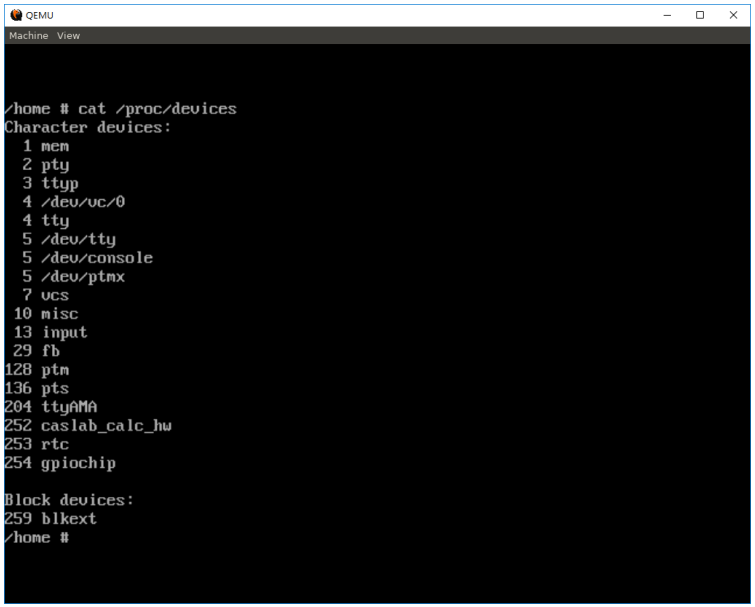
- 完成driver之後，應用程式就可以透過此driver去使用我們所設計的calculator。此應用程式我們亦放在program資料夾中
- 開啟device
  - calc_fd = open(DevPath, O_RDWR);
  - 其中DevPath 為 “ /dev/calculator” ，是我們在linux中所新增的node，後續會提到
- 寫入operand
  - write(calc_fd, (void*)buffer, sizeof(unsigned int) * 2);
- 寫入operator
  - ioctl(calc_fd, CMD_OP_SET, OP_LIST[i]);
- 通知硬體執行
  - ioctl(calc_fd, CMD_CTRL, 0x01);
- 讀回Result
  - read(calc_fd, (void*)&(result[i]), sizeof(int));
- 最後我們一樣透過arm-none-linux-gnueabi-gcc編譯
  - `arm-none-linux-gnueabi-gcc calculate.c –o calculate`
  - 執行檔為`calculate`

Simulation Flow

- 首先我們先將編譯出來的應用程式`calculate`放入先前busybox的_install/home中，並且產出相對應的initrd。
- Linux kernel則是4.14.85版本編譯完成的zImage。
- 之後就可以利用qemu-system-arm來進行全系統模擬。
  - ./qemu-system-arm -M realview-eb -kernel zImage -initrd initrd.gz -cpu arm1136
- 開機完成後必須掛載device driver
  - # insmod cal_drv.ko

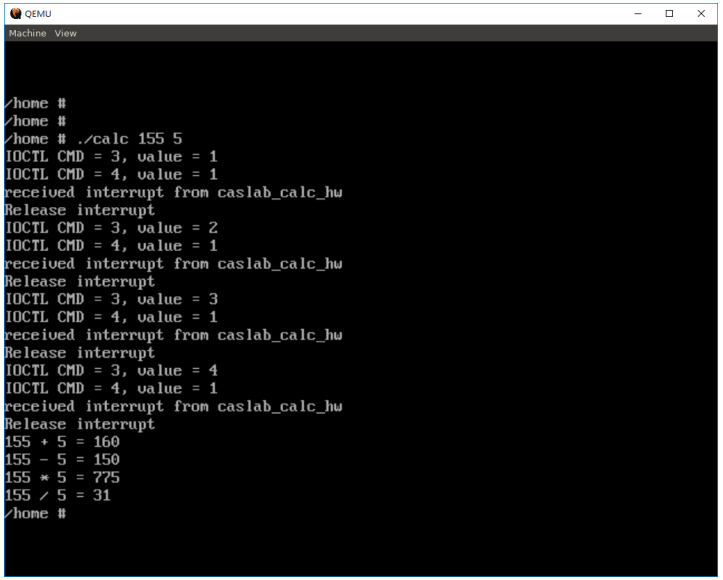
Simulation Flow

- 接著我們必須察看Linux kernel設定我們driver的掛載編號
  - # cat /proc/devices
  - 從圖我們看到掛載編號是252(因為我們的驅動程式採靜態登記法)。
- 之後我們就要替此device宣告一個node，應用程式要讀寫的時候是對此node做讀寫
  - # mknod /dev/calculator c 252 0
  - c表char device, 252為剛剛查到的編號(major number), 0為minor number。



Simulation Flow

- 此時應用程式開啟/dev/calculator才能成功並且可以順利操作driver來控制calculator。
- 執行應用程式並觀察回傳值是否正確
  - # ./calculate 115 5
  - 在範例程式中，後面接的參數為operand並且執行了加減乘除四項運算後把結果一次展現，途中會發現執行了四次ISR，以及觀察應用程式與硬體的互動為我們所預期。



Conclusion

- 在前面的LAB，我們以C語言設計了一套簡單的硬體calculator並且採用QEMU模擬。
- 為了讓應用程式可以順利使用此虛擬硬體，我們亦設計了對應的驅動程式。
- 但這些僅止於對硬體的行為(behavior)模擬，因為虛擬硬體完全不具有時間的概念，因此此種全系統模擬方法並無法滿足有時序的模擬硬體(cycle accurate)。
- 然而若將所有平台上的硬體(包含CPU)皆採用硬體模擬語言(HDL)開發，那其模擬速度將無法在可接受時間內啟動Linux OS。因此我們之後將會結合QEMU以及SystemC做分部模擬，讓QEMU全力執行Linux，而SystemC只模擬開發硬體(在此就是calculator)，冀望達到開發中硬體具有時間概念，並且可以做全系統模擬來發展應用程式與驅動程式，以此達到更高度的軟硬體整合。

Conclusion

- 為了使用上方便，我們可以建立靜態連結，如此一來不管是qemu重編或是linux kernel及initrd重編，都可以在我們建立靜態連結的資料夾中直接使用最新的檔案。
  - #cd ~/workspace
  - #mkdir ~/Exec
  - #cd ~/Exec
  - #ln -s {qemu-path}/arm-softmmu/qemu-system-arm .
  - #ln -s {busybox-path}/initrd.gz .
  - #ln -s {linux-kernel-path}/arch/arm/boot/zImage .
  - #ln -s {linux-kernel-path}/arch/arm/boot/dts/arm-realview-eb.dtb .
  - ./qemu-system-arm -M realview-eb -kernel zImage -initrd initrd.gz -dtb arm-realview-eb.dtb -cpu arm1136