ΗΥ-225: Οργάνωση Υπολογιστών
Άνοιξη 2019

Τμ. Επ. Υπολογιστών
© Πανεπιστήμιο Κρήτης

Σειρά Ασκήσεων 5: Διακλαδώσεις - Άλματα, If-Then-Else, Βρόχοι, Switch, Συγκρίσεις

Προθεσμία έως Δευτέρα 4 Μαρτίου 2019 (βδ. 5.1) ώρα 23:59 (από βδ. 3.3)

Βιβλίο: Διαβάστε την §2.7: σελ. 145-151 στο Ελληνικό, ή σελ. 92-97 στο Αγγλικό. Μπορείτε επίσης να δείτε, κάπως εγκυκλοπαιδικά, τις σελ. 115-120 του Αγγλικού βιβλίου· στο Ελληνικό βιβλίο, οι αντίστοιχες σελίδες ήταν οι 171-175, αλλά αυτές δεν σας βοηθάνε ιδιαίτερα, διότι αλλάζουν σημαντικά από τον MIPS στον RISC-V.

5.1 Διακλαδώσεις υπό Συνθήκη στον RISC-V

Οι εντολές μεταφοράς ελέγχου (CTI, control transfer instructions) καθορίζουν να εκτελεστεί σαν επόμενη εντολή --πάντοτε ή υπό ορισμένες συνθήκες μόνο-- μια έντολη άλλη από την "επόμενη από κάτω" τους εντολή. Όταν η μεταφορά ελέγχου γίνεται υπό συνθήκη, οι εντολές συνήθως ονομάζονται **διακλαδώσεις** (branch). Όταν η μεταφορά γίνεται πάντοτε, δηλαδή χωρίς συνθήκη, οι εντολές συνήθως λέγονται **άλματα** (jump). Μία παραλλαγή άλματος είναι το κάλεσμα διαδικασίας, και μιά άλλη η επιστροφή από διαδικασία.

Το πιό δημοφιλές *addressing mode*, δηλαδή τρόπος διευθυνσιοδότησης ή τρόπος προσδιορισμού της διεύθυνσης προορισμού των διακλαδώσεων, είναι το "**PC-Relative** addressing mode, δηλαδή ο προορισμός της διακλάδωσης δηλώνεται μέσα στην εντολή *σε σχέση με την εντολή διακλάδωσης αυτή καθεαυτή*, δηλαδή σαν σχετική απόσταση (μετακίνηση) από εδώ που βρισκόμαστε τώρα προς τα εκεί όπου θέλουμε να πάμε. Ο RISC-V χρησιμοποιεί και αυτός PC-Relative mode γιά τις εντολές CTI. Το PC-relative branch addressing mode mode προσφέρει δύο σημαντικά πλεονεκτήματα. Πρώτον, επειδή η απόσταση αυτή είναι συνήθως μικρή, αυτή μπορεί να χωρέσει σε λιγότερα bits μέσα στην εντολή· και είναι συνήθως μικρή επειδή οι διακλαδώσεις είναι εντός των if-then-else και εντός των βρόχων, που και τα δύο είναι εντός της ίδιας διαδικασίας (procedure), και όπου όλα αυτά είναι συνήθως σχετικά μικρά. Δεύτερον, όταν ο Linker συνενώνει αρχεία που έχουν γίνει compiled χωριστά γιά να φτιάξει ένα ενιαίο εκτελέσιμο (ή όταν κάνουμε dynamic linking at runtime), τότε η καθε διαδικασία καταλήγει να ξεκινά από μιά "καινούργια" διεύθυνση που δεν την ξέραμε νωρίτερα, άρα ο Linker έχει σχετικές διορθώσεις διευθύνσεων μέσα στον κώδικα που πρέπει να κάνει. Οι διορθώσεις αυτές είναι λιγότερες όπου χρησιμοποιείται PC-relative addressing mode, διότι όταν μετακινείται *ολόκληρη* μιά διαδικασία σε νέες διευθύνσεις, οι *σχετικές* αποστάσεις των εντολών *μέσα* στη διαδικασία *δεν* αλλάζουν.

Στη γλώσσα Assembly, η διεύθυνση προορισμού της διακλάδωσης δηλώνεται απλά με μια ετικέτα (label), και αναλαμβάνει ο Assembler να υπολογίσει και να βάλει τη σωστή δυαδική τιμή. Στη γλώσσα μηχανής του RISC-V, οι εντολές **διακλάδωσης** ακολουθούν το **B-format**, που όπως είπαμε είναι παραλλαγή του I-format και αφιερώνει, όπως κι εκείνο, 12 bits γιά μιά σταθερή ποσότητα, Imm12, που χρησιμοποιείται γιά να προσδιοριστεί η διεύθυνση προορισμού, ως εξής:

if (condition) then: `PC_new ← PC_br + (2*(signed)Imm12)` *else:* `PC_new ← PC_br + 4`

όπου `PC_br` είναι η διεύθυνση της ίδιας της εντολής διακλάδωσης, `Imm12` είναι η σταθερή ποσότητα των 12 bits μέσα στην εντολή, θεωρούμενη ως προσημασμένος αριθμός σε συμπλήρωμα ως προς 2 (δηλαδή sign-extended), και `PC_new` είναι η διεύθυνση της εντολής προορισμού σε περίπτωση επιτυχίας της διακλάδωσης. Ο πολλαπλασιασμός του Imm12 επί 2 γίνεται γιά να εκμεταλλευτούμε το γεγονός ότι η διεύθυνση όλων των εντολών του RISC-V είναι ακέραιο πολλαπλάσιο του 2 (συχνά, που δεν έχουμε και μικρές (16-μπιτες) εντολές, η διευθύνσεις όλων των εντολών είναι και ακέραια πολλαπλάσια του 4, αλλά αυτό δεν είναι πάντα εγγυημένο, κι έτσι ο RISC-V δεν στηρίζεται σε αυτό). Με τον τρόπο αυτό, διπλασιάζουμε το "βεληνεκές" των διακλαδώσεων –με άλλα λόγια, ο αριθμός Imm12 μετράει πλήθος "μικρών" (16-μπιτων) εντολών

μπροστά (θετικός) ή πίσω (αρνητικός), αντί να μετρά πλήθος Bytes μπροστά ή πίσω.

Η συνθήκη διακλάδωσης αφορά πάντοτε τη σύγκριση των δύο καταχωρητών πηγής, **rs1** και **rs2**, που περιέχει το B-format. Οι εντολές διακλάδωσης που υπάρχουν στον βασικό RISC-V, το συντακτικό τους σε Assembly, και οι συγκρίσεις που κάνουν είναι οι εξής:

- **`beq rs1, rs2, label`** # διακλάδωση εάν: `rs1 == rs2`
- **`bne rs1, rs2, label`** # διακλάδωση εάν: `rs1 ≠ rs2`
- **`blt rs1, rs2, label`** # διακλάδωση εάν: `rs1 < rs2` –θεωρώντας τους rs1, rs2 σαν προσημασμένους (signed)
- **`bge rs1, rs2, label`** # διακλάδωση εάν: `rs1 ≥ rs2` –θεωρώντας τους rs1, rs2 σαν προσημασμένους (signed)
- **`bltu rs1, rs2, label`** # διακλάδωση εάν: `rs1 < rs2` –θεωρώντας τους rs1, rs2 σαν απρόσημους (unsigned)
- **`bgeu rs1, rs2, label`** # διακλάδωση εάν: `rs1 ≥ rs2` –θεωρώντας τους rs1, rs2 σαν απρόσημους (unsigned)

Οι συντομογραφίες των ονομάτων προκύπτουν από τα: eq = equal; ne = not equal; lt = less than; ge = greater of equal. Οι συγκρίσεις ισότητας και ανισότητας (≠) είναι προφανώς οι ίδιες γιά αριθμούς signed και unsigned, αφού η ισότητα απαιτεί όλα τα bits να είναι ίδια, ένα προς ένα, ανεξαρτήτως του τι παριστάνει ο κάθε συνδυασμός από bits, άρα περιττεύουν εντολές "bequ" ή "bneu" αφού αυτές θα ήταν εντελώς ισοδύναμες με τις beq και bne που υπάρχουν.

Υπάρχει ένα περίφημο κόλπο γιά να ελέγξει κανείς την διπλή ανίσωση **`0 ≤ i < N`**, που είναι ο έλεγχος ορίων γιά το index ενός πίνακα μεγέθους N, χρησιμοποιώντας μία μόνο εντολή διακλάδωσης. Έστω ότι το array index i είναι προσημασμένος (signed) ακέραιος και βρίσκεται στον καταχωρητή x20, και το μέγεθος N ειναι θετικός ακέραιος και βρίσκεται στον x11. Τότε η εντολή: **`bgeu x20, x11, indexOutOfBounds`** πραγματοποιεί και τις δύο συγκρίσεις ταυτοχρόνως, διότι ερμηνεύει και συγκρίνει τον (προσημασμένο) ακέραιο i (x20) σαν να ήταν unsigned: Όταν το κάνουμε αυτό, σε αναπαράσταση συμπληρώματος ως προς 2, τότε, ως γνωστό, όλοι οι αρνητικοί αριθμοί μοιάζουν σαν πολύ μεγάλοι θετικοί αριθμοί – μεγαλύτεροι από τον μεγαλύτερο προσημασμένο θετικό· έτσι, εάν το index i είναι <0, τότε στη *απρόσημη (unsigned)* σύγκριση το i θα μοιάζει σαν πολύ μεγάλος θετικός αριθμός, σαφώς μεγαλύτερος από το N, άρα θα εμφανίζεται να παραβιάζει τον περιορισμό μας ξανά από την "επάνω" πλευρά, όπως και όταν i≥N.

<u>Άσκηση 5.2:</u> Απουσία Διακλαδώσεων ble, bgt, και Καταχωρητή-Σταθεράς

(α) Ο RISC-V *δεν* έχει εντολή ble (branch if less or equal), γιά περιπτώσεις όπως π.χ. if (i≤j), ούτε εντολή bgt (branch if greater than), γιά περιπτώσεις όπως π.χ. if (i>j). Γιατί; Θεωρήστε π.χ. ότι η μεταβλητή i βρίσκεται στον καταχωρητή x22 και η μεταβλητή j στον x23, και γράψτε τις παραπάνω διακλαδώσεις μέσω εντολών που έχει ο RISC-V.

(β) Επίσης, οι εντολές διακλάδωσης του RISC-V συγκρίνουν πάντα και μόνον δύο καταχωρητές μεταξύ τους –δεν υπάρχουν εντολές διακλάδωσης όπως π.χ. "`beqi rs1, Immediate, Label`" που να συγκρίνουν το περιεχόμενο ενός καταχωρητή με μία σταθερή ποσότητα. Γιατί άραγε;

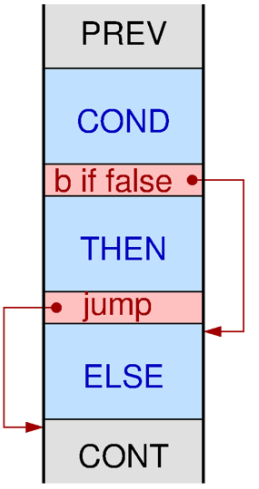
(γ) Δεδομένων λοιπόν όλων των παραπάνω που δεν υπάρχουν, πώς θα μεταφράζατε σε Assembly του RISC-V τον ψευδοκώδικα "branch if i≤13" (ενδεχομένως με τη χρήση κάποιου ενδιάμεσου προσωρινού καταχωρητή);
Δώστε τις απαντήσεις σας, με τις εξηγήσεις τους, γραπτώς.

5.3 Μετάφραση των If-Then-Else

Στο σχήμα δίπλα φαίνεται η γενική δομή της μετάφρασης σε Assembly ενός if-then-else statement, με τη χρήση μιάς εντολής διακλάδωσης (branch if... - "b if") και μιάς εντολής άλματος (χωρίς συνθήκη - "jump"). Το μπλοκ "PREV" παριστά τον κώδικα που υπάρχει στο πρόγραμμα πριν το if-then-else, και το μπλοκ "CONT" (continue) τον κώδικα μετά το if-then-else. Το μπλοκ "COND"

(condition) παριστά τον κώδικα (έκφραση - expression) συνθήκης που υπάρχει μέσα στην παρένθεση του "if", που μπορεί να είναι από πολύ μικρή εώς πολύ μεγάλη· όταν αυτή η συνθήκη είναι πολύ απλή (π.χ. "if (i==j)") τότε η μετάφρασή της μπορεί να γίνεται με μόνο την εντολή διακλάδωσης, ενώ όταν η συνθήκη είναι πιό πολύπλοκη τότε απαιτείται κώδικας "προετοιμασίας" πριν φτάσουμε στη διακλάδωση –τον οποίο κώδικα προετοιμασίας τον παριστούμε σαν "COND" στο σχήμα. Τα μπλοκ κώδικα THEN και ELSE συνήθως βολεύει να τα τοποθετήσει ο compiler στη μνήμη εντολών με την ίδια σειρά με την οποία τα διαβάζει από τον πηγαίο κώδικα, δηλαδή πρώτα το THEN και μετά το ELSE.

```
PREV;
if (COND) {
    THEN;
} else {
    ELSE;
}
CONT;
```



Η κάτω πλευρά του σχήματος παριστά τη μνήμη εντολών, με τις διευθύνσεις να αυξάνουν από πάνω προς τα κάτω. Αμέσως μετά την εκτέλεση των εντολών PREV, ο επεξεργαστής *πάντα* αρχίζει να υπολογίζει τη συνθήκη COND, της οποίας οι εντολές είναι τοποθετημένες "κολλητά" μετά τις εντολές PREV. Αμέσως μετά από αυτό τον υπολογισμό της συνθήκης, η εκτέλεση πρέπει να συνεχίσει είτε στο μπλοκ THEN είτε στο μπλοκ ELSE· αυτό επιτυγχάνεται με μιάν εντολή διακλάδωσης, η οποία πρέπει να οδηγήσει τον επεξεργαστή στο μπλοκ ELSE (που είναι σε απόσταση) εάν η συνθήκη είναι ψευδής, αλλιώς θα συνεχίσει στο μπλοκ THEN (που είναι τοποθετημένο "κολλητά από κάτω"). Εάν εκτελεστεί το μπλοκ THEN, τότε μετά το τέλος του πρέπει να *παρακάμψει* ο επεξεργαστής το μπλοκ ELSE, άρα χρειαζόμαστε μιάν εντολή άλματος (χωρίς συνθήκη) που να μας οδηγήσει κατευθείαν στο μπλοκ CONT, δηλαδή αμέσως μετά την έξοδο από το if-then-else. Αντίθετα, εάν εκτελεστεί το μπλοκ ELSE, τότε μετά το τέλος του βγαίνουμε από το if-then-else και συνεχίζουμε κανονικά στο μπλοκ CONT, που είναι τοποθετημένο "κολλητά" από κάτω, χωρίς να απαιτείται προς τούτο καμιά διακλάδωση ή άλμα. Στο βιβλίο υπάρχει το εξής παράδειγμα (με μία προσθήκη, εδώ, κώδικα "CONT" (μηδενισμός της μεταβλητής h) γιά περισσότερη σαφήνεια):

```
if ( i==j ) { f = g+h; } else { f = g-h; } h=0;
```

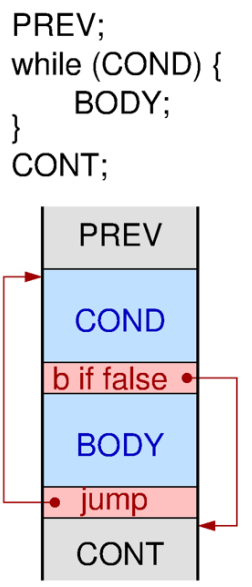
όπου οι μεταβλητές f, g, h, i, j θεωρείται ότι βρίσκονται στους καταχωρητές x19, x20, x21, x22, και x23, αντίστοιχα. Τότε αυτό μεταφράζεται σε Assembly του RISC-V ως εξής (τα ονόματα των labels είναι εμπνευσμένα από την αρίθμηση της παραγράφου, §5.3) (η (ψεύδο-)εντολή "j" (jump) υλοποιεί το άλμα που χρειαζόμαστε, και θα μιλήσουμε γιά αυτήν παρακάτω, στην §5.7):

```
            ...{PREV}...            # instructions before entering into "if"
            bne  x22, x23, else53   # if ( i!=j ) goto ELSE
            add  x19, x20, x21      # f = g+h; {THEN}
            j    cont53             # skip over "ELSE"
    else53: sub  x19, x20, x21      # f = g-h; {ELSE}
    cont53: add  x21, x0, x0        # h=0;  {CONT}
```

5.4 Μετάφραση των Βρόχων While

Κατ' αναλογία προς το προηγούμενο, στο εδώ σχήμα φαίνεται η μετάφραση σε Assembly ενός βρόχου while. Και πάλι, τα μπλοκ PREV και CONT είναι ό,τι υπάρχει πριν και μετά το βρόχο. Το μπλοκ COND είναι η συνθήκη γιά να εκτελέσουμε το σώμα, BODY, και γιά την πρώτη και γιά όλες τις (τυχόν) υπόλοιπες φορές. Η τοποθέτηση του κώδικα συνθήκης, COND, πρώτα και ύστερα του σώματος, BODY, είναι αυτή που βολεύει τον compiler, δηλαδή η ίδια με τη σειρά που αυτός τα διαβάζει από τον πηγαίο κώδικα –δείτε όμως και την αμέσως επόμενη άσκηση 5.5. Η τοποθέτηση των εντολών στη μνήμη, όπως στο κάτω μέρος του σχήματος, ακολουθεί τη λογική του βρόχου: αμέσως μετά την εκτέλεση των εντολών PREV, ο επεξεργαστής συναντά "κολλητά" τον υπολογισμό της συνθήκης COND, και αρχίζει *πάντα* με αυτόν. Εάν η συνθήκη είναι ψευδής – είτε την πρώτη είτε οιαδήποτε επόμενη φορά– τότε βγαίνουμε από το while μέσω της εντολής διακλάδωσης, και συνεχίζουμε με το CONT· αλλιώς, δηλαδή εάν η διακλάδωση αποτύχει, άρα όταν η συνθήκη είναι αλληθής, συνεχίζουμε ακριβώς από κάτω, δηλαδή μπαίνουμε στο σώμα

του βρόχου και εκτελούμε τον κώδικα BODY. Μετά την εκτέλεση του σώματος BODY, πρέπει *πάντα* να ρωτήσουμε "να το ξανακάνω;", δηλαδή πρέπει πάντα να επανυπολογίσουμε τη συνθήκη COND γιά να δούμε εάν θα επαναλάβουμε το βρόχο ή θα βγούμε από αυτό. Γι' αυτό, στο τέλος του BODY τοποθετούμε μιάν εντολή άλματος που μας γυρίζει πάντα πίσω στο COND· προφανώς, η (μόνη) έξοδος από αυτό το βρόχο είναι όποτε πετύχει η διακλάδωση, δηλαδή όταν η συνθήκη COND γίνει ψευδής.



Στο βιβλίο υπάρχει το παρακάτω παράδειγμα, ελαφρώς παραλλαγμένο εδώ. Ψάχνουμε σε έναν πίνακα ακεραίων double: `long long int table[SIZE]` μέχρι να βρούμε εκεί μίαν επιθυμητή τιμή `value`, η οποία υποτίθεται ότι ξέρουμε ότι υπάρχει στον πίνακα. Η μεταβλητή i βρίσκεται, όπως και πριν, στον καταχωρητή x22, η τιμή value που αναζητούμε βρίσκεται στον x24, και η διεύθυνση βάσης του πίνακα (που ως συνήθως *δεν* χωρά στη 12-μπιτη σταθερά Imm12 της εντολής ld) βρίσκεται στον καταχωρητή x25. Αφού τα στοιχεία του πίνακα είναι double words, μεγέθους 8 Bytes καθένα, το στοιχείο `table[i]` θα βρίσκεται στη διεύθυνση: *x25 + 8·i*.

```
i=0;  while (table[i] != value)  { i = i+1; }
```

Θα χρησιμοποιήσουμε τον καταχωρητή x10 γιά τα ενδιάμεσα, προσωρινά αποτελέσματά μας. Η εντολή `slli` (shift left logical immediate) προκαλεί αριστερή ολίσθηση των bits του πρώτου τελεστέου πηγής (εδώ του x22) κατά το πλήθος θέσεων (bits) που ορίζει ο δεύτερος (σταθερός - immediate) τελεστέος πηγής (εδώ: 3 θέσεις), γεμίζοντας τις κενούμενες θέσεις με μηδενικά ("logical"), και γράφει το αποτέλεσμα στον καταχωρητή προορισμού (εδώ: x10)· όπως ξέρουμε (§6.1 Ψηφιακής Σχεδίασης), μιά τέτοια αριστερή ολίσθηση ισοδυναμεί με πολλαπλασιασμό επί 2 είς την δύναμη του πλήθους των θέσεων ολίσθησης –άρα εδώ πολλαπλασιασμός επί $2^3 = 8$. Έτσι, ο παραπάνω κώδικας μεταφράζεται σε:

```
        add  x22, x0, x0       # i=0;
loop54: sll  x10, x22, 3       # tmp = 8 * i (start COND evaluation)
        add  x10, x25, x10     # tmp = διεύθυνση του table[i]
        ld   x10, 0(x10)       # tmp = table[i]
        beq  x10, x24, cont54  # if table[i] == value goto cont54 (loop exit)
        addi x22, x22, 1       # i = i+1 (loop BODY)
        j    loop54            # repeat the loop
cont54: ....                   # contine with the rest of the program
```

Άσκηση 5.5: Βελτιστοποίηση Βρόχου "While"

Στο παράδειγμα που δόθηκε αμέσως παραπάνω, στην §5.4, παρατηρήστε ότι σε κάθε ανακύκλωση αυτού του βρόχου εκτελούνται τόσο μια διακλάδωση υπό συνθήκη (branch), όσο και ένα άλμα χωρίς συνθήκη (jump). Όμως, μόνον οι απλοί μεταφραστές θα έφτιαχναν τέτοιο κώδικα, δεδομένου ότι αυτός μπορεί να βελτιωθεί –οι συνηθισμένοι μεταφραστές παράγουν κώδικα που τρέχει πιο γρήγορα στη *συνηθισμένη περίπτωση*. Η συνηθισμένη περίπτωση είναι ο βρόχος να επαναλαμβάνεται αρκετές φορές πριν βγούμε από αυτόν –γύρω στις 10 φορές κατά μέσον όρο, λένε οι στατιστικές.

Ξαναγράψτε το βρόχο σε Assembly, ούτως ώστε **μόνο ένα** branch ή jump να εκτελείται σε κάθε ανακύκλωση. Κατά την είσοδο ή την έξοδο από το βρόχο επιτρέπεται να εκτελούνται δύο εντολές μεταφοράς ελέγχου –αυτό που μας ενδιαφέρει είναι να γλυτώνουμε τη μια από αυτές κατά τις υπόλοιπες επαναλήψεις του βρόχου, που αποτελούν και την πλειοψηφία των φορών που αυτός εκτελείται.

• Έστω ότι το σώμα (BODY) αυτού του βρόχου επαναλαμβάνεται (εκτελείται) 10 φορές. Τότε, πόσες εντολές συνολικά εκτελούνταν με τον παλαιό κώδικα, και πόσες συνολικά με τον νέο;

5.6 Σύνθετες Συνθήκες με υπολογισμό Short-Circuit

Ας θεωρήσουμε τώρα μιά λίγο πιό σύνθετη συνθήκη διακλάδωσης, με λογικό ΚΑΙ μέσα της. Θεωρήστε ένα παράδειγμα ανάλογο με το προηγούμενο, όπου ψάχνουμε να βρούμε κάποιο "value", αλλά τώρα το ψάχνουμε σε linked list αντί σε array· επιπλέον, τώρα ενδέχεται και να μην υπάρχει το στοιχείο αυτό στη λίστα, άρα πρέπει να προβλέψουμε και την περίπτωση που η λίστα τελειώνει πριν το βρούμε. Στον παρακάτω κώδικα, έστω ότι η τιμή value που αναζητούμε βρίσκεται και πάλι στον καταχωρητή x24, και ότι ο pointer **p** βρίσκεται στον καταχωρητή x25 (στη θέση της διεύθυνσης βάσης του προηγούμενου πίνακα). Ο p δείχνει σε structures με δύο στοιχεία καθένα, μεγέθους 8 Bytes καθένα επειδή βρισκόμαστε σε 64-μπιτο RISC-V: το πρώτο στοιχείο είναι μία τιμή "key" σε offset 0 σε σχέση με την αρχή του structure, και το δεύτερο στοιχείο είναι ένας pointer "next" στο επόμενο structure της λίστας, σε offset 8 σε σχέση με την αρχή του structure.

```
while ( p!=NULL && p->key != value)  { p = p->next; }
```

Εδώ, ως γνωστόν, τα semantics της C μας βοηθάνε να *μην* προσπαθήσει ο επεξεργαστής να διαβάσει το p->key όταν ο pointer p τύχει να είναι NULL: η C προδιαγράφει "short circuit evaluation" του λογικού ΚΑΙ, δηλαδή εάν είναι ψευδές το πρώτο σκέλος, p!=NULL, τότε παρακάμπτεται και *δεν* υπολογίζεται το δεύτερο σκέλος, p->key != value. Γιά να υλοποιηθεί αυτό, απαιτούνται *δύο* διακλαδώσεις υπό συνθήκη σε αυτό το βρόχο:

```
loop56: beq  x25, x0, cont56  # if p==NULL then exit the loop
        ld   x10, 0(x25)      # tmp = p->key
        beq  x10, x24, cont56 # if p->key == value then exit the loop
        ld   x25, 8(x25)      # p = p->next (loop BODY)
        j    loop56           # repeat the loop
cont56: ....                  # contine with the rest of the program
```

5.7 Άλματα χωρίς συνθήκη στον RISC-V

Όπως είπαμε, οι εντολές μεταφοράς ελέγχου (CTI, control transfer instructions) κάνουν ώστε να εκτελεστεί σαν επόμενη εντολή μια έντολη άλλη από την "από κάτω" τους εντολή. Όταν αυτή η μεταφορά ελέγχου γίνεται υπό συνθήκη τότε αυτές λέγονται διακλαδώσεις (branches), ενώ όταν γίνεται χωρίς συνθήκη, δηλαδή πάντα, τότε αυτές λέγονται **άλματα** (jump). Η απλή εντολή άλματος χρειάζεται να περιέχει μέσα της μόνον έναν τελεστέο πέραν από τον opcode –τη διεύθυνση προορισμού– άρα το instruction format θα μπορούσε να αφιερώσει πολλά bits γιά τον προσδιορισμό αυτής της διεύθυνσης. Ο RISC-V χρησιμοποιεί και γιά τα άλματα το PC-Relative addressing mode, γιά τους ίδιους λόγους (πλεονεκτήματα) που το χρησιμοποιεί και γιά τις διακλαδώσεις. Δεδομένου πάντως ότι τα απλά άλματα (δηλ. όχι τα καλέσματα διαδικασίας) χρησιμοποιούνται μόνον στα if-then-else και βρόχους, δηλαδή *μέσα* στις διαδικασίες, και μαζί με τη χρήση PC-relative mode, δεν υπάρχει σοβαρή ανάγκη γιά πολλά bits στη σχετική απόσταση που έχει να κωδικοποιήσει η σταθερή ποσότητα μέσα στην εντολή άλματος.

Έτσι, ο RISC-V καταλήγει να μην χρειάζεται (και *δεν* έχει) ειδική εντολή γιά το απλό άλμα (jump), αλλά υλοποιεί τέτοια απλά άλματα σαν ειδικές περιπτώσεις μιάς γενικότερης εντολής, της jump-and-link (jal), η οποία, όπως θα πούμε στο επόμενο κεφάλαιο (6η σειρά διαλέξεων), χρησιμοποιείται γιά κάλεσμα διαδικασιών, και έχει έναν καταχωρητή προορισμού. Η απλή jump κάνει την ίδια δουλειά με εκείνην, αλλά δεν χρειάζεται να γράψει τίποτα στον καταχωρητή προορισμού, άρα απλώς χρησιμοποιεί τον x0 σε εκείνη τη θέση (αφού ο x0 δεν αλλάζει ποτέ τιμή, και μένει πάντα 0, όσο και αν "γράφουμε" σε αυτόν). Έτσι, τόσο η "`jal rd, label`", όσο και η *ψευδοεντολή* "**`j`** `label`" (απλή **jump**) που είναι ισοδύναμη με "`jal x0, label`" χρησιμοποιούν και οι δύο το J-format (§4.3), άρα προσδιορίζουν τη διεύθυνση προορισμού τους μέσω της 20-μπιτης σταθερής ποσότητας Imm20. Ο τρόπος που χρησιμοποιείται αυτό το πεδίο γιά να αλλάξει τον PC είναι πανομοιότυπος με εκείνον γιά τις διακλαδώσεις –απλώς εδώ το Immediate έχει 20 bits αντί 12 εκεί:

- `PC_new` ← `PC_j + (2*(signed)Imm20)`

όπου `PC_j` είναι η διεύθυνση της ίδιας της εντολής άλματος, `Imm20` είναι η σταθερή ποσότητα

των 20 bits μέσα στην εντολή, θεωρούμενη ως προσημασμένος αριθμός σε συμπλήρωμα ως προς 2 (δηλαδή sign-extended), και `PC_new` είναι η διεύθυνση της εντολής όπου πηδάμε. Όπως και με τις διακλαδώσεις, ο πολλαπλασιασμός επί 2 αξιοποιεί το γεγονός ότι ο PC έχει τιμή πάντοτε ακέραιο πολλαπλάσιο του 2, κι έτσι διπλασιάζει το βεληνεκές των αλμάτων σε $\pm 2^{19}$ = 512 K half-words (μικρές, 16-μπιτες εντολές) = ±1 MByte.

Indirect (indexed) Jump – `jr rs1` (ψευδοεντολή):

Εκτός από άλματα σε σταθερές, πάντα τις ίδιες διευθύνσεις, οι υπολογιστές χρειάζονται και άλματα σε **μεταβλητούς** (runtime variable) προορισμούς, γιά μία σειρά από σκοπούς, όπως αναλύουμε εδώ. Στον RISC-V, τον ρόλο αυτό τον παίζει η ψευδοεντολή **`jr` (jump-register)**· όπως και με την απλή jump, η jump-register συντίθεται σαν ειδική περίπτωση της γενικότερης jump-and-link-register που θα δούμε στο επόμενο κεφάλαιο. Η `jr rs1` κάνει την εξής απλή αλλά εξαιρετικά "ισχυρή" λειτουργία:
• `PC_new ← rs1`
με άλλα λόγια κάνει ώστε η επόμενη εντολή να είναι η εντολή στην *οιαδήποτε αυθαίρετη* διεύθυνση μνήμης είχαμε υπολογίσει προηγουμένως και είχαμε τοποθετήσει στον καταχωρητή rs1. Η εντολή αυτή μας επιτρέπει να μεταφέρουμε τον έλεγχο (την εκτέλεση του προγράμματος) σε αυθαίρετη θέση μνήμης, η οποία μπορεί και να ποικίλει κατά την εκτέλεση του προγράμματος (run-time variable) και πιθανόν να εξαρτάται και από τα δεδομένα (data dependent). Χρησιμοποιείται γιά:
• Επιστροφή από διαδικασία, όπως θα πούμε στο επόμενο κεφάλαιο.
• Μεταφορά του ελέγχου οσοδήποτε μακριά, αφού μπορούμε με προηγούμενες εντολές να φορτώσουμε την οποιαδήποτε διεύθυνση θέλουμε στον rs1.
• γιά μετάφραση του **`switch`** statement, όπως περιγράφουμε επιγραμματικά εδώ:

Γιά το `switch` statement, όπως περιγράφεται στο τέλος της §2.7 του βιβλίου, η βασική ιδέα είναι η εξής: ο κώδικας της κάθε περίπτωσης (`case`) γράφεται σε κάποια περιοχή μνήμης, και ο compiler θυμάται σε ποιά διεύθυνση αρχίζει αυτός ο κώδικας, γιά την καθεμία περίπτωση. Στη συνέχεια, ο compiler κατασκευάζει έναν πίνακα (array), τον "jump table", ο οποίος περιέχει σε κάθε θέση του αυτές τις διευθύνσεις όπου αρχίζει ο κώδικας της κάθε περίπτωσης, με τη σειρά, γιά όλες τις (αριθμητικές) τιμές της κάθε περίπτωσης. Όταν είναι να εκτελεστεί το `switch` statement, παίρνουμε την τιμή της μεταβλητής του switch (αυτήν της οποίας την κάθε περίπτωση τιμών εξετάζουμε), και χρησιμοποιούμε αυτή την τιμή σαν **index** στον jump table, γιά να επιλέξουμε ένα από τα περιεχόμενα του jump table –αυτό που αντιστοιχεί στην περίπτωση της τωρινής τιμής της μεταβλητής του switch. Το στοιχείο του πίνακα που διαλέξαμε είναι η διεύθυνση του κώδικα που πρέπει να εκτελεστεί γιά την περίπτωση της μεταβλητής που τώρα μας ήλθε. Αυτό ακριβώς το στοιχείο του jump table το φέρνουμε σε έναν καταχωρητή R, και στη συνέχεια εκτελούμε την εντολή `jr R`, η οποία κάνει ώστε η επόμενη εντολή που θα εκτελεστεί να είναι εκεί που δείχνει ο R, δηλαδή εκεί που μας είπε ο jump table να πάμε γιά την περίπτωση της τιμής που τώρα είχε η μεταβλητή.

5.8 Πράξεις Σύγκρισης με αποτέλεσμα Boolean

Εκτός από τα if-then-else, στις γλώσσες προγραμματισμού, αριθμητικές συγκρίσεις εμφανίζονται και σε εκχωρήσεις (assignments) του τύπου: `myBool = (a<b)`, όπου οι μεν μεταβλητές a και b είναι αριθμοί, το δε αποτέλεσμα της σύγκρισης που εκχωρείται είναι τύπου *Boolean*. Στην C, ο τύπος Boolean κρατιέται μεν μέσα σε μεταβλητές τύπου ακεραίου, αλλά οι μόνες νόμιμες τιμές του είναι το 0 γιά "ψευδές" και το 1 γιά "αλληθές" (αυτό έχει πολλά μηδενικά αριστερά, και έναν μόνον άσο δεξιά). Εκχωρήσεις όπως η παραπάνω θα μπορούσαν βέβαια να υλοποιηθούν μέσω `if (a<b) {myBool=1;} else {myBool=0;}`, όμως μιά τέτοια υλοποίηση αφ' ενός θα απαιτούσε κάμποσες εντολές, και αφ' ετέρου θα προκαλούσε σημαντικές καθυστερήσεις κατά την εκτέλεση του προγράμματος επειδή οι εντολές διακλάδωσης είναι αργές ή και δύσκολες στους επεξεργαστές με ομοχειρία (pipelining) όπως θα δούμε αργότερα στο μάθημα. Γιά να υποστηρίξει εκχωρήσεις όπως η παραπάνω, ο RISC-V έχει την εντολή **set if less than (slt)**, μαζί με άλλες τρείς παραλλαγές της, που υλοποιεί ουσιαστικά την παραπάνω εκχώρηση. Πρόκειται γιά εντολές ανάλογες προς την πρόσθεση ή την αφαίρεση, μόνο που το αποτέλεσμά τους είναι τύπου Boolean αντί τύπου ακέραιος:

- **slt rd, rs1, rs2** # rd ← (rs1<rs2) – signed comparison, Boolean result
- **sltu rd, rs1, rs2** # rd ← (rs1<rs2) – unsigned comparison, Boolean result
- **slti rd, rs1, Imm12** # rd ← (rs1<Imm12) – signed Immediate, signed comparison
- **sltiu rd, rs1, Imm12** # signed Immediate (sign-extended), but unsigned comparison

Οι εντολές αυτές γράφουν στον καταχωρητή rd είτε τον αριθμό 1 (πολλά μηδενικά bits αριστερά και ένας άσος δεξιά) εάν το αποτέλεσμα της σύγκρισης είναι αληθές, είτε τον αριθμό 0 (όλο μηδενικά bits) εάν το αποτέλεσμα είναι ψευδές. Οι δύο πρώτες παραπάνω εντολές συγκρίνουν δύο καταχωρητές μεταξύ τους, και ακολουθούν το R-format, οι δε δύο επόμενες συγκρίνουν καταχωρητή (αριστερά από το <) προς σταθερή ποσότητα (δεξιά από το <), και ακολουθούν το I-format. Η μεν πρώτη και η τρίτη θεωρούν τους τελεστέους σαν προσημασμένους (signed) ακεραίους όταν κάνουν τη σύγκριση, οι δε δεύτερη και τέταρτη κάνουν απρόσημη (unsigned) σύγκριση. Στον RISC-V οι σταθερές ποσότητες "Immediate" θεωρούνται πάντα προσημασμένες, άρα το Imm12 επεκτείνεται από 12 σε 32 ή 64 bits πάντα σαν προσημασμένη (signed) σταθερά, ακόμα και όταν η σύγκριση, στη συνέχεια, είναι unsigned, δηλαδή θεωρεί τον 32-μπιτο ή 64-μπιτο δεύτερο τελεστέο σαν unsigned. Όσον αφορά εκφράσεις όπως `(a<b)&&(c<d)`, είτε σε εκχωρήσεις Booolean είτε μέσα σε if, εάν μεν η γλώσσα ή οι περιστάσεις δεν απαιτούν short-circuit evaluation τότε αυτές μπορούν να υλοποιηθούν με δύο εντολές set-if-less-than ακολουθούμενες από μία εντολή (bit-wise) AND, που δίνει συνήθως και την γρηγορότερη εκτέλεση· εάν, από την άλλη, απαιτείται short-circuit evaluation, τότε θα χρειαστεί και η χρήση διακλαδώσεων.

Άσκηση 5.9: Σύνθεση άλλων Συγκρίσεων

Ο RISC-V έχει μόνον τις παραπάνω τέσσερεις εντολές συγκρίσεων, που όλες τους κάνουν σύγκριση "less than", και δεν έχει συγκρίσεις ≤ ή > ή ≥ επειδή αυτές μπορούν να συντεθούν από τη σύγκριση less than –ας δούμε πώς. Έστω ότι η μεταβλητή **i** βρίσκεται στον καταχωρητή **x22** η μεταβλητή **j** στον καταχωρητή **x23**, η μεταβλητή **B** στον καταχωρητή **x26**, και ότι **CONST** σημαίνει μια αυθαίρετη προσημασμένη σταθερή ποσότητα που χωρά σε 12 bits. Συνθέστε τις παρακάτω εκχωρήσεις, όπου όλες θεωρούνται signed, χρησιμοποιώντας την εντολή **slt** ή την **slti**, και, όπου χρειάζεται, και την εντολή **xori** (ανάλογη της addi, αλλά κάνει exclusive-OR αντί γιά πρόσθεση) με σταθερή immediate τον αριθμό 1, η οποία επομένως έχει σαν συνέπεια να αντιστρέφει το δεξιό bit του άλλου τελεστέου της, δηλαδή ισοδυναμεί με Boolean NOT:

i. `B = ( i <  j )` (μικρότερο)
ii. `B = ( i >= j )` (μεγαλύτερο ή ίσο)
iii. `B = ( i >  j )` (μεγαλύτερο)
iv. `B = ( i <= j )` (μικρότερο ή ίσο)
v. `B = ( i <  CONST )` (μικρότερο)
vi. `B = ( i >= CONST )` (μεγαλύτερο ή ίσο)
vii. `B = ( i >  CONST )` (μεγαλύτερο)
viii. `B = ( i <= CONST )` (μικρότερο ή ίσο)

Υπόδειξη: παίξτε πρώτα με τη σειρά που βάζετε τους δύο τελεστέους πηγής όταν αυτοί είναι καταχωρητές, καθώς και με ενδεχόμενη άρνηση του αποτελέσματος. Όταν ο δεύτερος τελεστέος είναι σταθερή ποσότητα, τότε δεν μπορείτε μεν να αντιστρέψετε τη σειρά που τους δίνετε στην εντολή slti, μπορείτε όμως να παίξετε με τις σταθερές `CONST, CONST+1, CONST-1`. Θα εκτιμήστε το πόσα πολλά μπορεί να κάνει το software, εκμεταλλευόμενο λίγους, προσεκτικά επιλεγμένους δομικούς λίθους hardware!

Στη συνέχεια παρατηρήστε ότι η εντολή: **sltu rd, x0, rs2** ελέγχει εάν ο rs2 είναι μεγαλύτερος του μηδενός, *αλλά* τον συγκρίνει ως unsigned, δηλαδή σαν πάντα ≥0, άρα ουσιαστικά ελέγχει εάν ο rs2 είναι **διάφορος του μηδενός**. Επίσης η εντολή: **sltiu rd, rs1, 1** ελέγχει εάν ο rs1 είναι μικρότερος του 1, *αλλά* τον συγκρίνει ως unsigned, δηλαδή σαν πάντα ≥0, άρα ουσιαστικά ελέγχει εάν ο rs1 είναι **ίσος με μηδέν**.

Με βάση αυτές τις παρατηρήσεις, χρησιμοποιήστε μιάν εντολή αφαίρεσης (sub) ή πρόσθεσης του `-CONST`, και στη συνέχεια μιάν από τις παραπάνω συγκρίσεις ισότητας/ανισότητας προς μηδέν, γιά να συνθέσετε και τις ππαρακάτω εκχωρήσεις. Σαν προσωρινό καταχωρητή παρατηρήστε ότι

μπορείτε να χρησιμοποιήσετε τον ίδιο τον καταχωρητή προορισμού, και ότι αυτό θα ίσχυε ακόμα και εάν τύχαινε ο καταχωρητής αυτός να ήταν ο ίδιος με έναν από τους καταχωρητές πηγής. Δώστε τις απαντήσεις σας γραπτά.

a. `B = ( i == j )` (ίσο)
b. `B = ( i != j )` (άνισο)
c. `B = ( i == CONST )` (ίσο)
d. `B = ( i != CONST )` (άνισο)

Τρόπος Παράδοσης:
Παραδώστε μαζί την προηγούμενη άσκηση 4 και αυτήν εδώ τη σειρά 5, on-line, σε μορφή **PDF** (μόνον) (μπορεί να είναι κείμενο μηχανογραφημένο ή/και "σκαναρισμένο" χειρόγραφο, αλλά *μόνον* σε μορφή PDF). Παραδώστε μέσω **`turnin ex045@hy225 [directoryName]`** ένα αρχείο ονόματι **`ex045.pdf`** που θα περιέχει τις απαντήσεις σας σε όλες τις ασκήσεις, 4 και 5.
