

Ιόνιο Πανεπιστήμιο – Τμήμα Πληροφορικής
ΠΜΣ «Δικτυωμένα Συστήματα Μεγάλου Όγκου Δεδομένων»
Μάθημα: «Αλγόριθμοι Βελτιστοποίησης και Παράλληλη Επεξεργασία»
Μέρος Β΄: Παράλληλη Επεξεργασία
2022-23

Αναδιατάξεις δεδομένων επεξεργασίας

(οι λειτουργίες gather και scatter)

<http://mixstef.github.io/courses/pms-parcomp/>

Μ.Στεφανιδάκης

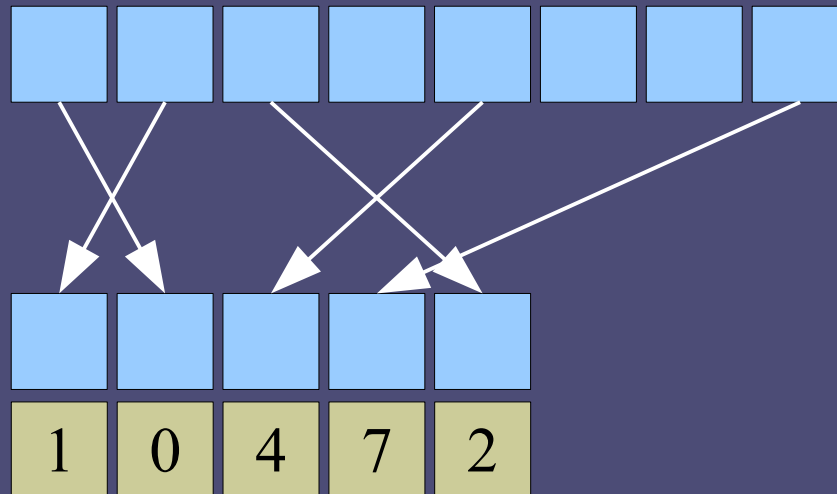


Αναδιάταξη δεδομένων

- Στη διάρκεια της παράλληλης/κατανεμημένης επεξεργασίας
 - Μεταξύ των φάσεων υπολογισμού
 - Μετακίνηση δεδομένων για τη διευκόλυνση του επόμενου σταδίου υπολογισμού
 - Η μετακίνηση των δεδομένων επηρεάζει καθοριστικά την απόδοση
 - Πολλές φορές περισσότερο από τον ίδιο τον υπολογισμό
 - Η αναδιάταξη των δεδομένων είναι το σημαντικότερο μέρος του φόρτου στην αναζήτηση και την ταξινόμηση
 - Παρά ο ίδιος ο υπολογισμός

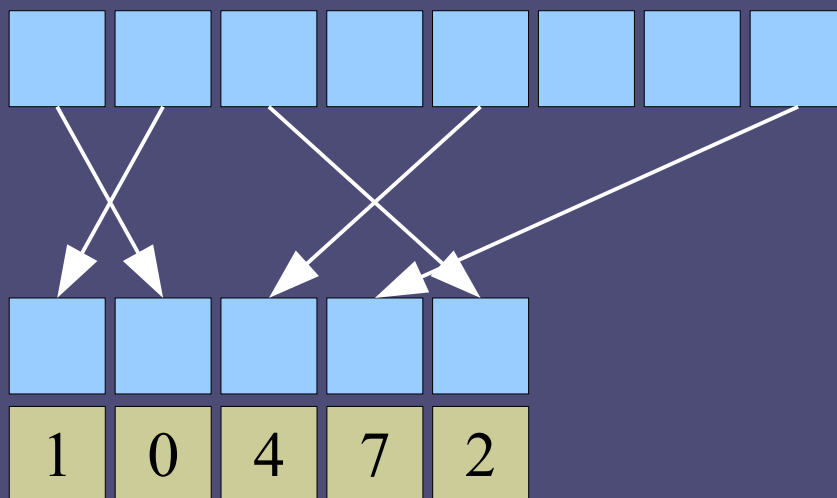
Λειτουργία gather

- Παράλληλες τυχαίες αναγνώσεις
 - Χρησιμοποιώντας έναν πίνακα δεικτών (διευθύνσεων)
 - Τα δεδομένα εξόδου έχουν τιμές από τα δεδομένα εισόδου αλλά τις διαστάσεις του πίνακα δεικτών



Λειτουργία gather

- **Παράλληλες τυχαίες αναγνώσεις**
 - Αναγνώσεις από την ίδια θέση επιτρέπονται
 - Συνήθως ευνοούνται από την οργάνωση του υλικού (π.χ. κρυφές μνήμες)
 - Δεν υπάρχει σύγκρουση κατά την αποθήκευση
 - Ποτέ δεν γράφουμε στην ίδια θέση πάνω από μία φορά

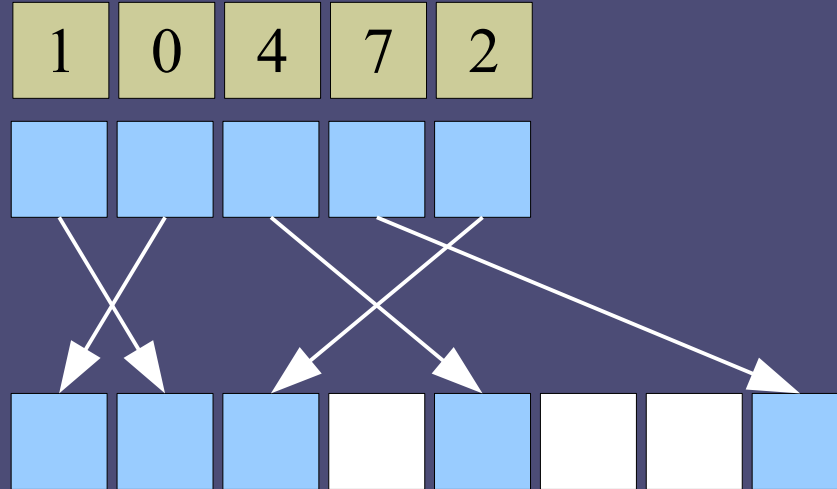


Λειτουργία gather

- **Επεκτάσεις και υποπεριπτώσεις**
 - Η λειτουργία gather επεκτείνεται και σε περισσότερες διαστάσεις
 - Η έξοδος θα έχει τις διαστάσεις του πίνακα δεικτών
 - Ειδικές υποπεριπτώσεις της gather
 - shift: ολίσθηση ή περιστροφή δεδομένων
 - Αποδοτική εκτέλεση με εντολές vector (sse)
 - zip/unzip: συνδυασμός πολλαπλών συλλογών/πινάκων εισόδου σε έναν μοναδικό πίνακα εξόδου

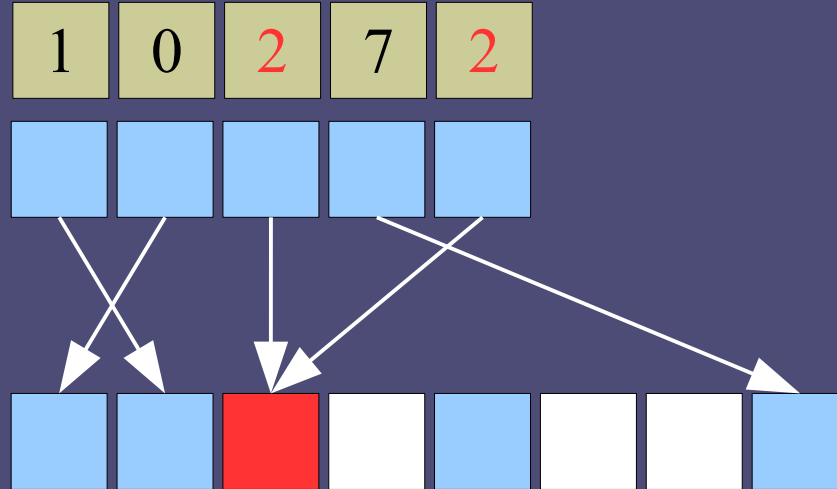
Λειτουργία scatter

- Παράλληλες τυχαίες εγγραφές
 - Χρησιμοποιώντας έναν πίνακα δεικτών (διευθύνσεων)
 - Τι θα συμβεί αν προσδιορίζεται η ίδια θέση εγγραφής πάνω από μία φορά;



Λειτουργία scatter

- Συγκρούσεις κατά την εγγραφή
 - Όταν ο ίδιος δείκτης εμφανίζεται πάνω από μία φορά
 - Πρέπει να διευθετηθεί προγραμματιστικά

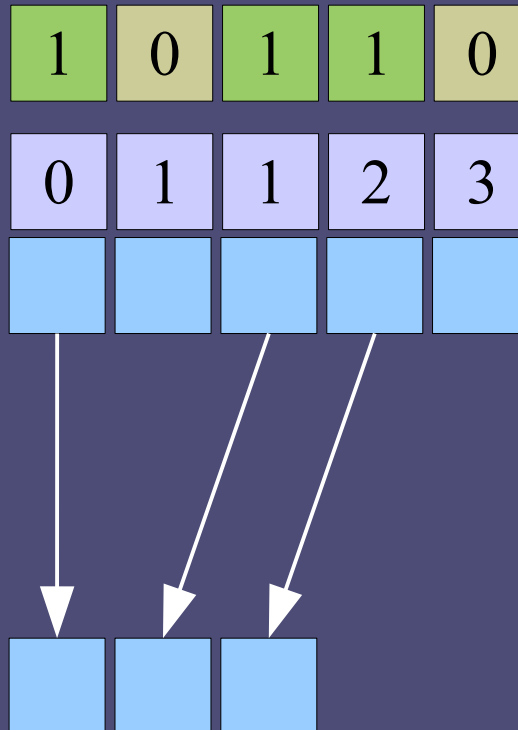


Διευθέτηση συγκρούσεων

- **Παραλλαγές**
 - Απαγόρευση εμφάνισης του ίδιου δείκτη
 - Χρήση πρόσθετης πληροφορίας «προτεραιότητας»
 - Συγχώνευση (merge) τιμών
 - Ατομικές εγγραφές
 - Μία θα επικρατήσει – δεν ξέρουμε ποια (μη ντετερμινιστική εγγραφή)
 - Χρήσιμο σε περιπτώσεις flags
 - Χρειάζεται υποστήριξη από το υλικό
- **Η λειτουργία scatter μπορεί να μετατραπεί σε gather**
 - Αρκεί να ξέρουμε τους δείκτες εκ των προτέρων
 - Με τη βοήθεια πρόσθετου υπολογισμού

Λειτουργία rack

- Συνεχόμενη έξοδος επιλεγμένων στοιχείων
 - Συνδυασμός exclusive scan και scatter υπό συνθήκη
 - Δεν υπάρχει σύγκρουση κατά την εγγραφή



Βιβλιογραφία

- Michael McCool, James Reinders, and Arch Robison. 2012. *Structured Parallel Programming: Patterns for Efficient Computation* (1st ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.