How to Compose Premium-quality Figures for Academic Publications

Po-Chun Huang (黃柏鈞)

Department of Computer Science and Engineering,
Yuan Ze University

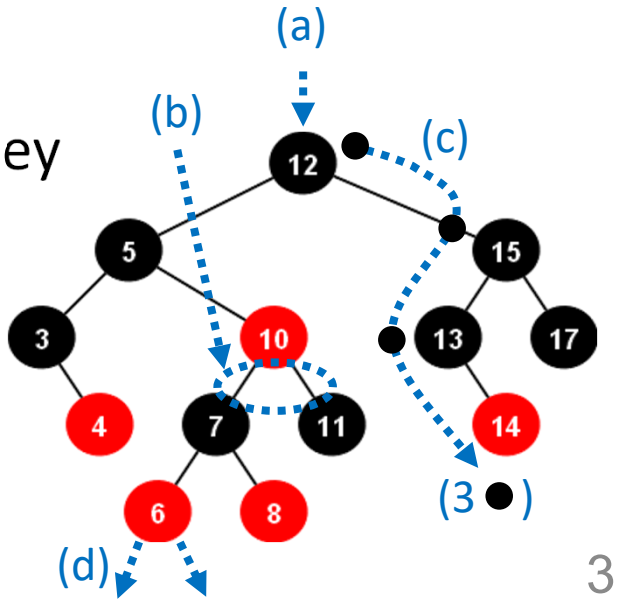
Agenda

- Why this talk?
- Common metaphors for figures in academic publications
- Modeling concepts into figures
- Tricks for space tweaking in short articles
- Editing tools for raster & vector graphics
- Conclusion

Why This Talk?

- Researchers focus on writing quality of papers for better readability.
- Although figures are often a better means to deliver complex ideas, their importance is largely ignored *by researchers*. Why?
  - Related techniques are harder to learn.
  - Senses of aesthetics are not listed as must-have in engineering education. They simply do not recognize aesthetics. ☹
- Motivation: Imagine how you can describe a binary search tree with a paragraph of texts or a figure.

Example: Reb–Black tree



http://cs.lmu.edu/~ray/images/redblacktree.png

Intended Audiences

- Main target audiences: graduate students and research assistants majoring science or engineering
- Other intended audiences: faculties, undergraduate students, and graduate students in other areas
- **Assumptions**
  - You want to make your papers and reports look better.
  - You are not experts in *digital typography* (or you don't really care that).
  - You have limited time to take actions.

Conceptual Visual-modeling Process

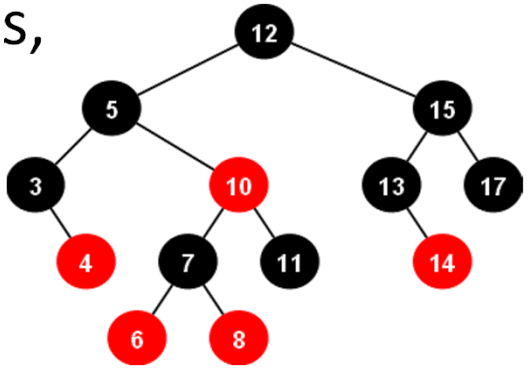
- How to convert the concepts in our mind into graphics and accompanying texts in paper?

http://www.yanahealing.com/Uploads/ckupload/1408936975.jpg; http://marketeer.kapost.com/wp-content/uploads/2014/04/innovative_ideas.jpg; https://www.darencademy.com/images/backend/article/4575b84022785b1d5e0baf0a59dfd22f.jpg

Raster Graphics vs. Vector Graphics

- **Raster graphics (bitmaps)** are suitable for images of experimental environments or image inputs/outputs for computer vision studies.
  - Raster graphics are good at reproducing delicate light and shadow details, but limited by resolution.
  - Typically, no modification of bitmaps is allowed for academic publications.

- **Vector graphics** are suitable for most figures, e.g., system architecture, data structures, working examples, and program flow diagrams.
  - Vector graphics have unlimited resolution and smaller file sizes.
  - Most academic publications welcome vector graphics for most figures in manuscripts (except for experimental environments or image inputs/outputs for computer vision studies).
  - Today we primarily focus on vector graphics.



https://en.wikipedia.org/wiki/Image_tracing; http://cs.lmu.edu/~ray/images/redblacktree.png;

Tools for Editing Raster & Vector Graphics

- **Raster graphics tools**
  - Proprietary: Adobe Photoshop, Corel Photo-Paint, Microsoft PowerPoint, and 小畫家
  - Free: GIMP, Paint.NET, Krita, and Autodesk Pixlr
- **Vector graphics tools**
  - Proprietary: Adobe Illustrator, Corel Designer, CorelDraw, Microsoft PowerPoint, and Canvas
  - Free: Inkscape, LibreOffice Draw, FreeMind, Microsoft Expression, and LaTeXDraw
- **Personal recommendation: ((Adobe Illustrator** OR **Corel Designer** OR **CorelDraw)** AND **PowerPoint)**
  - Not ad! Why?
  - Trial versions available on internet!

Weakness of Each Tool

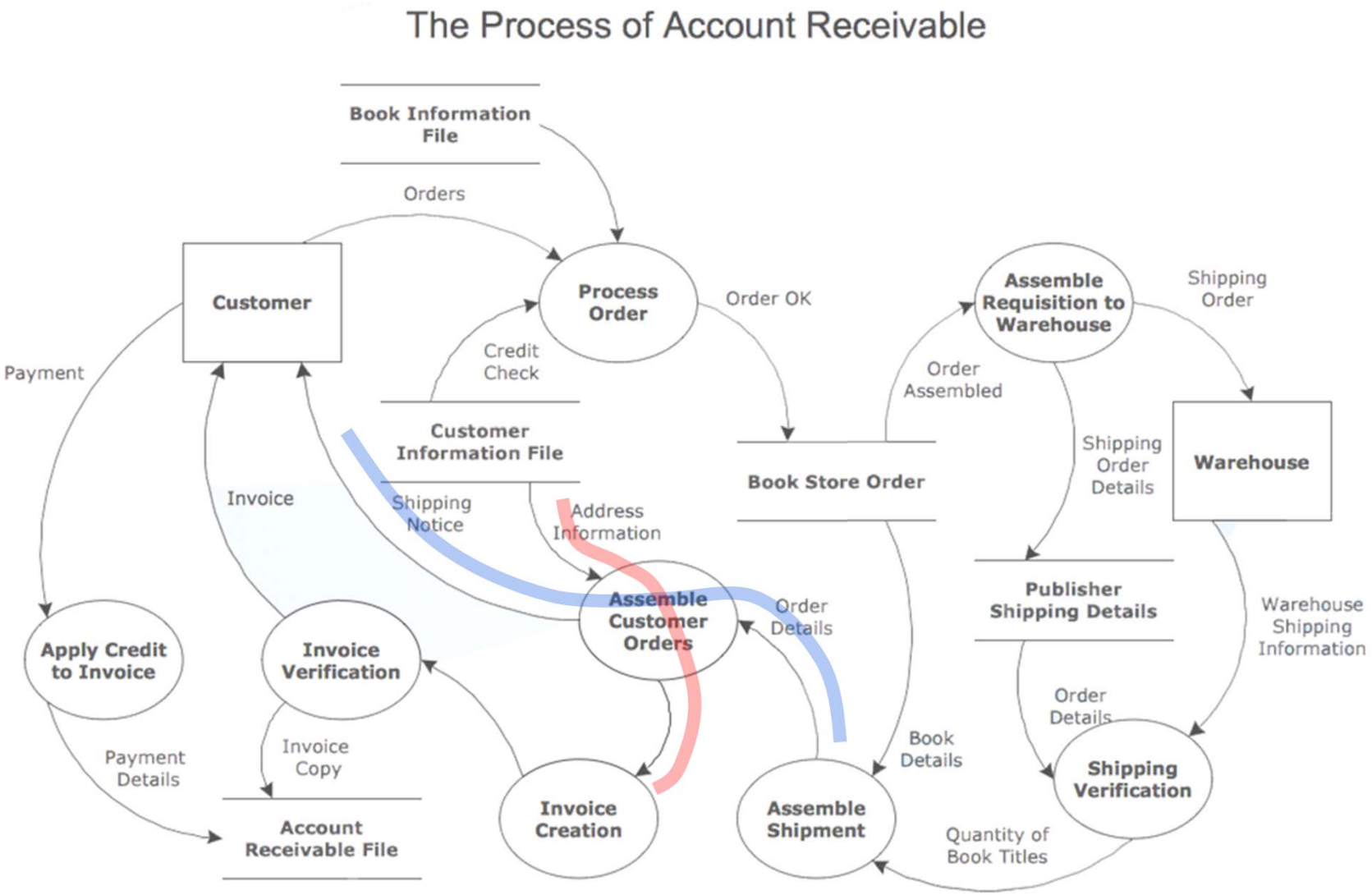
- PowerPoint: The output vector graphics are sometimes ugly, especially for complex curves and shadows!
- CorelDraw and Corel Designer: They have no intrinsic supports for gradient line colors. We need to (stupidly) convert lines to areas before we can apply gradient fills.
- Illustrator: The learning curve is steeper than other tools.

Visual Memories

- Graphics are hashed then remembered; this is the reason of déjà vu (既視感).
- Thus, to impress the readers, graphics must be designed with very distinct “features.”
  - Not necessary but good to have.

http://www.incredibilia.es/wp-content/uploads/2014/10/cuccioli-gemelli-010.jpg

Example of “Visual Memories”

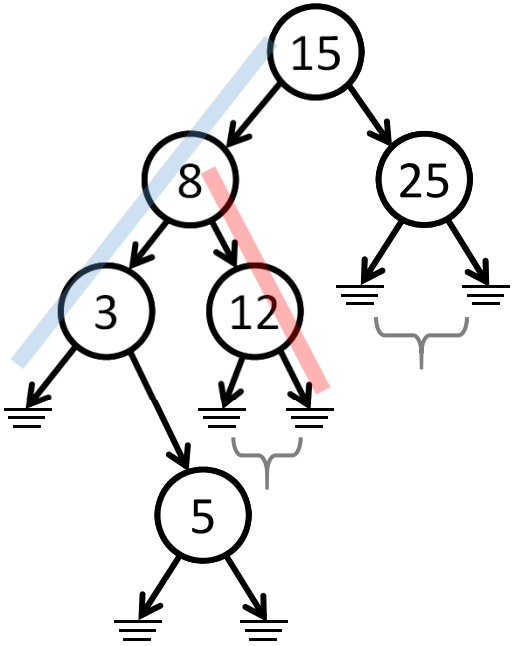


http://www.conceptdraw.com/samples/resource/images/solutions/business-process-diagrams/BUSINESS-PROCESS-DIAGRAMS-Business-Process-Data-Flow-Diagram.png

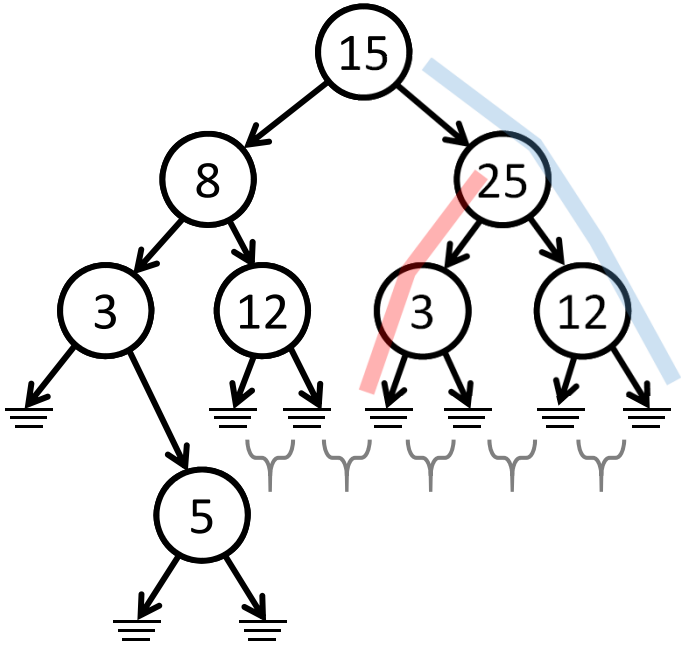
Example of "Visual Memories" (Cont'd)

- Usually there are some compromises…

"Keep it straight"



"Maintain a consistent gap"

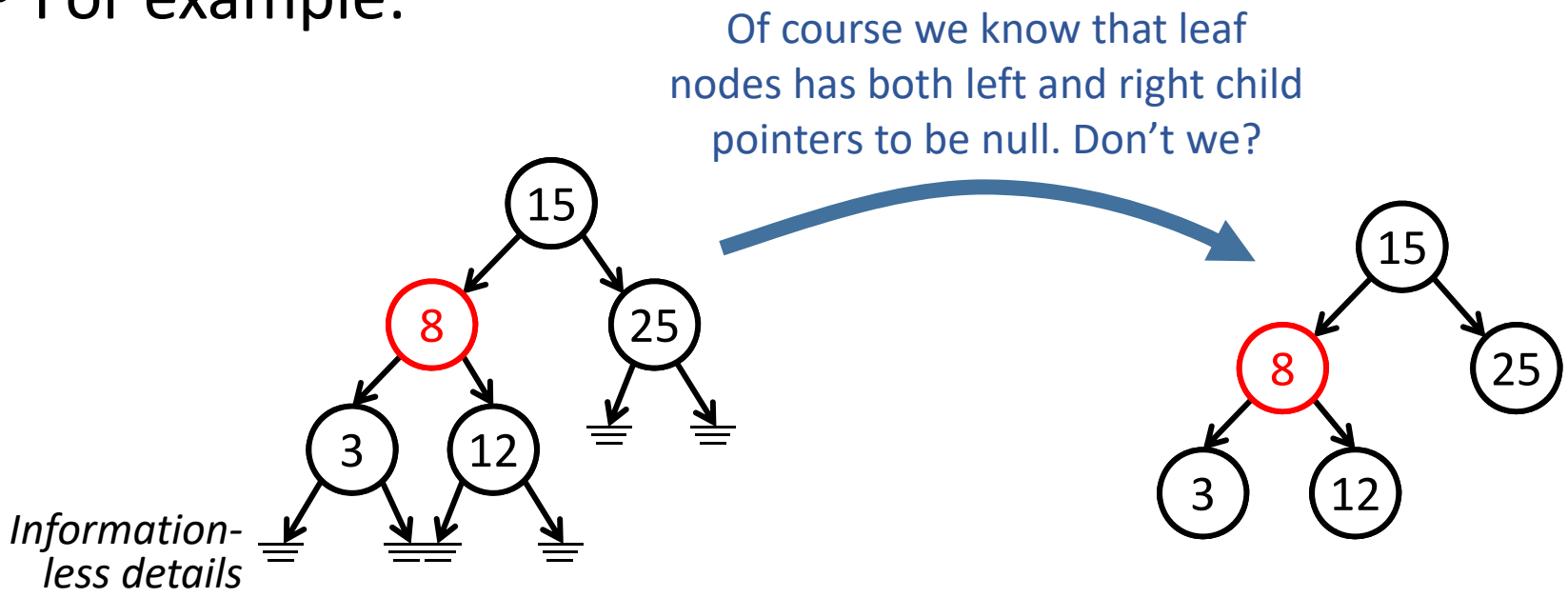


Distinct Characteristics of Figures in Academic Publications

- Common paper size: A4 (210×297 mm) and letter (8.5×11 in. = 215.9×279.4 mm)
- The size of in-text figures are 5×7 in. (full width) or 2.3×3.2 in. (half width). **The heights can be changed, but the widths are limited by 3.2 in. and 7 in.**
- **Many formal academic publications allow only greyscale figures.** (Thus we need to use metaphors to provide implications or emphasis.)
- **Since papers are normally printed out by laser printers, the effects that cannot be printed satisfactorily should not be used, for example, gradient fills.**

Reduce the Signal-to-noise Ratio (SNR)

- The figure space is limited; the patience and attention of readers are even more seriously limited.
- We must minimize the details presented on the figures, so as to maximize the **signal-to-noise ratio (SNR)** of the figure and to enhance comprehensiveness.
- For example:

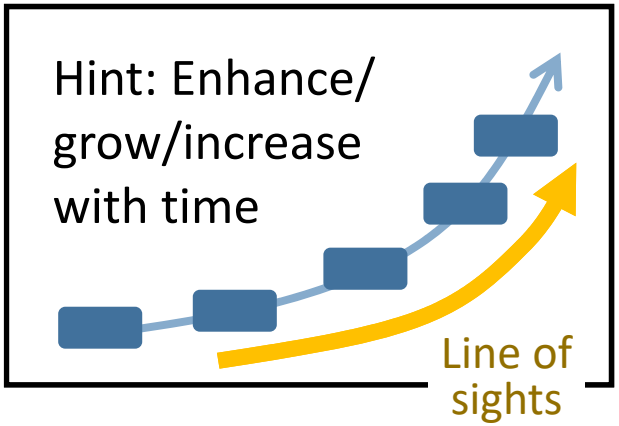


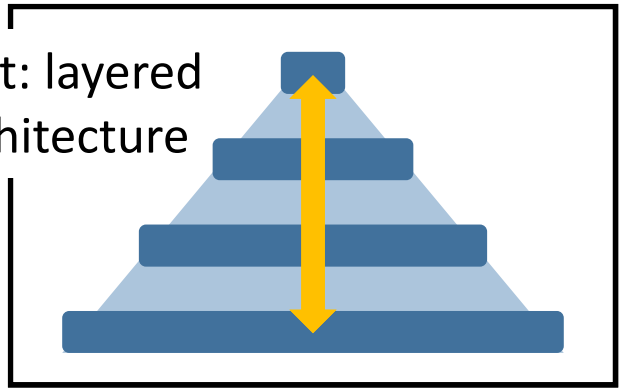
Reduce the SNR: Practices

- Use as few fonts (fonts, typefaces, or sizes) as possible.
- Use as few colors (greyscales) as possible.
- Use metaphors (conventions) to reduce the noises shown by the figure.
  - **Show only what are necessary; in other words,**
  - **Remove everything that are not necessary.**
- Keep texts in figures as concise as possible.

Control the Flow of "Line of Sights"

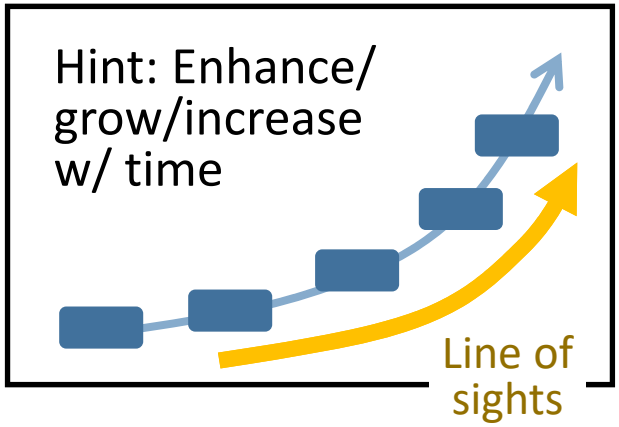
- **Each figure must have a central idea to deliver**, e.g., a flow of programs, hierarchical system architecture, or linked data structure.
  - If not, you'd better think again!
- We should not blindly draw what we have or we want on the figure; instead, we should intentionally use (or avoid) metaphors.

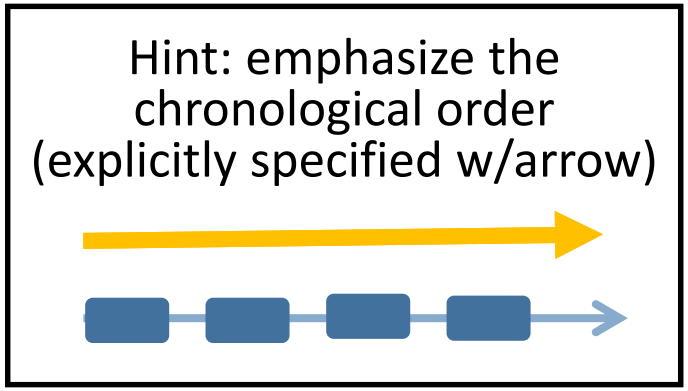


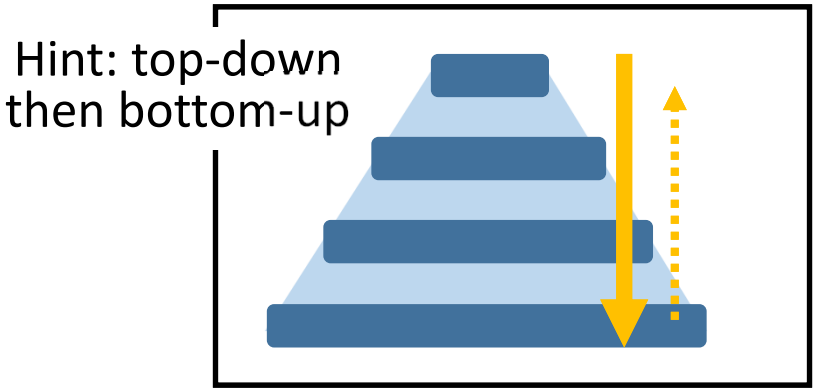


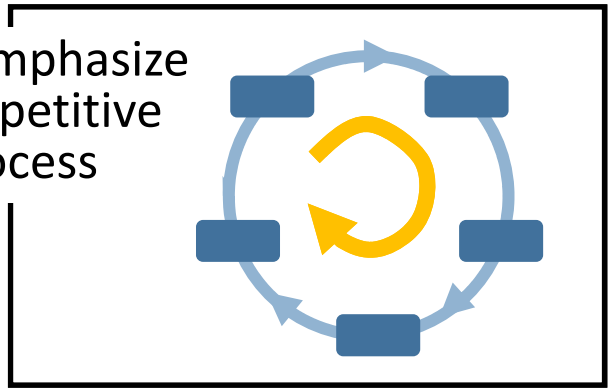
Hints Implied by Flow Direction

- Directions of flow of line of sights can provide useful implications to reduce the details of figures.

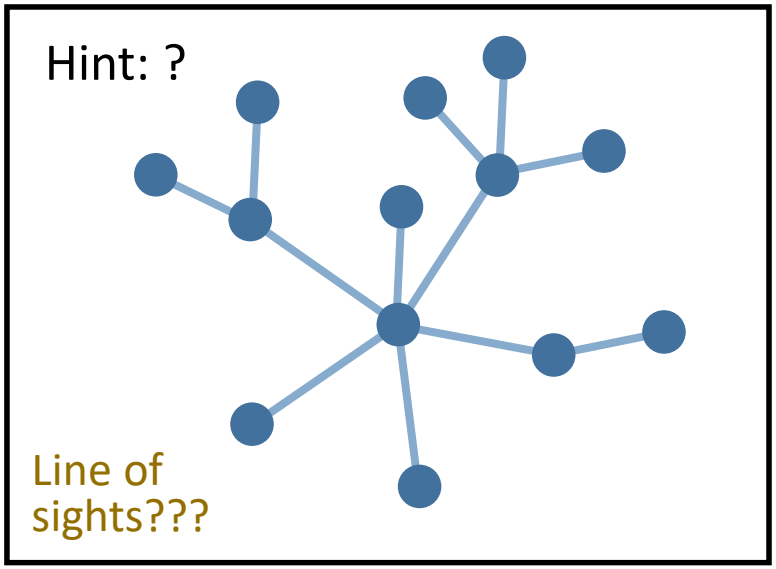






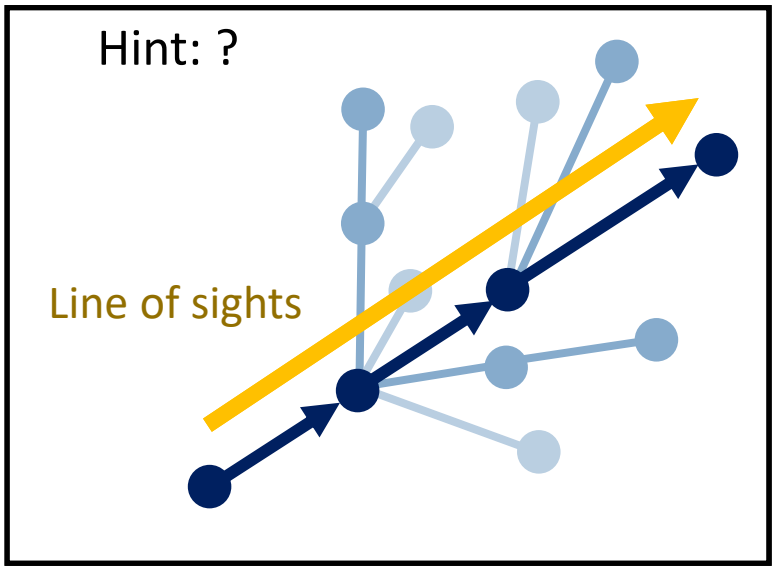


Good and Bad Examples



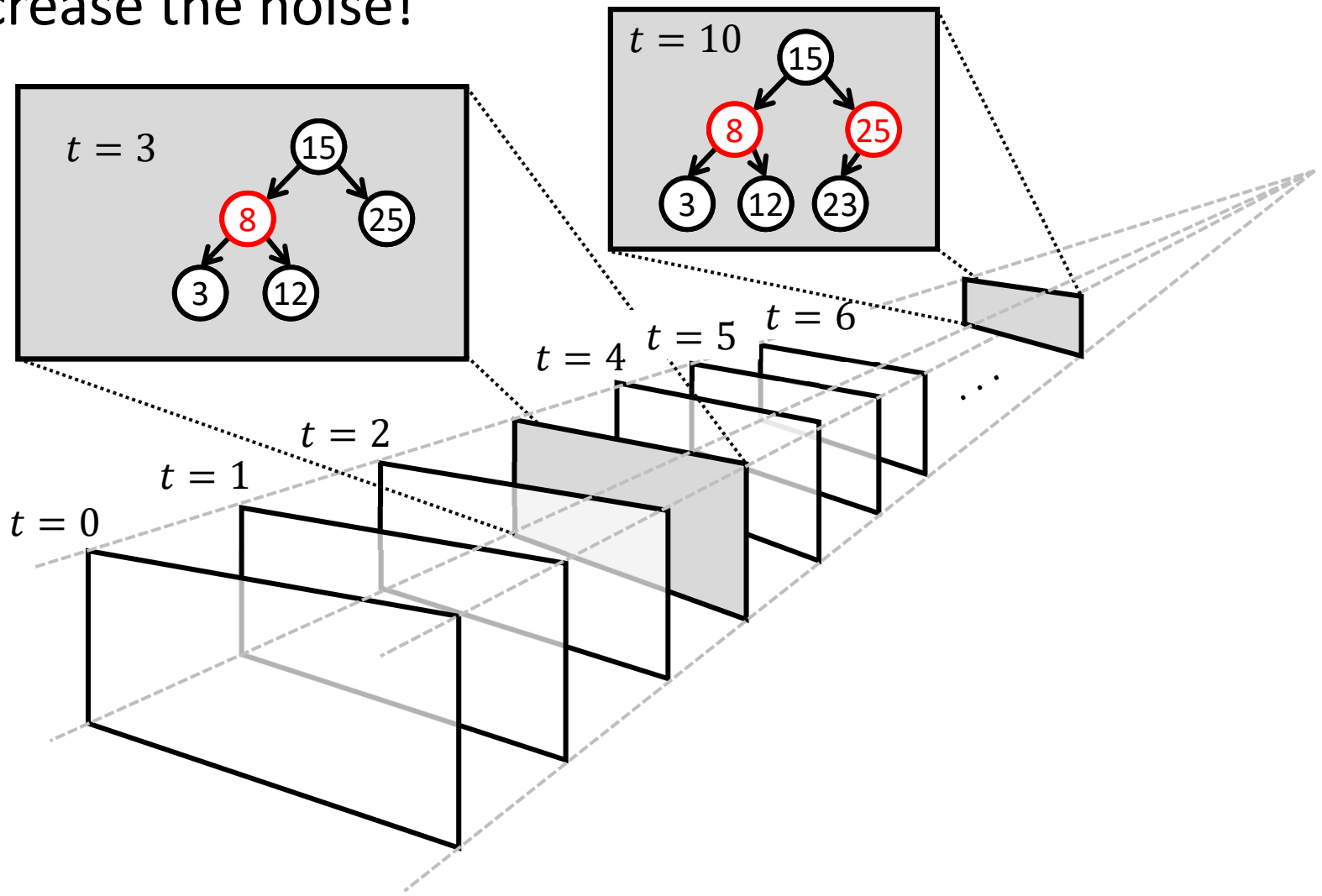
What metaphors did we use?

1. Brightness
2. Consistency
3. Direction
4. Arrow



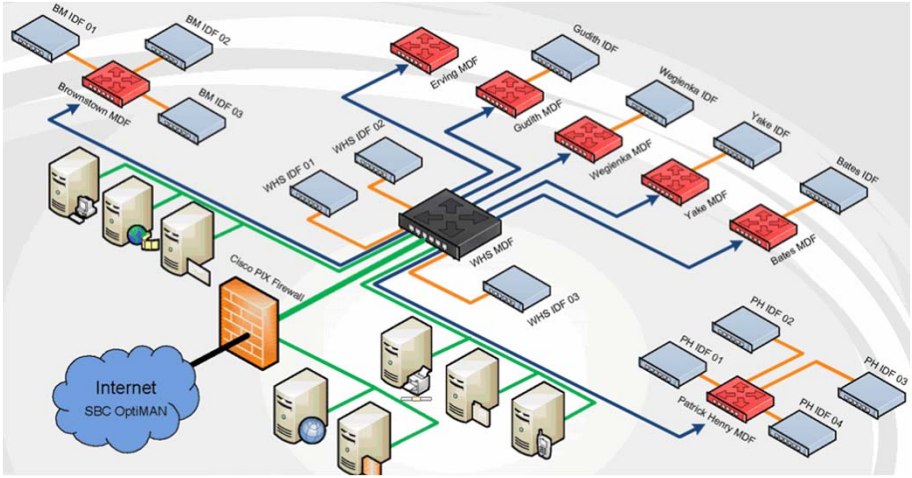
Advanced Hints Implied by Flow Direction

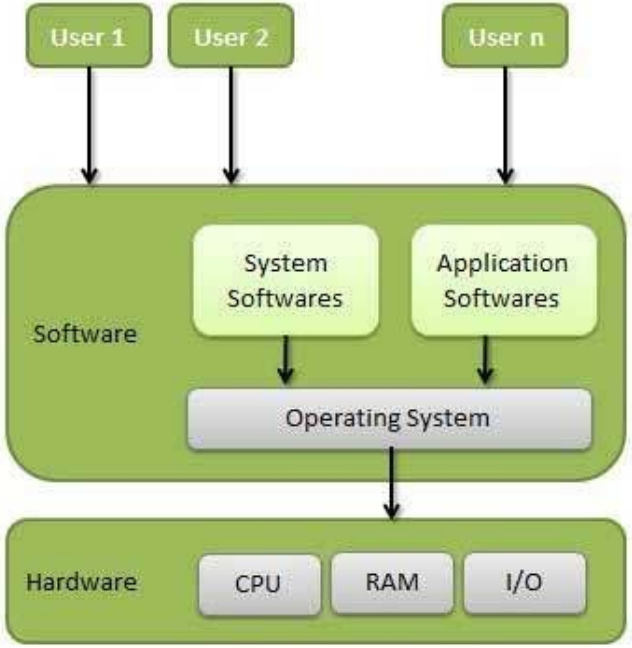
- Please be careful not to overuse such effects. They will increase the noise!



Hints Implied by Arrows

- Arrows often imply chronological order, (function) invocation order, or pointers in linked data structures.
- Arrows can reduce the usage of texts (such as numbering) in figures → enhanced SNR ratio.

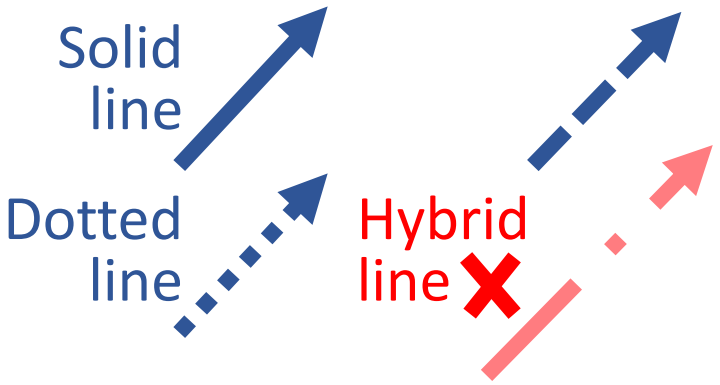


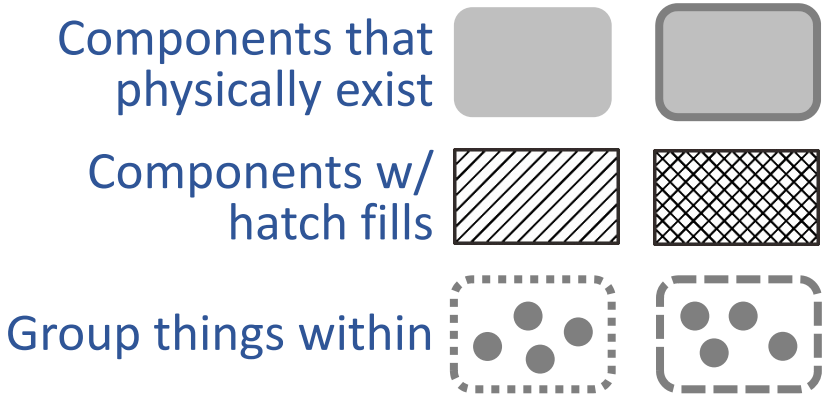


http://tr4.cbsistatic.com/hub/i/2010/09/03/41fc41c2-c3ae-11e2-bc00-02911874f8c8/7223a5dc109aca04746e9e547f73fdd0/hodgesbo1.gif, https://www.tutorialspoint.com/operating_system/images/conceptual_view.jpg

Hints Implied by Drawing Style

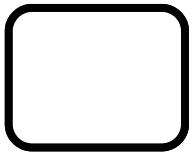
- **Line styles**
  - **Solid line**: things that physically exist; stronger relationship. Example: pointer; bus; communication channel.
  - **Dotted or dashed line**: logical or weaker relationship. Example: function invocation.
  - *We do not encourage the use hybrid lines (because it is complex and contains noises).*
- **Box styles**
  - **Filled box w/ or w/o solid border**: things that physically exist. Example: nodes in network; main memory; secondary storage.
  - **Transparent box w/ dashed border**: used to group things or create
  - We can use **hatch fills** to replace plain color fills; they are more vivid.



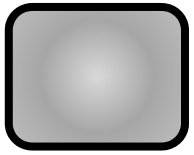


Styles Not Preferred by Academic Publications

- Most academic papers will be printed out by laser printers and read by readers.
- Effects not printable by laser printers should be avoided, for example:
  - Shadow effects:

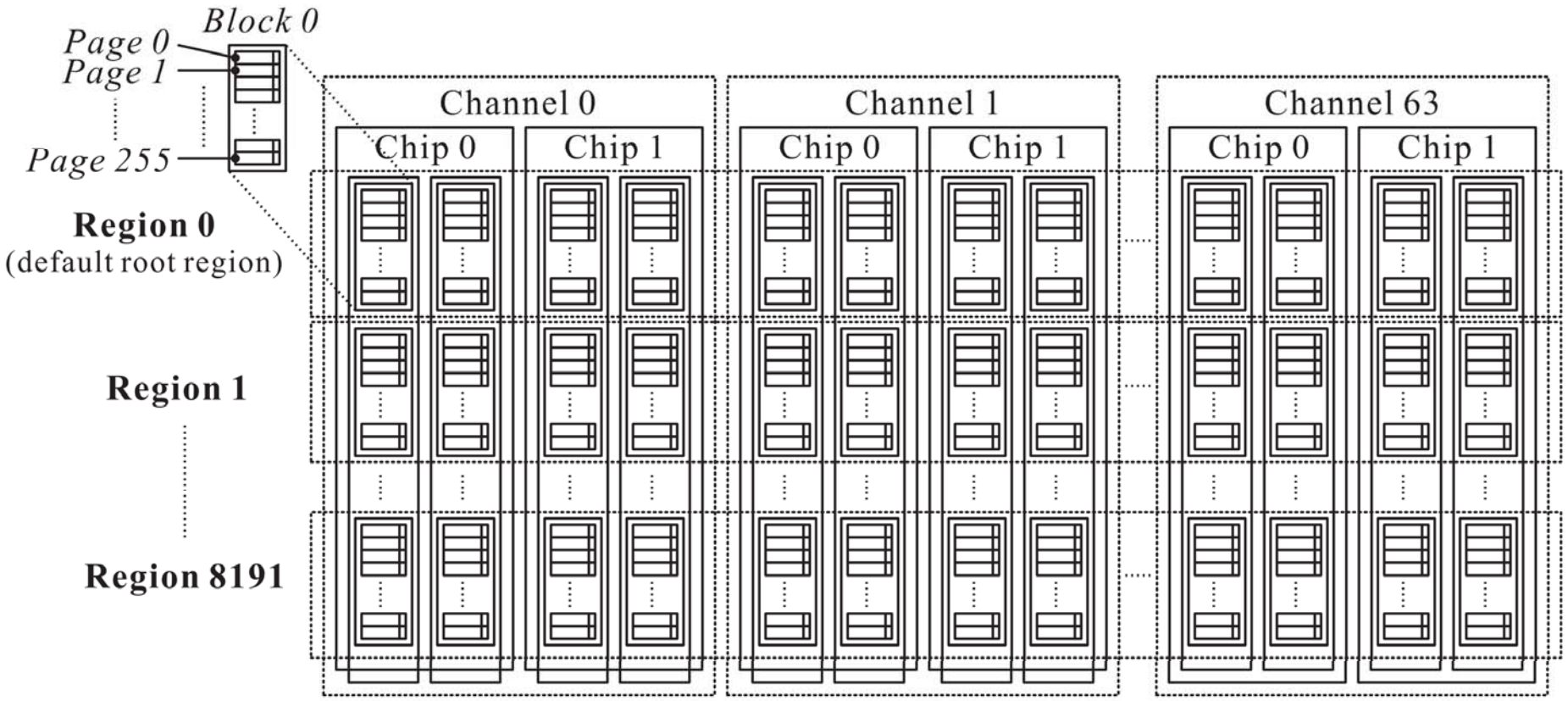


  - Gradient fills:



  - Very thin lines: See next page.

A Bad (Real-world) Example



Total device capacity: 4KBytes/Page * 256Pages/Block *
8192Blocks/Plane * 2Planes/Chip *
2Chips/Channel * 64 Channels/Device = **2TBytes**

Hints Imposed by Fonts

- Different fonts have different characteristics.
  - *Times New Roman: classical, formal, old-fashioned
  - Arial: modern, younger
  - *Chaparral Pro: typewriter-like, with good readability
  - Courier New: monospaced, professional, program code
  - ***Courier New (bold)**: easier to see than Courier New
  - Century Gothic: elegant and fashioned
  - *Calibri: default, soft, moderately casual
  - Comic Sans: casual, informal, relaxed, young (not suitable for formal academic publications; sometimes OK in slides)

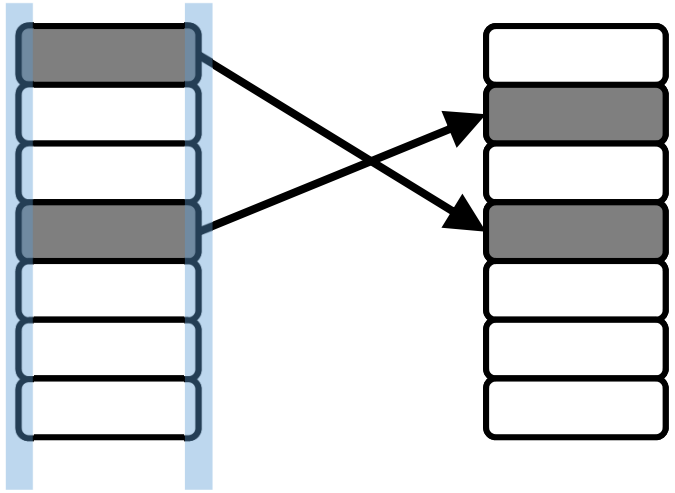
  (Fonts with * are my personal favorite!)
- **Use no more than 2 fonts in the same figure to reduce the total quantity of information shown by the figure.**
- **In formal academic publications, there is no bold face in the main texts; emphasis is done by italicizing the texts. The case could be different in figures (if we're careful).**

Advices of Font Usage

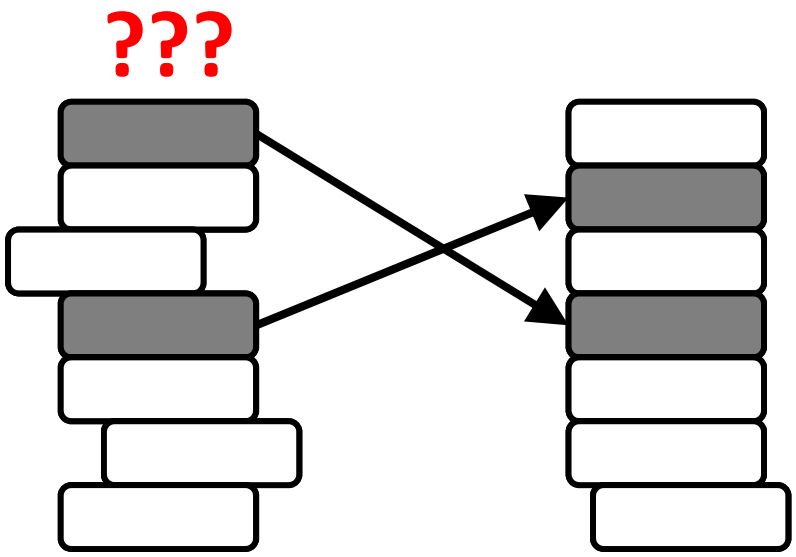
- **Use the same fonts as in the manuscript.** In case to highlight the figures, fonts with heavier weights can also be used.
  - Example 1: Times New Roman as manuscript font, and Times New Roman as figure font.
  - Example 2: Times New Roman as manuscript font, and **Calibri (bold)** as figure font.
- Do not use casual fonts like Comic Sans in formal papers, neither in main texts or in figures.
- Use only 1 or 2 font–size combinations in the same figure.
- The size of the fonts of in-text figures should be between 70%–100% of the main texts.
- The texts in figures should be kept short (except for the cases with good reasons).

The Power of Alignment

- Elements aligned to the same (invisible) straight line imply that they are related or belong to the same group.
- This is because that **human eyes will "complete" things automatically**.
- Example:



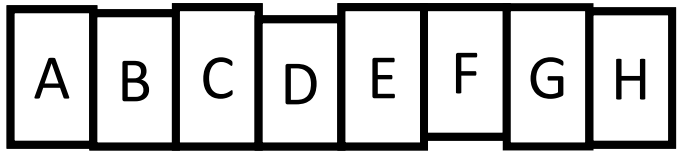
這是一組的！



“Auto-completion” of Human Eyes

The Power of Consistency (Sameness)

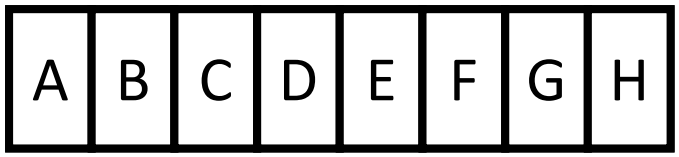
- Elements of the same style (and even size) imply that they are the same thing or belong to the same class.
  - Elements with the same style but different sizes can imply a closer or farer view of the same instance of the element, if they are connected through dotted lines or gradient fills.

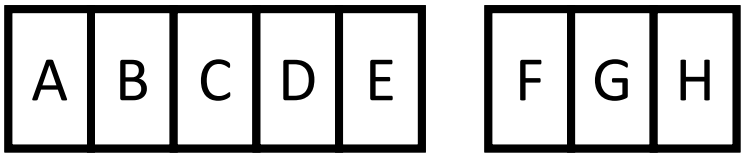


http://img05.tooopen.com/images/20150527/tooopen_sy_126682712132.jpg

The Power of Positioning

- **Nearby elements imply that they are related to each other. In addition, the closer they are in the same figure, the relationship is implied stronger.**
- **The same spacing between elements implies consistency.** In contrast, moving some elements to violate this metaphor provides extra information (no matter intentionally or unintentionally).





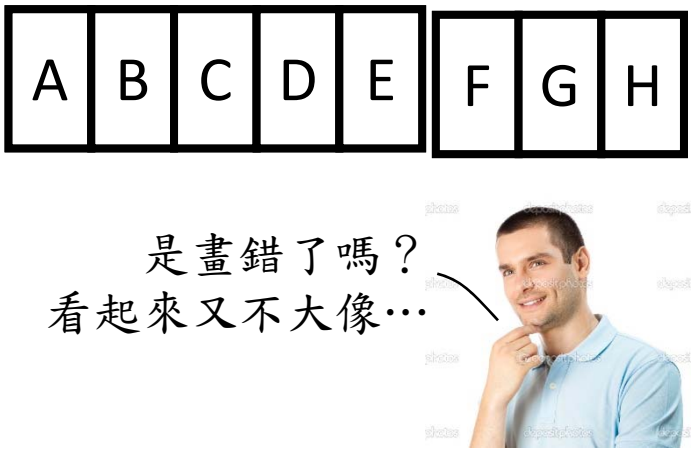
http://health.dahe.cn/jbzs/zbz/news/201309/W020130922328211636078.jpg

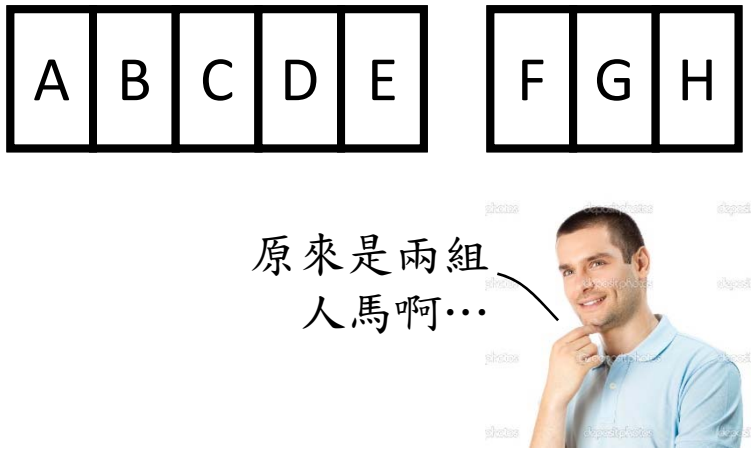
http://static6.depositphotos.com/1144472/632/i/950/depositphotos_6324296-stock-photo-portait-of-thinking-man-looking.jpg

Do It Intentionally

- If we want to do something intentionally, we should try to let it look so.

  (如果我們要有意做什麼樣的設計，就要讓他看起來是故意的，而不要模稜兩可！)





Quick Summary of the Metaphors

- **Arrows often imply chronological order, (function) invocation order, or pointers in linked data structures.**
- **Elements aligned to the same (invisible) straight line imply that they are related.**
- **Elements of the same style (and even size) imply that they are the same thing or belong to the same class.**
  - Elements with the same style but different sizes can imply a closer or farer view of the same instance of the element, if they are connected through dotted lines or gradient fills.
- **Nearby elements imply that they are related to each other.**
  - The closer they are in the same figure, the relationship is implied stronger.

Conversion Among File Formats

- PDF is the *de facto* standard format for academic publications. Other popular formats: HTML and DOC/DOCX.
- As we use LaTeX or Word to write papers, the supported figure formats are different:
  - Word supports (raster) GIF, PNG, JPEG, TIFF, BMP; (vector) WMF, EMF, EPS.
  - LaTeX (with latex–dvips–ps2pdf tool chain) supports only EPS.
  - LaTeX (with pdflatex tool chain) supports JPG, PNG, PDF, and EPS (included epstopdf package).
- **To minimize "missing font" problems, we suggest that fonts are converted to curve before inserted to the manuscript. This is a common problem reported by IEEE PDF Checker.**

Case Study 1

Technique: Highlight the changed part!

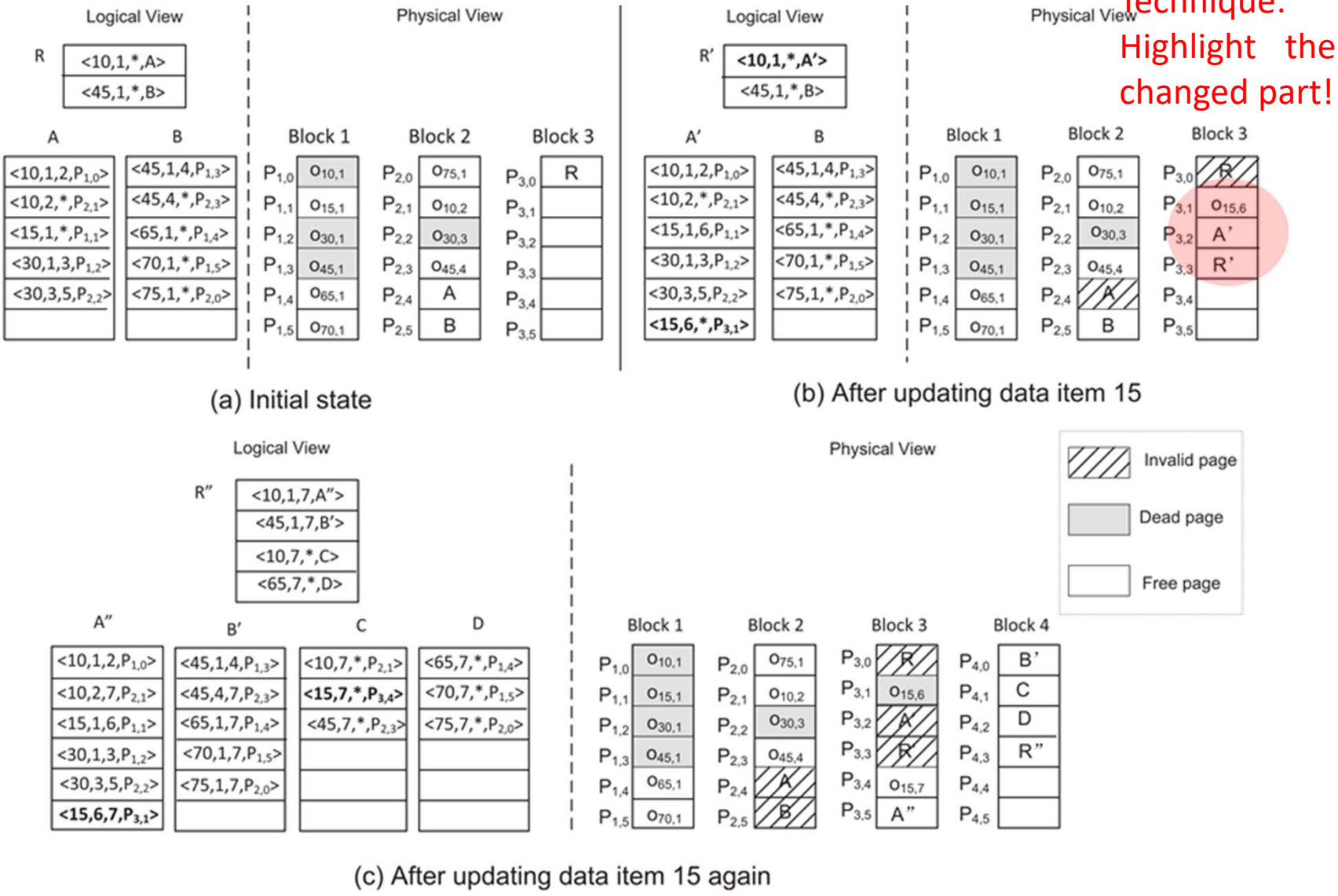
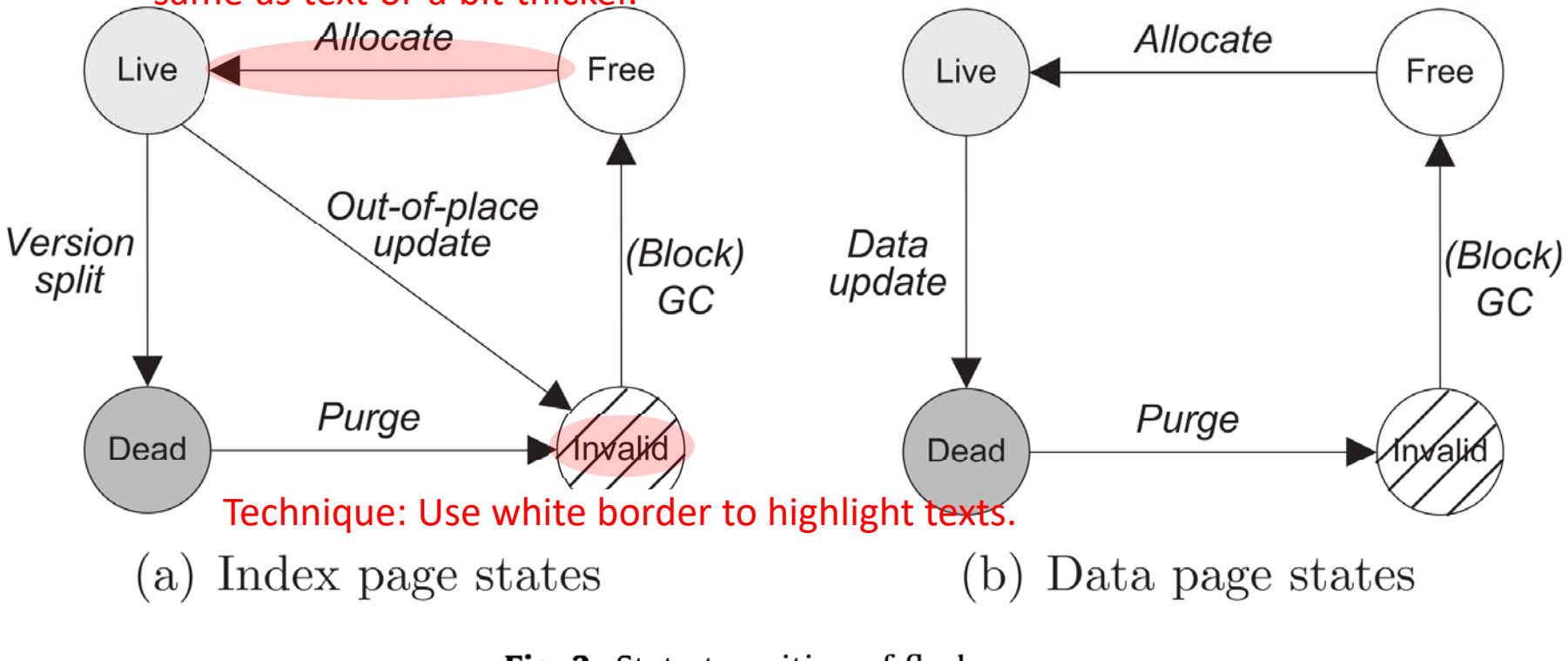


Fig. 2. The motivational example.

Case Study 2

Technique: Beware of the line width to be around the same as text or a bit thicker.

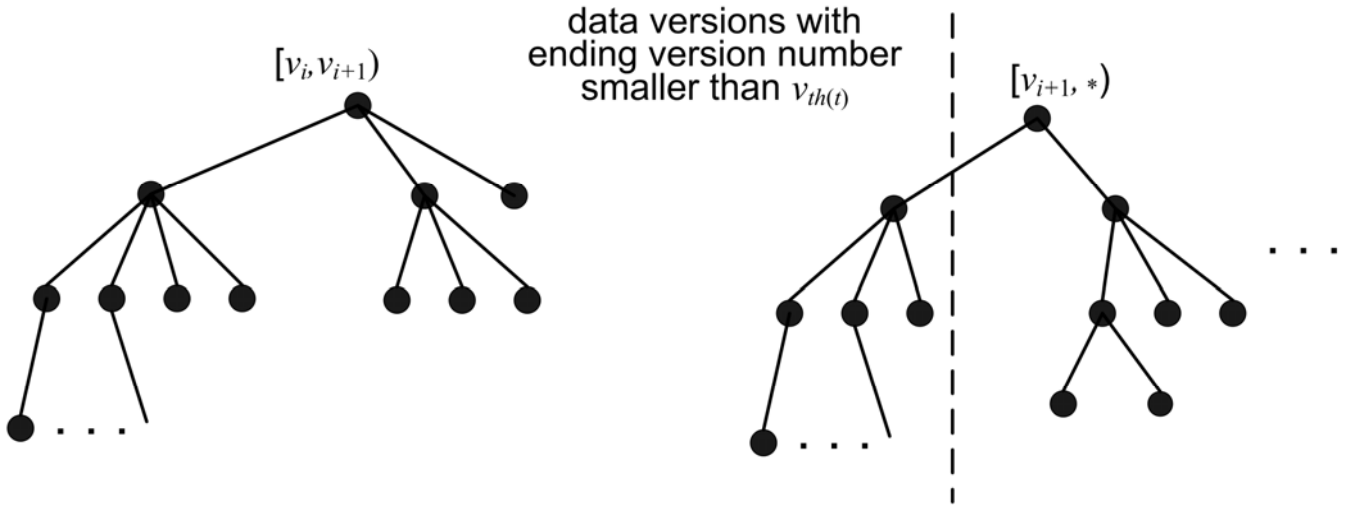


Technique: Use white border to highlight texts.

(a) Index page states (b) Data page states

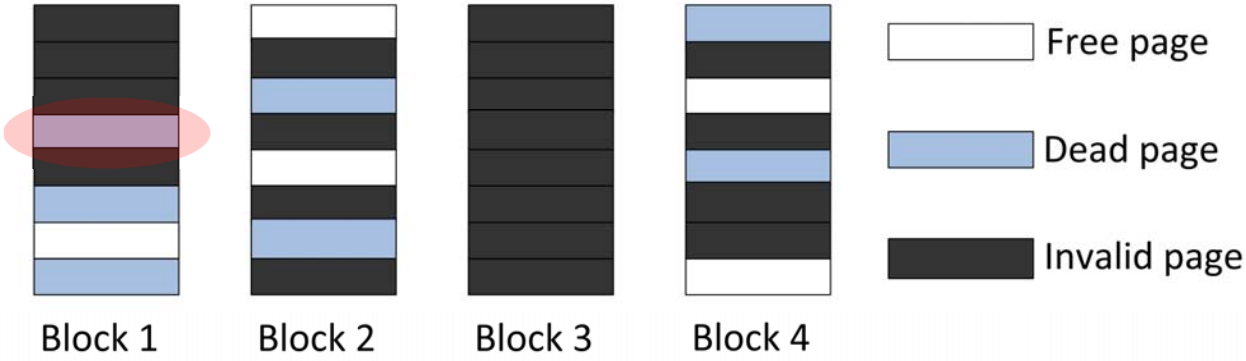
Fig. 3. State transition of flash pages.

Case Study 3



(a) An illustration of purging an index with purging version number $v_{th}(t)$

Technique: Do not use color.



(b) Physical layout after executing $Q(-\infty, v_h, *)$

Fig. 4. An illustration of the purging-range query $Q(v_l, v_h, *)$.

Case Study 4

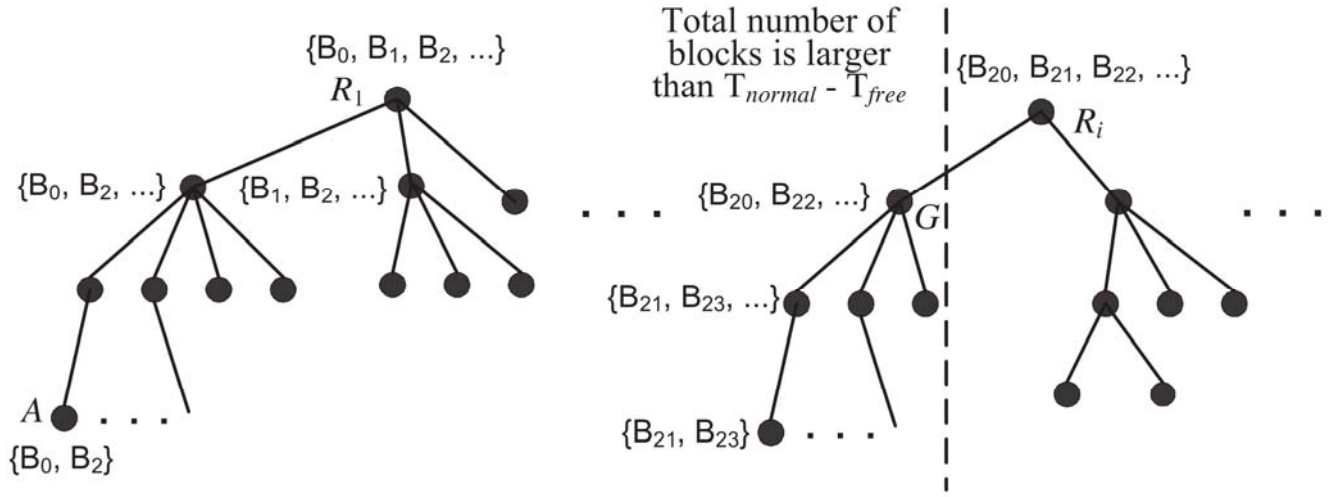


Fig. 5. An illustration of *PBL* of the index tree.

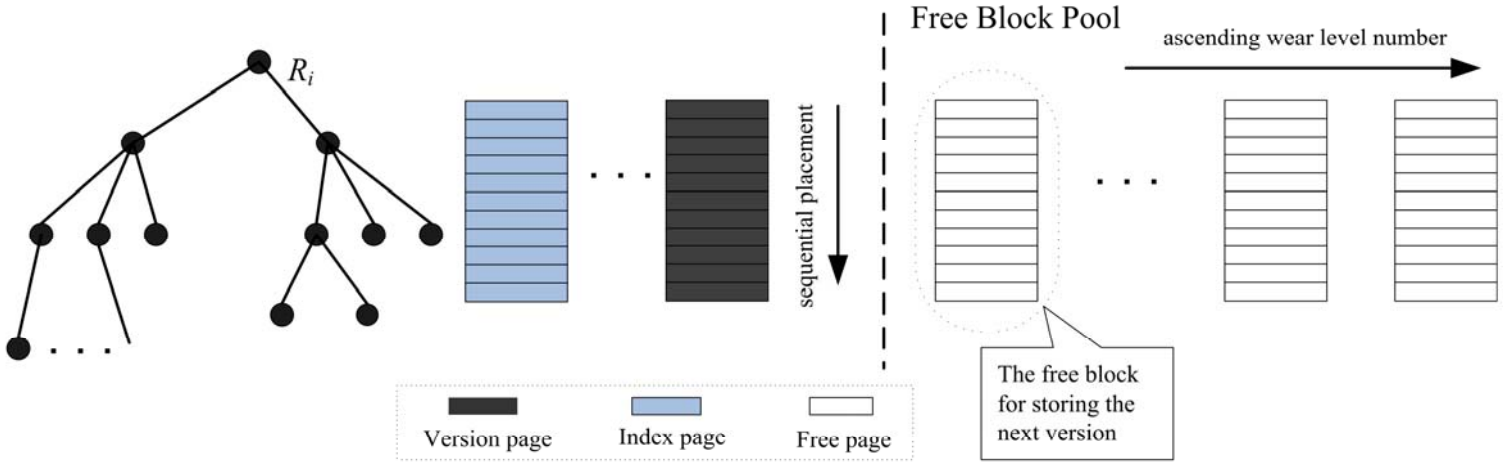
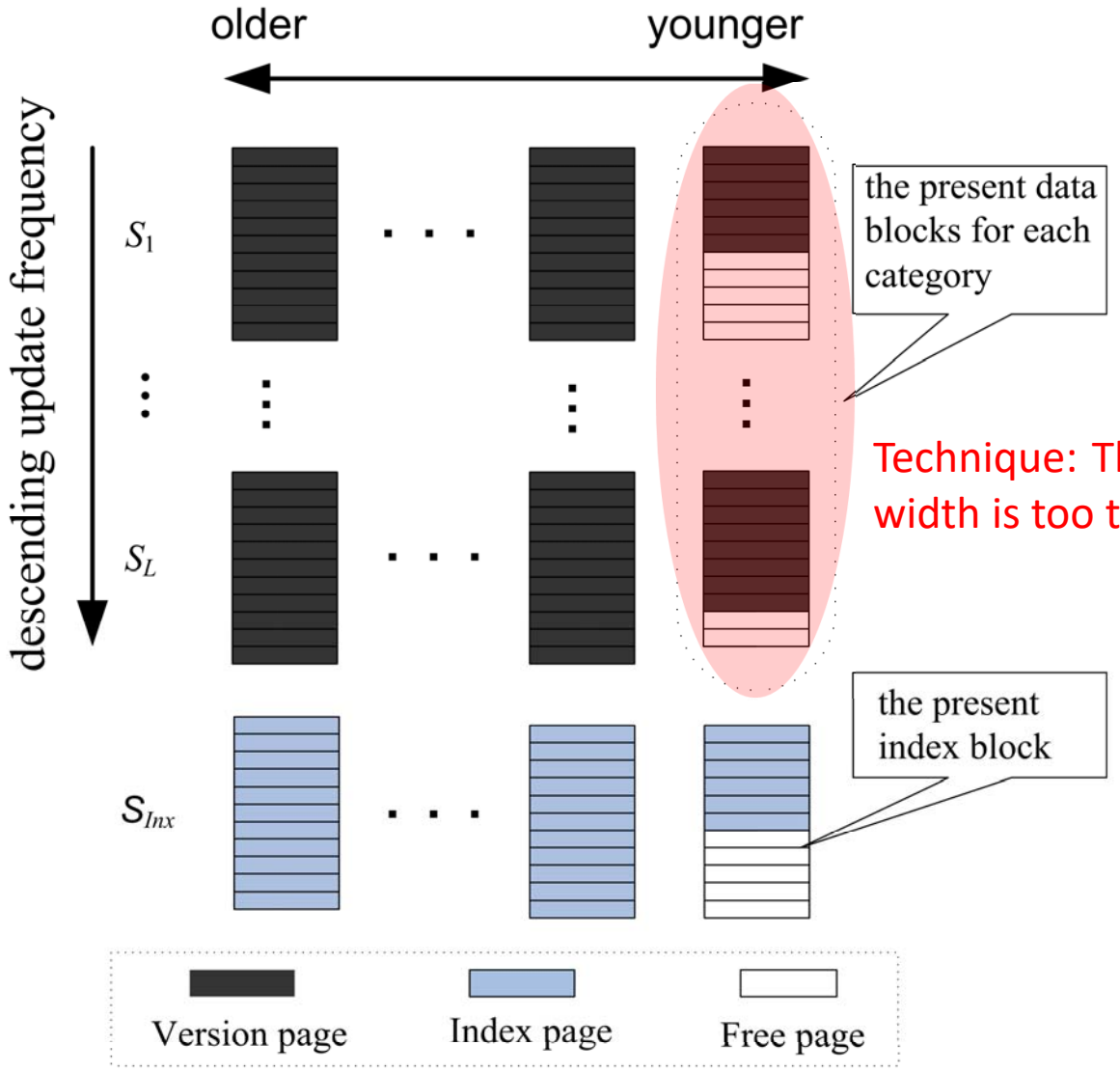


Fig. 6. An illustration of *SQ*.

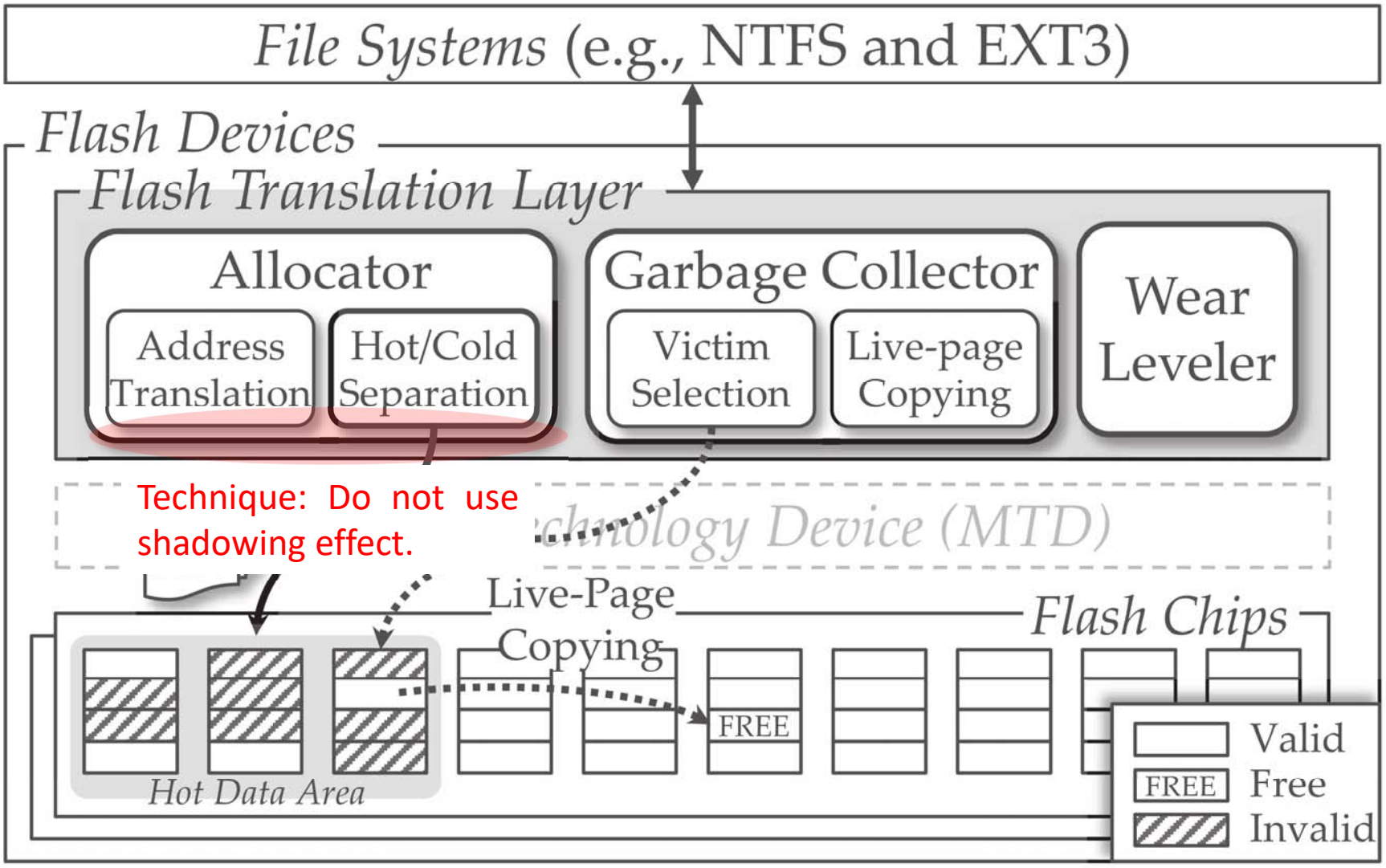
Case Study 5



Technique: The line width is too thin.

Fig. 8. An illustration of *FBP*.

Case Study 6


File Systems (e.g., NTFS and EXT3)
Flash Devices
Flash Translation Layer
Allocator
Address Translation
Hot/Cold Separation
Garbage Collector
Victim Selection
Live-page Copying
Wear Leveler
Technique: Do not use shadowing effect.
Live-Page Copying
Flash Chips
FREE
Hot Data Area
Valid
FREE
Free
Invalid

Case Study 7

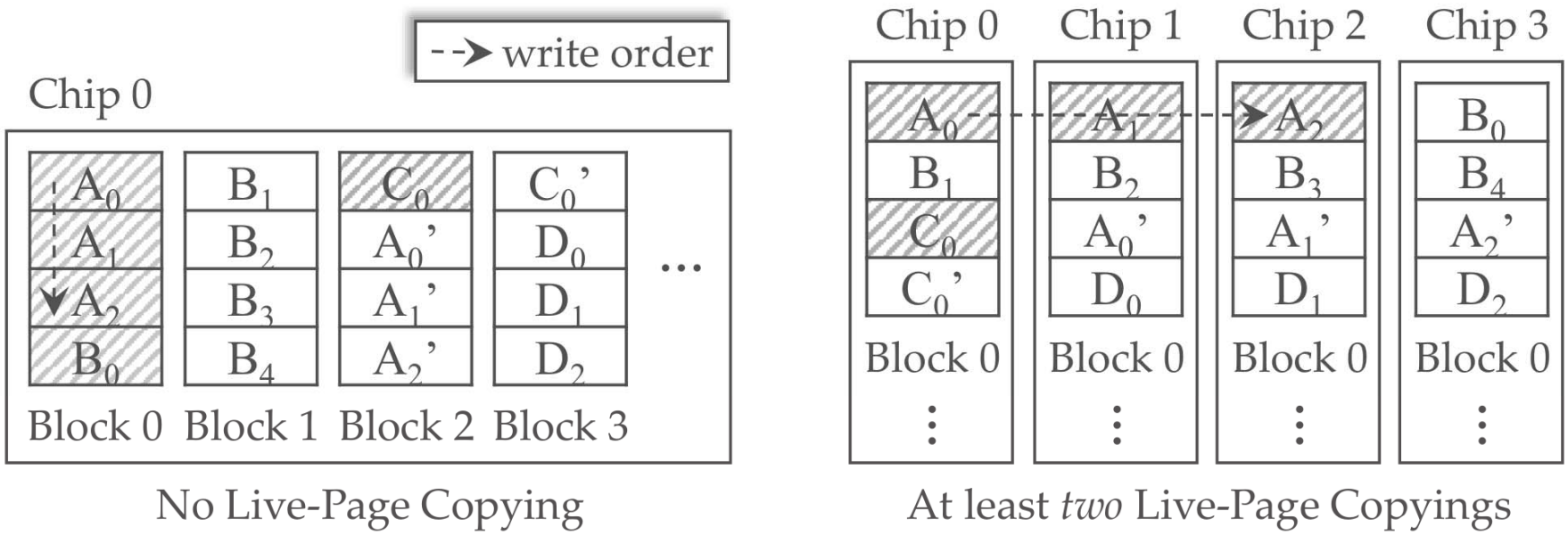
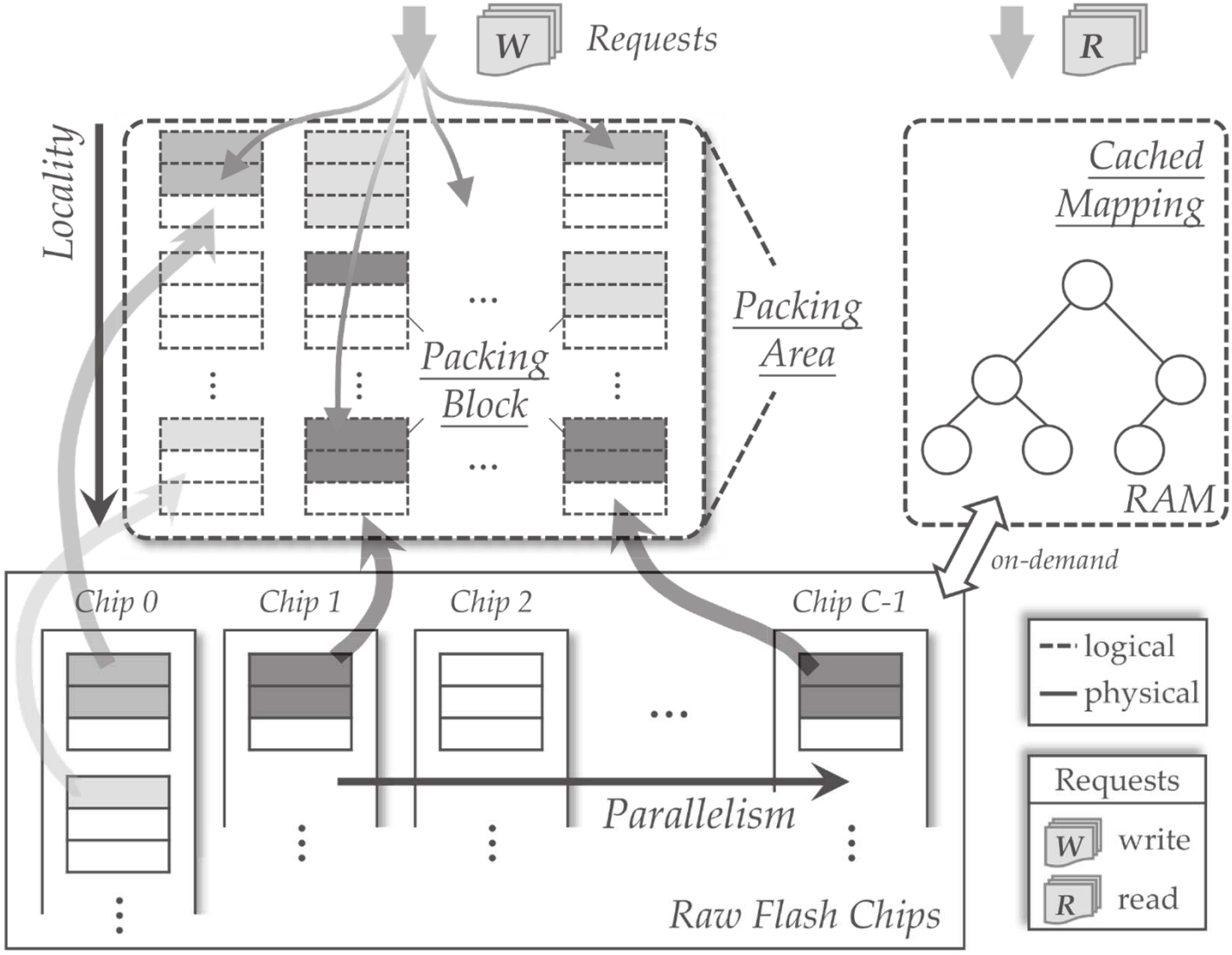


Fig. 2. An example of the potential conflict between the two development trends (*i.e.*, hot/cold separation and multi-chip architecture).

Case Study 8


W
Requests
R
Locality
Cached
Mapping
Packing
Block
Packing
Area
RAM
on-demand
Chip 0
Chip 1
Chip 2
Chip C-1
Parallelism
Raw Flash Chips
logical
physical
Requests
W
write
R
read

Case Study 9

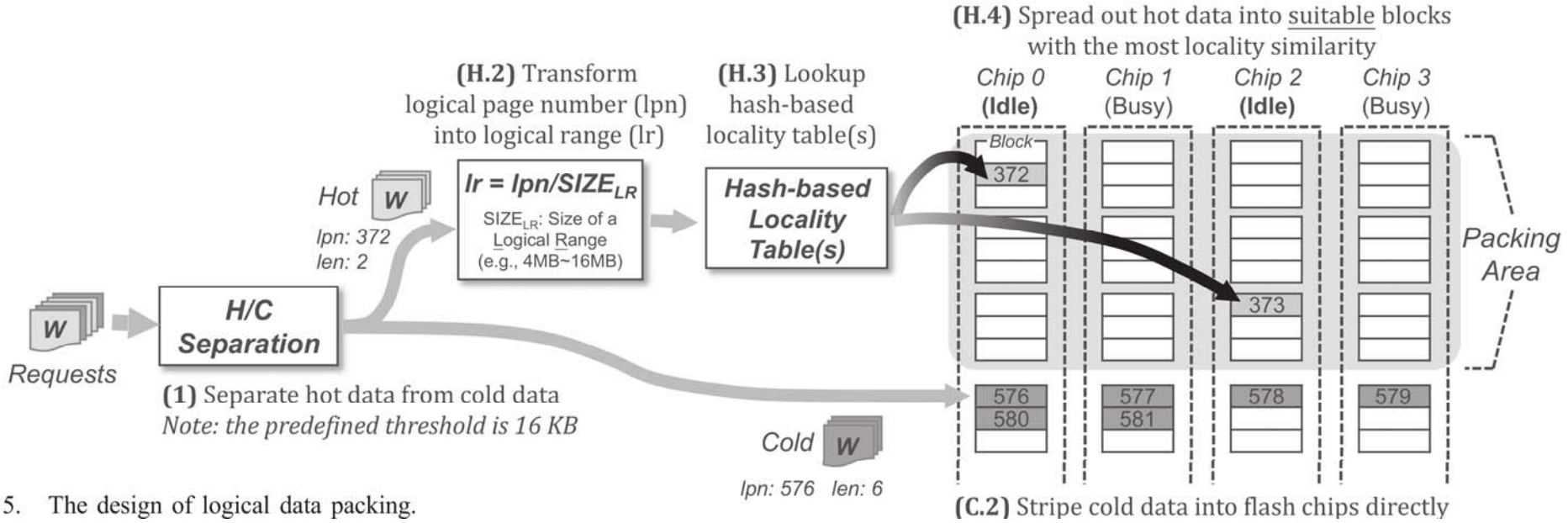
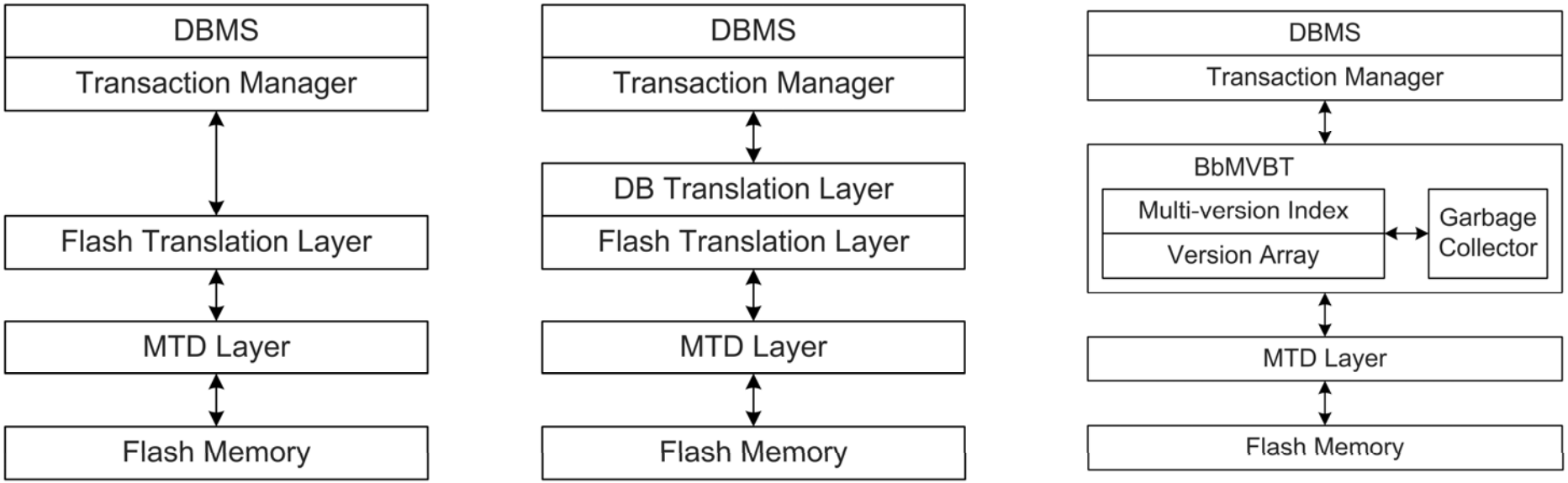


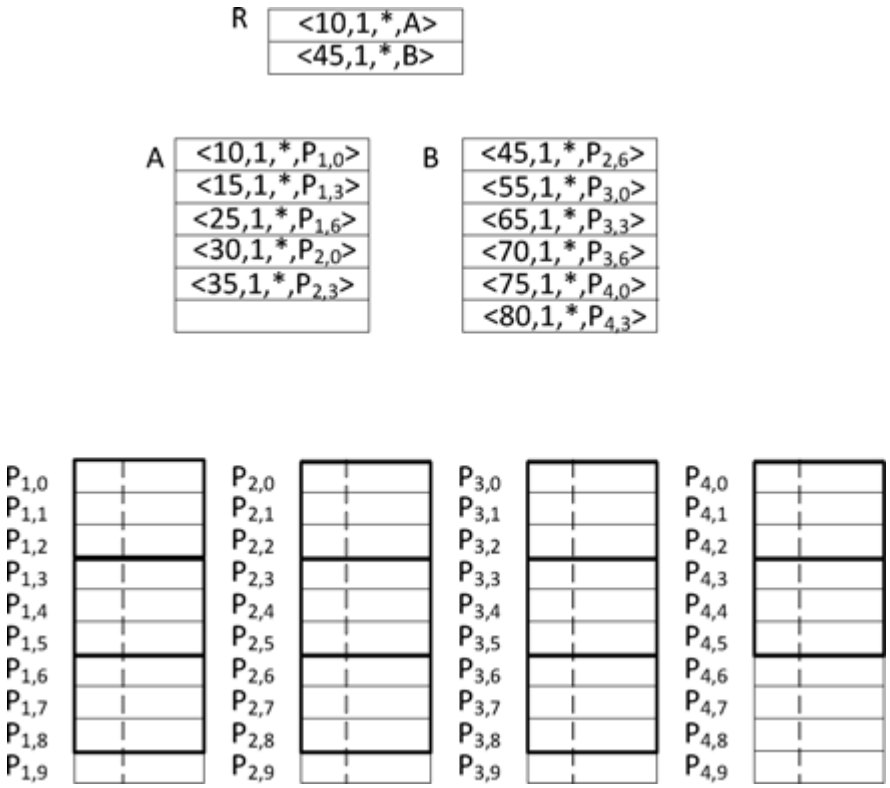
Fig. 5. The design of logical data packing.

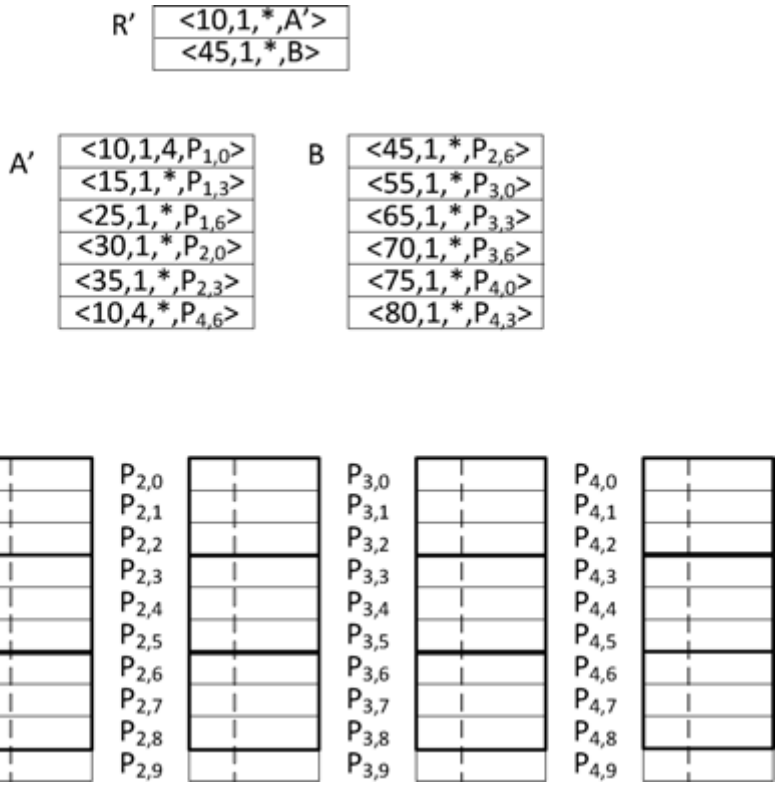
Case Study 10



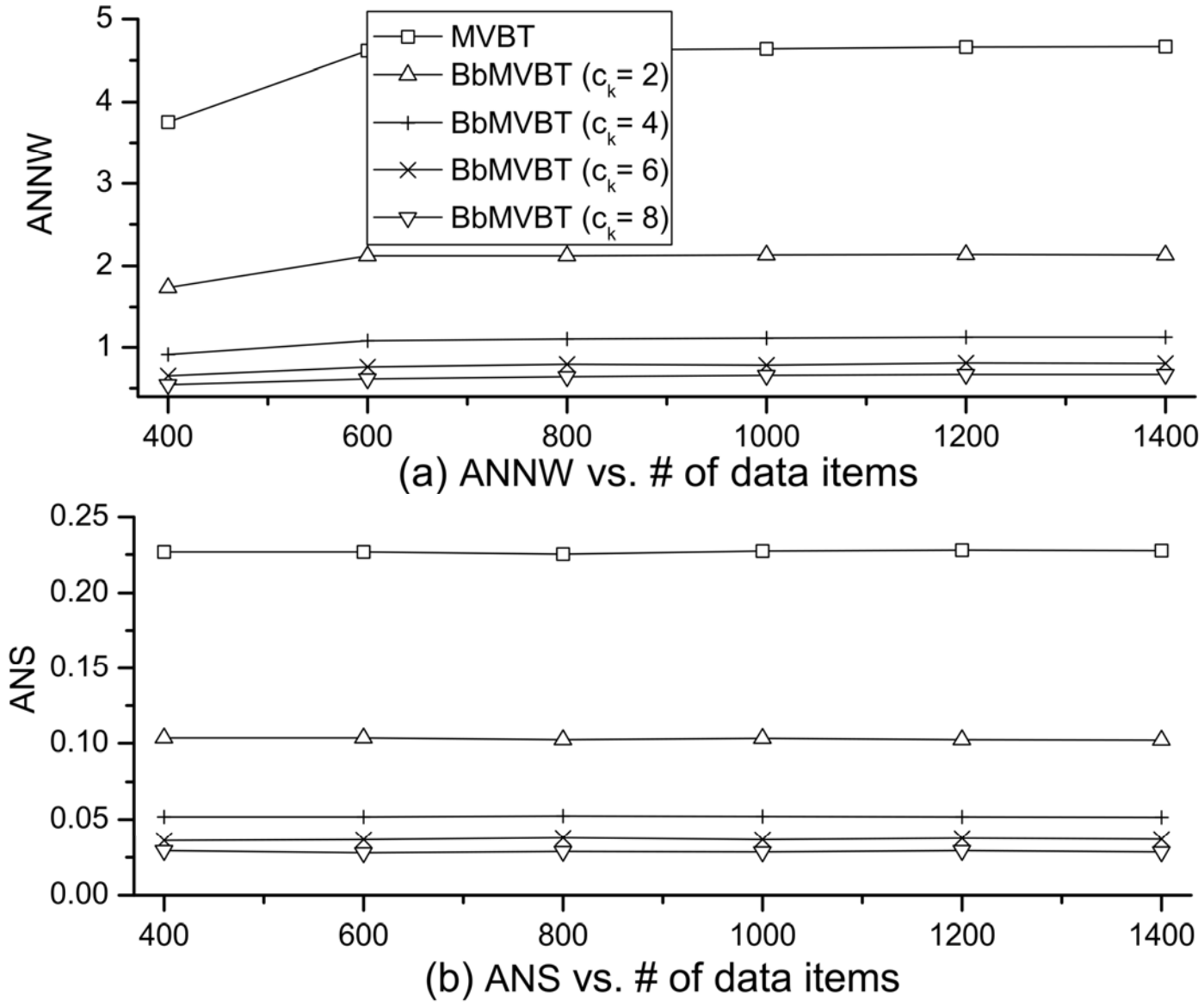
(a) Direct adoption (b) Layered design (c) Flash-aware design (our design)

Case Study 11

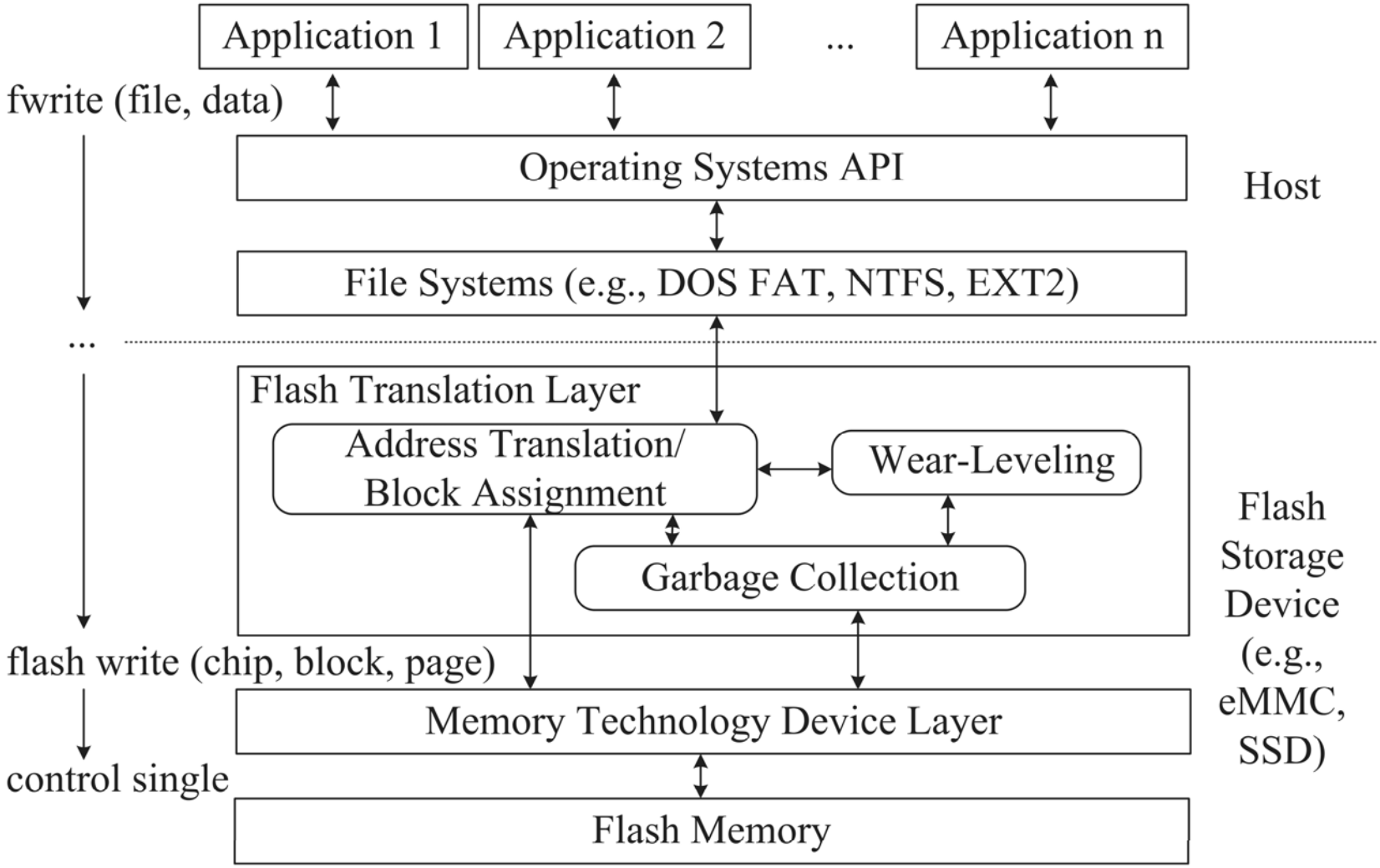




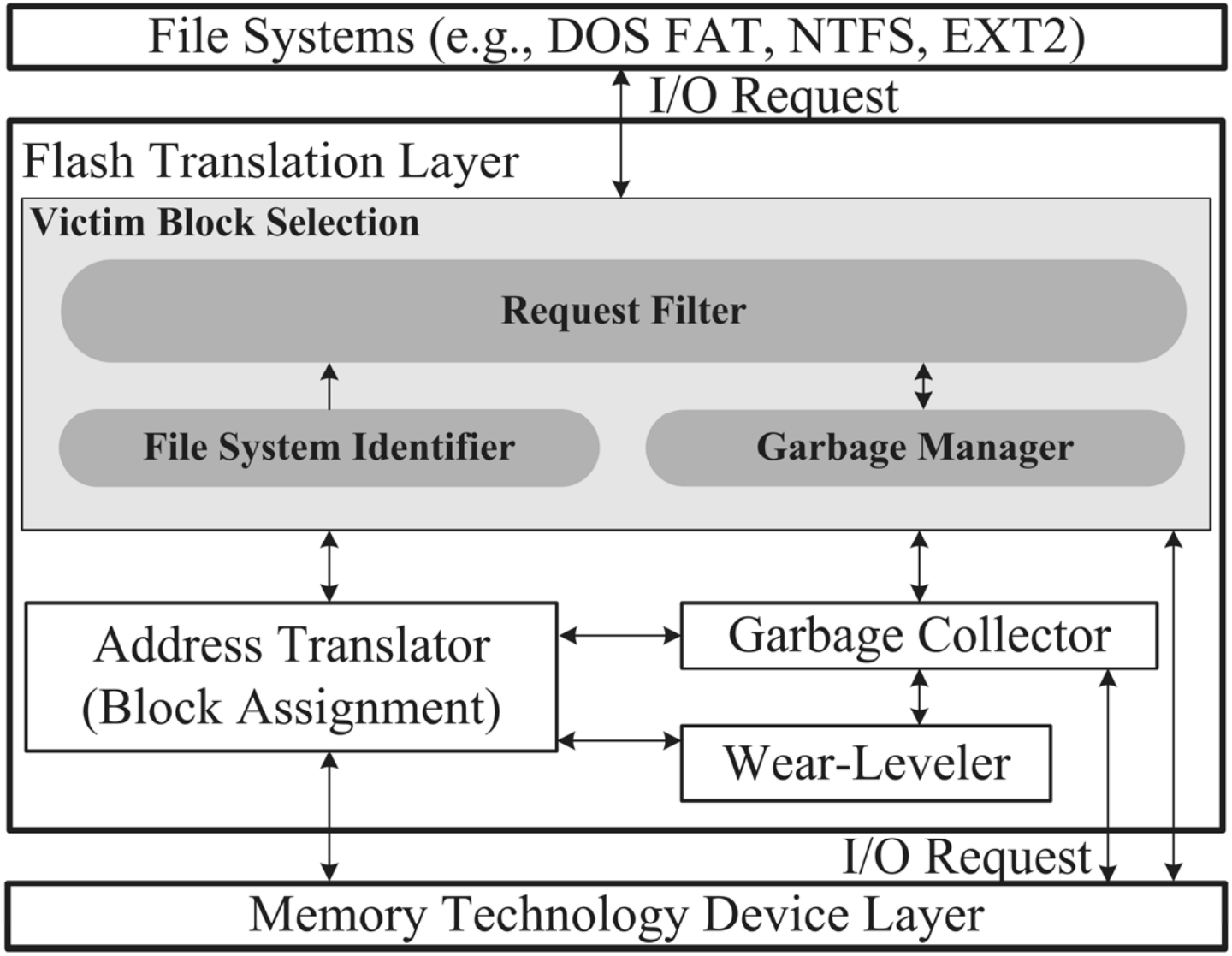
Case Study 12



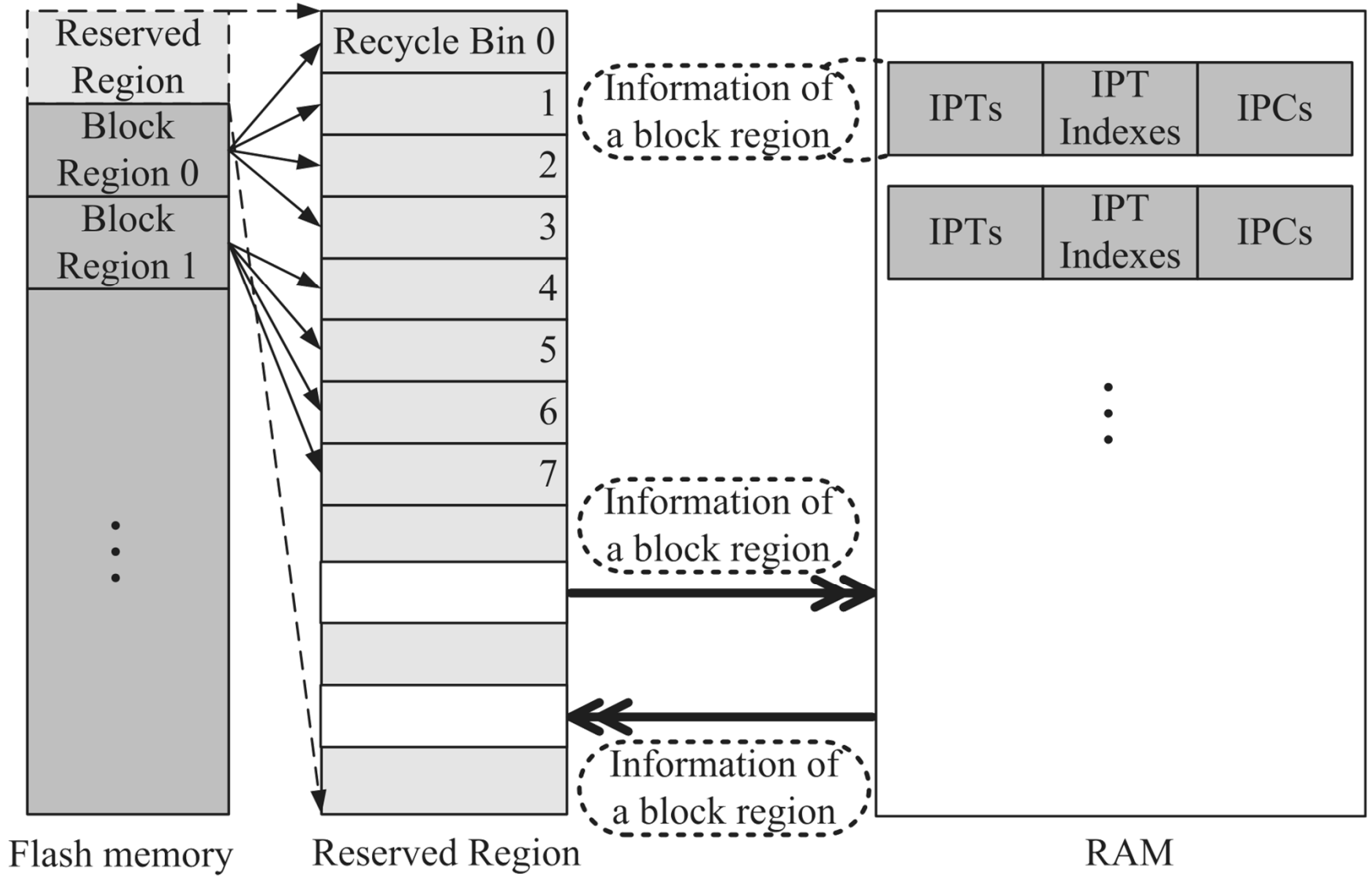
Case Study 13



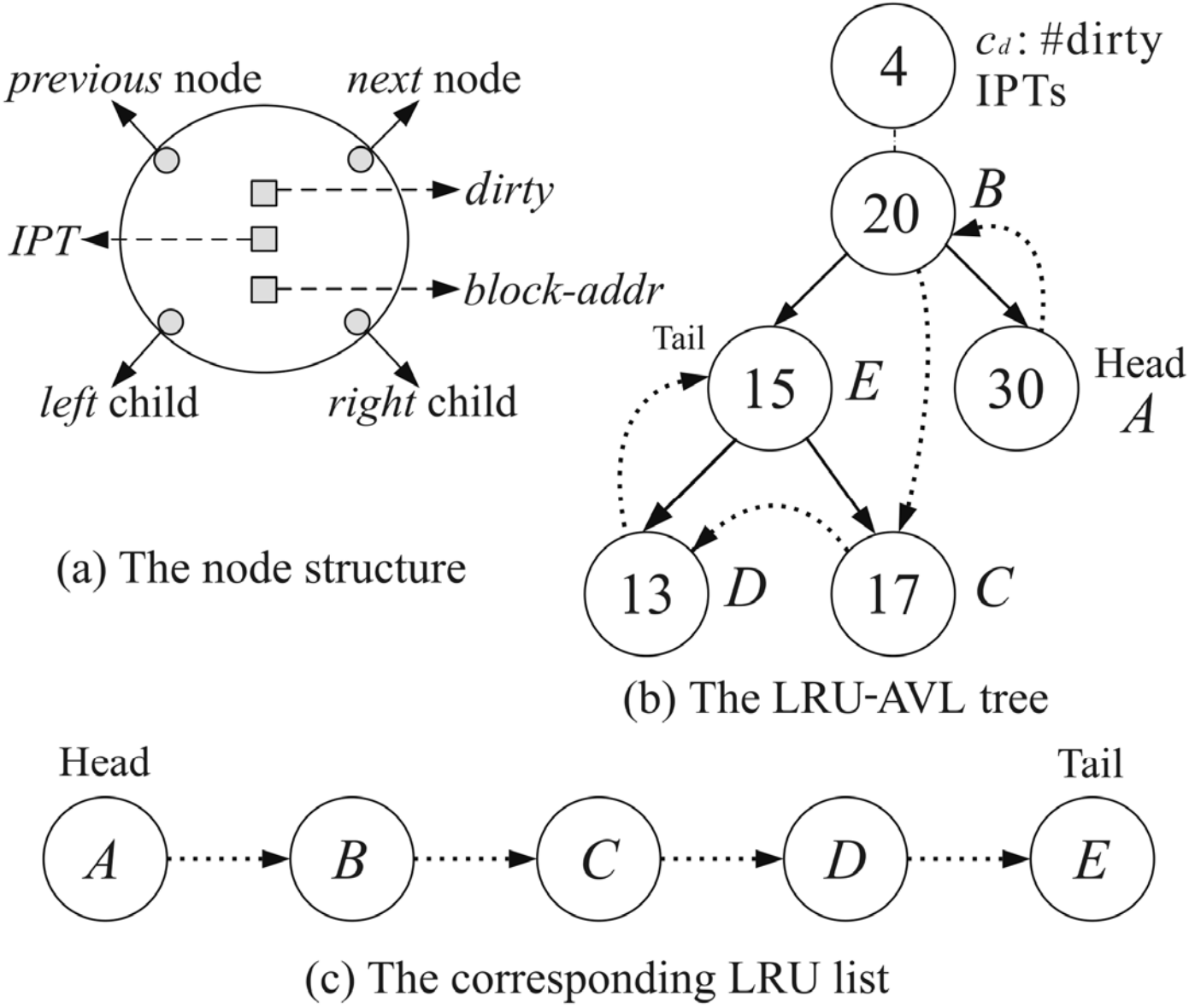
Case Study 14



Case Study 15



Case Study 16

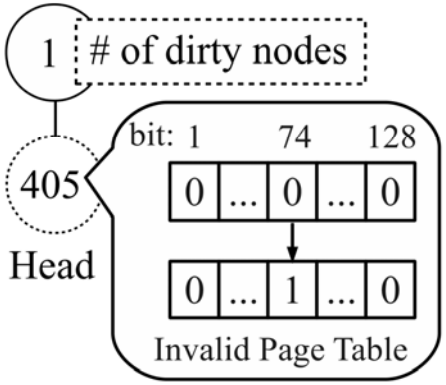


(a) The node structure

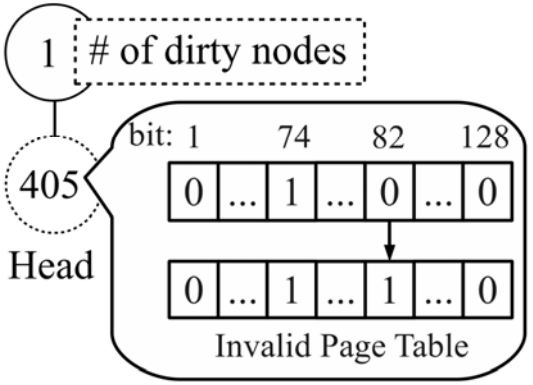
(b) The LRU-AVL tree

(c) The corresponding LRU list

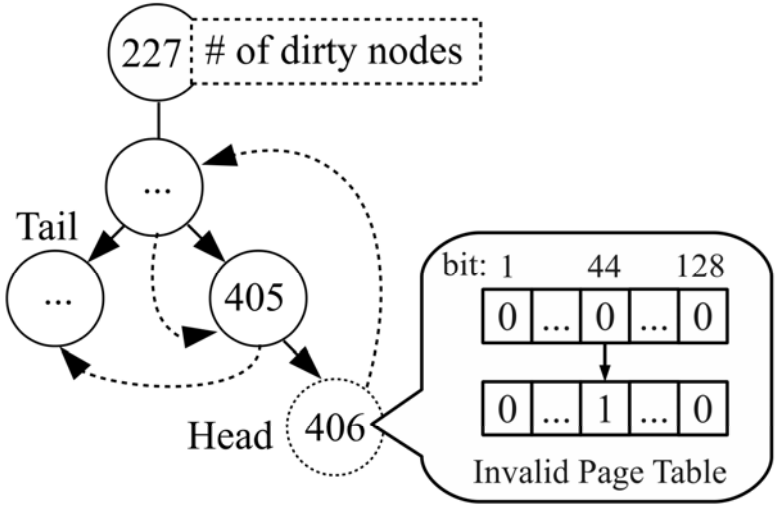
Case Study 17



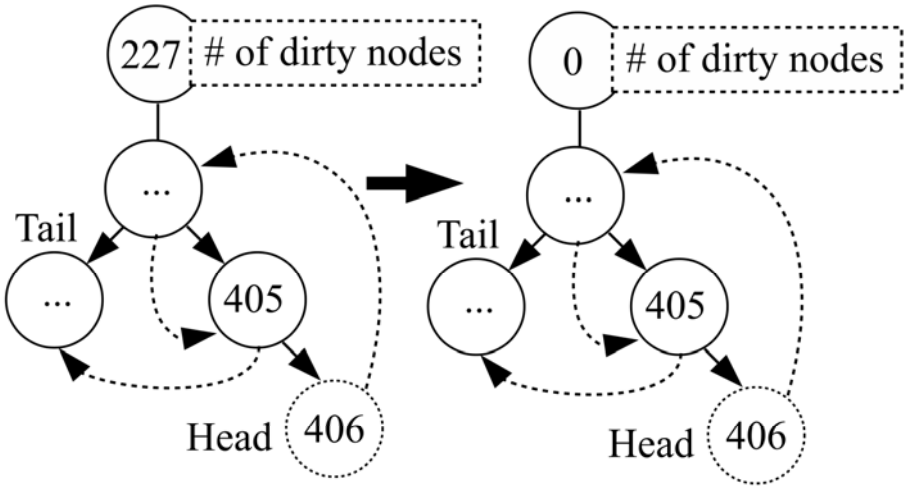
(a) Invalided page 576201



(b) Invalided page 576209

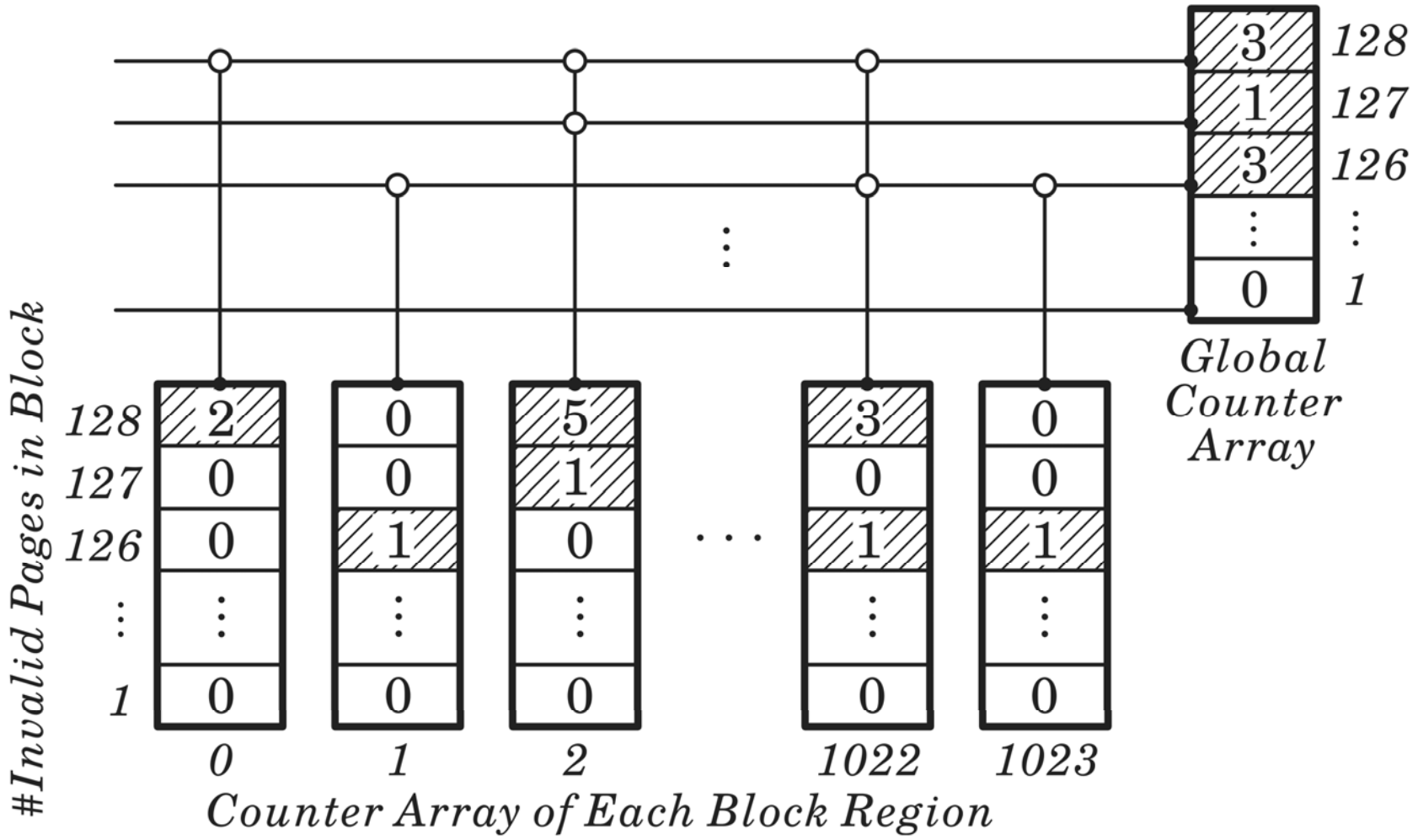


(c) Invalided page 576300

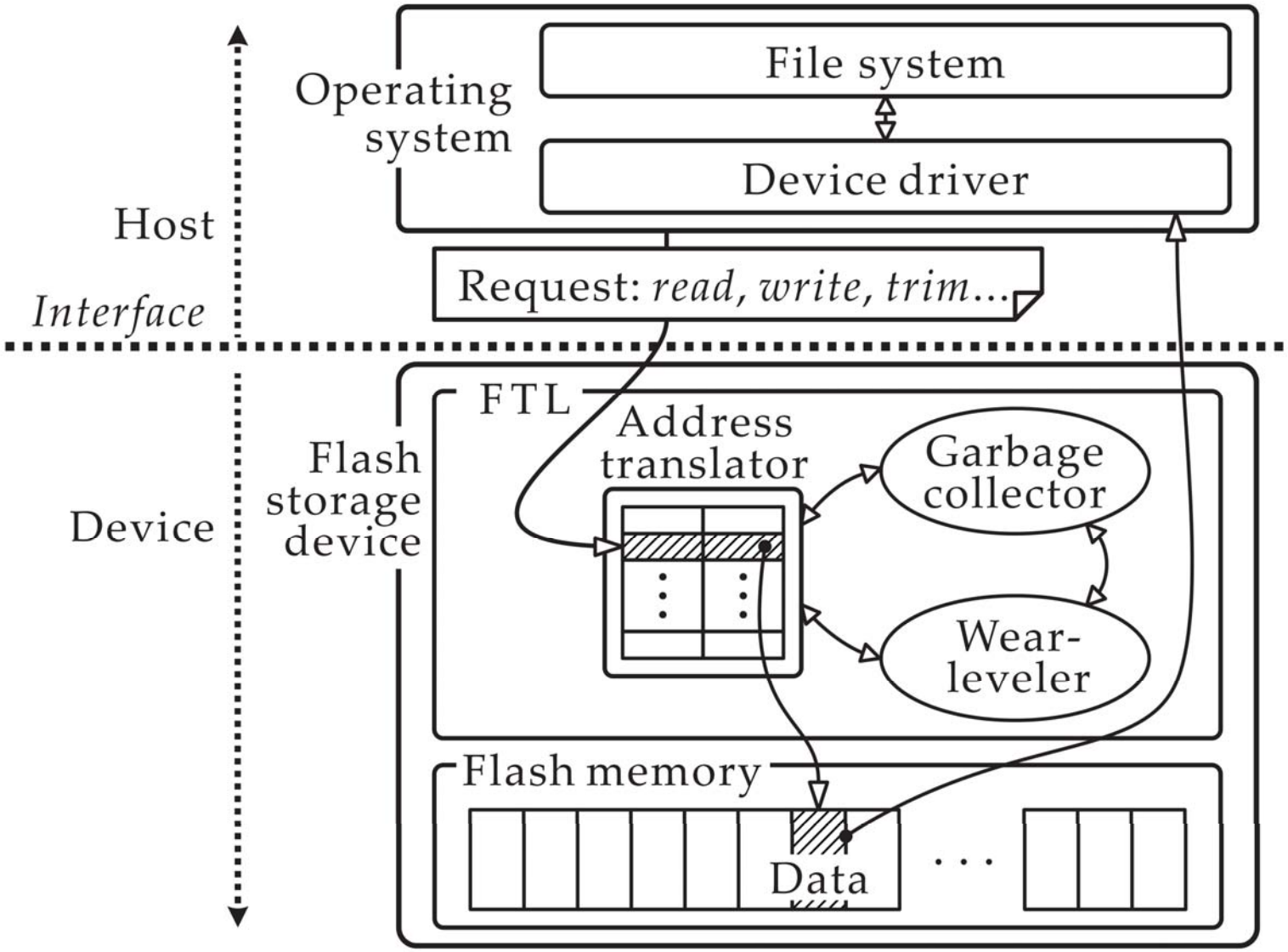


(d) Committing all node

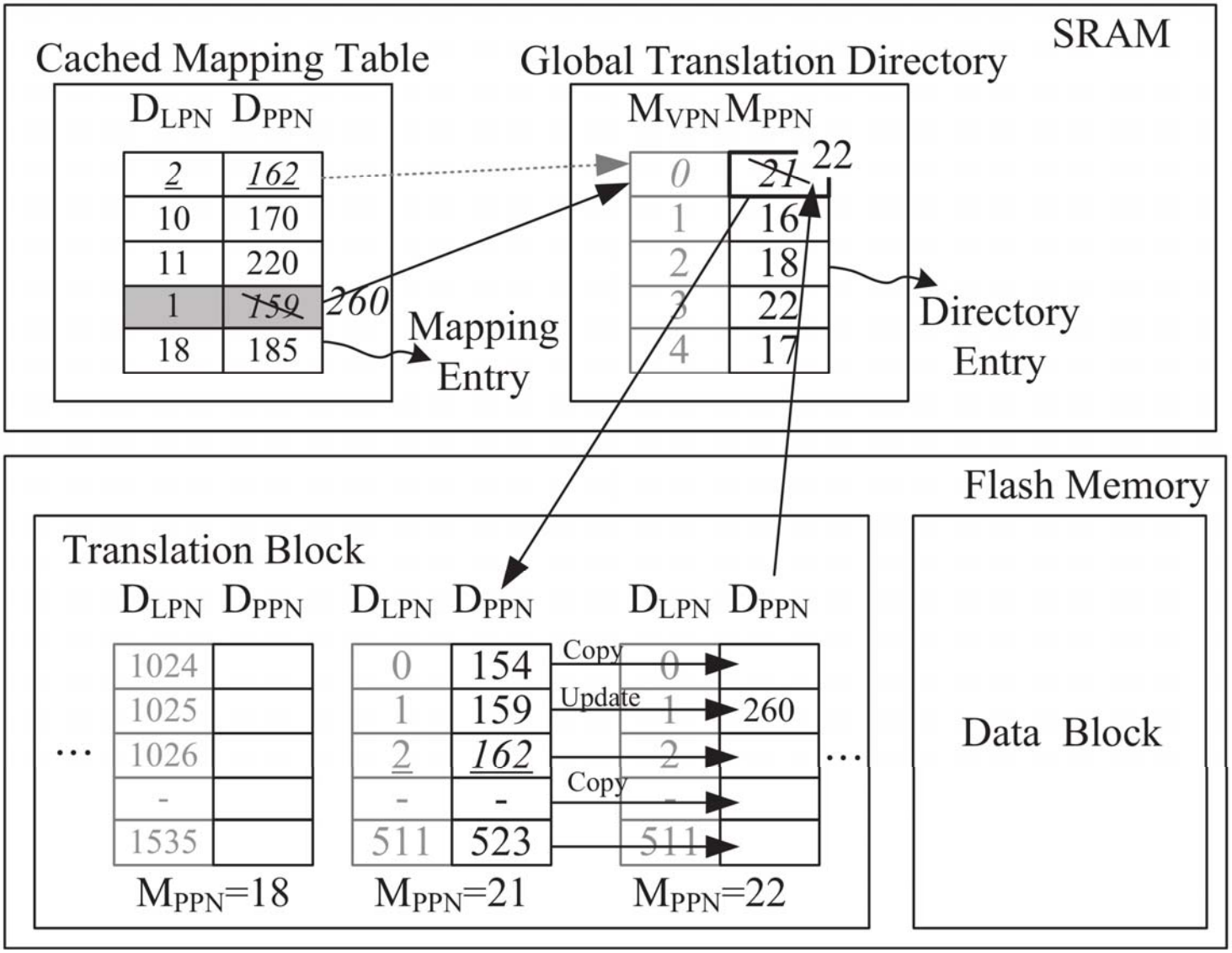
Case Study 18



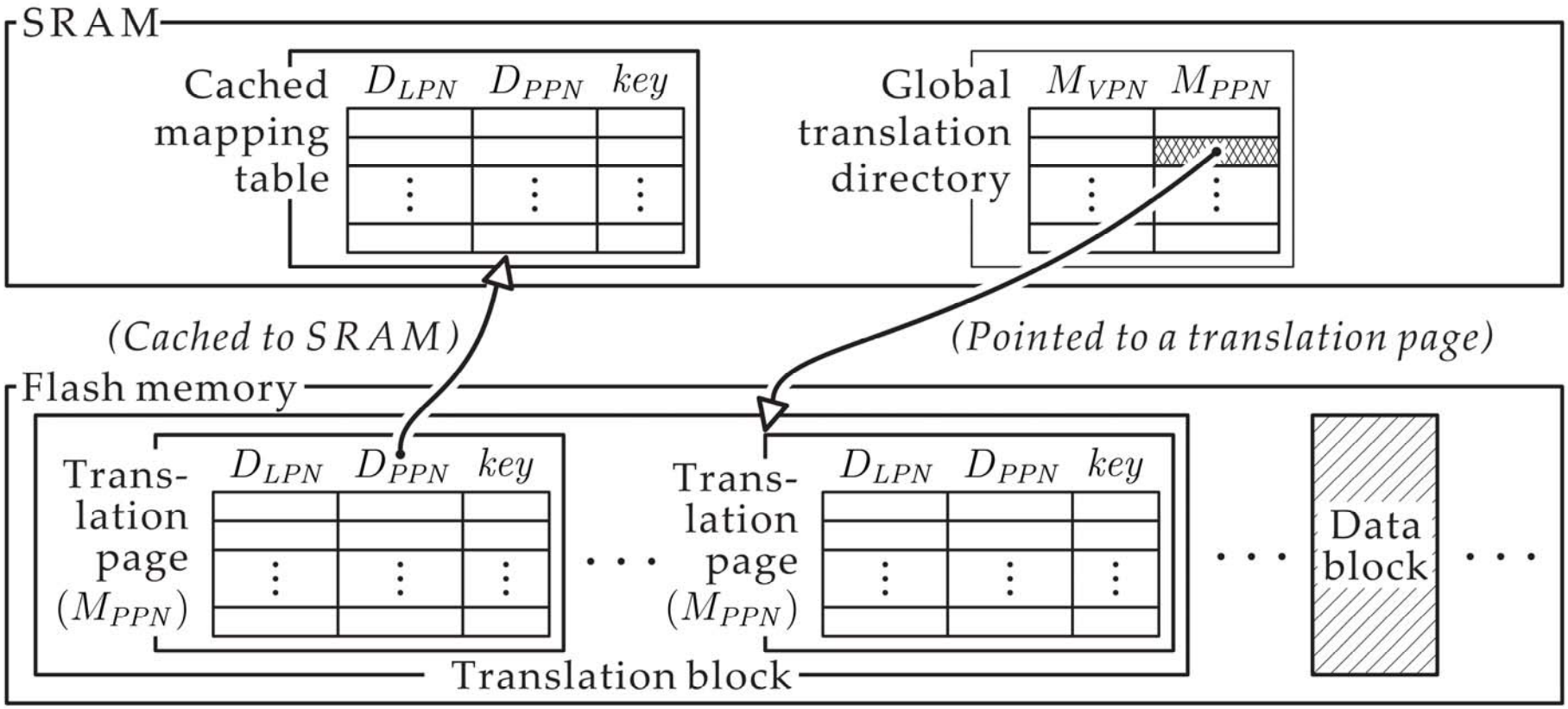
Case Study 19



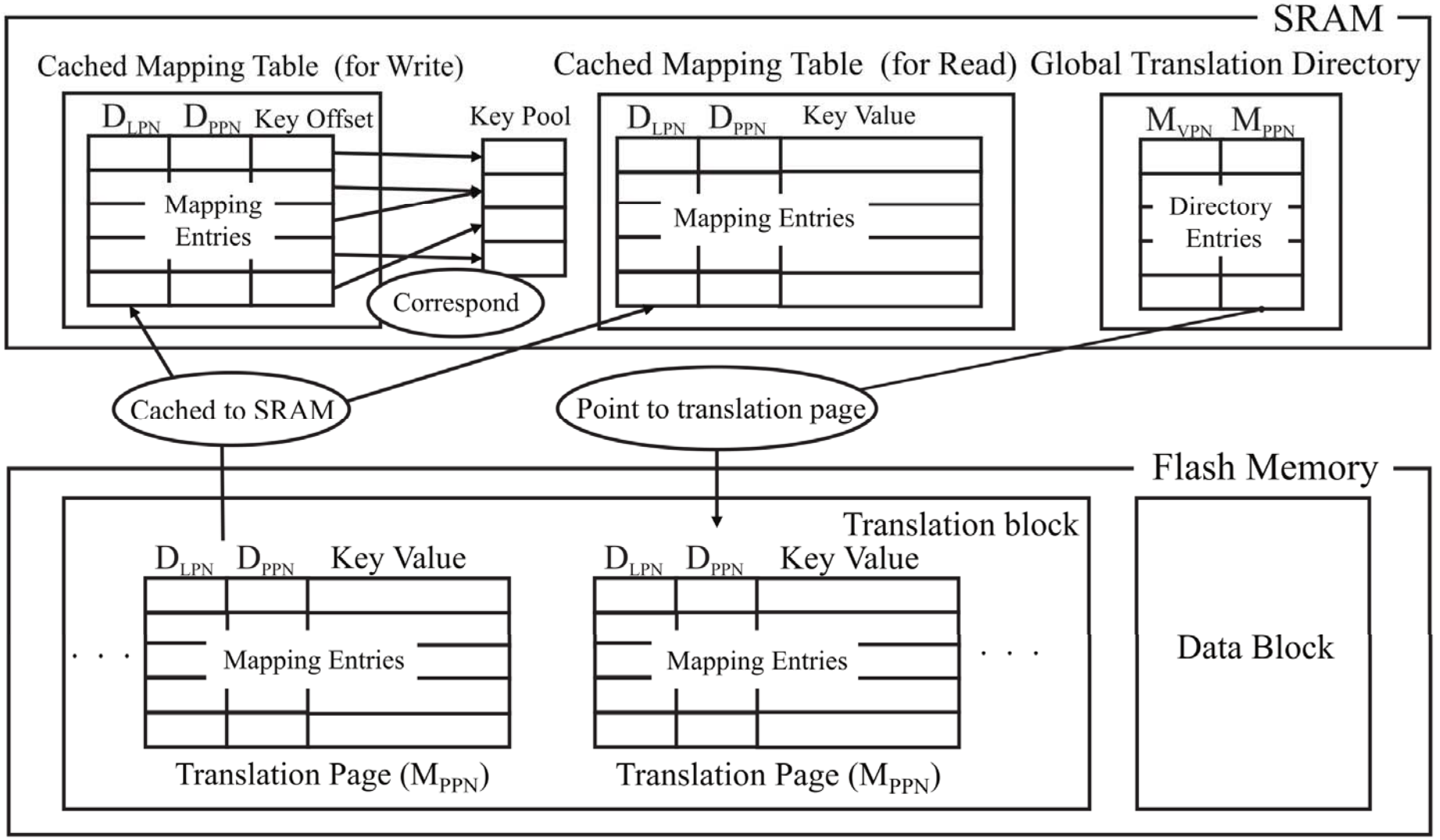
Case Study 20



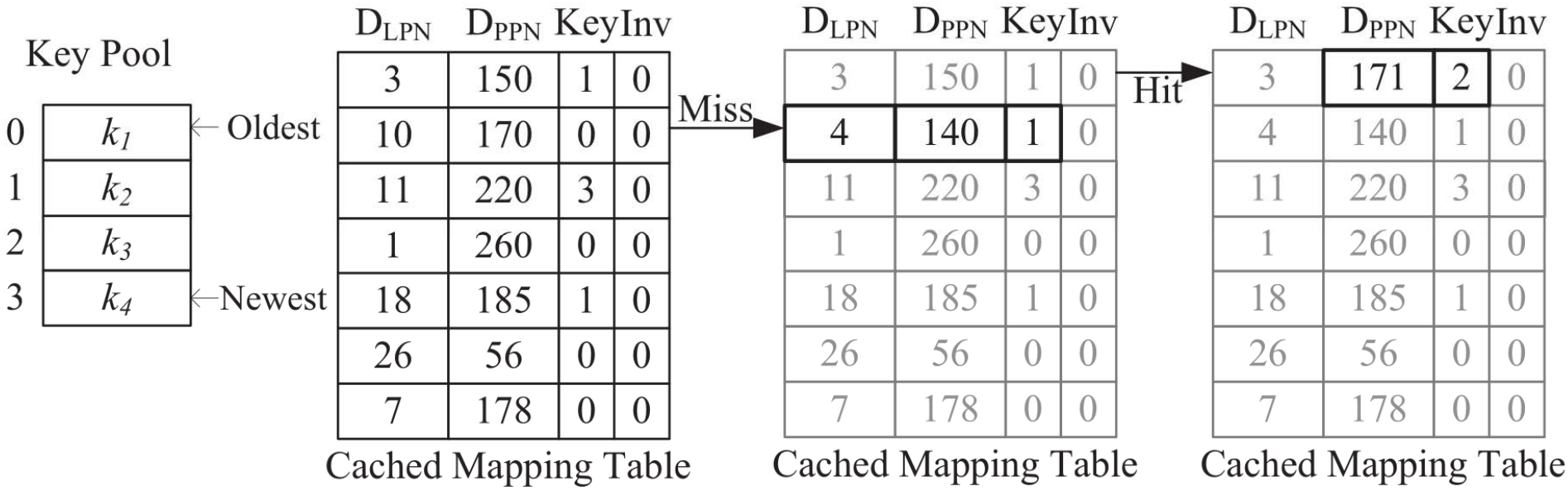
Case Study 21



Case Study 18

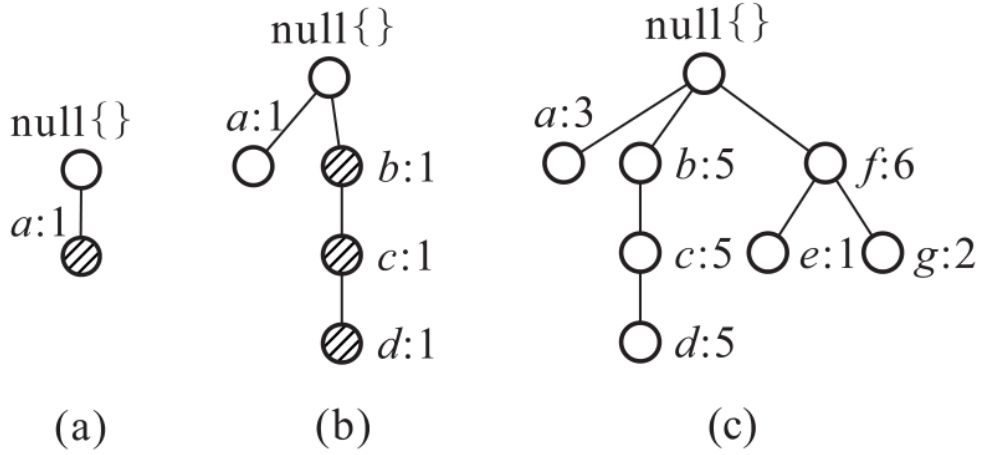


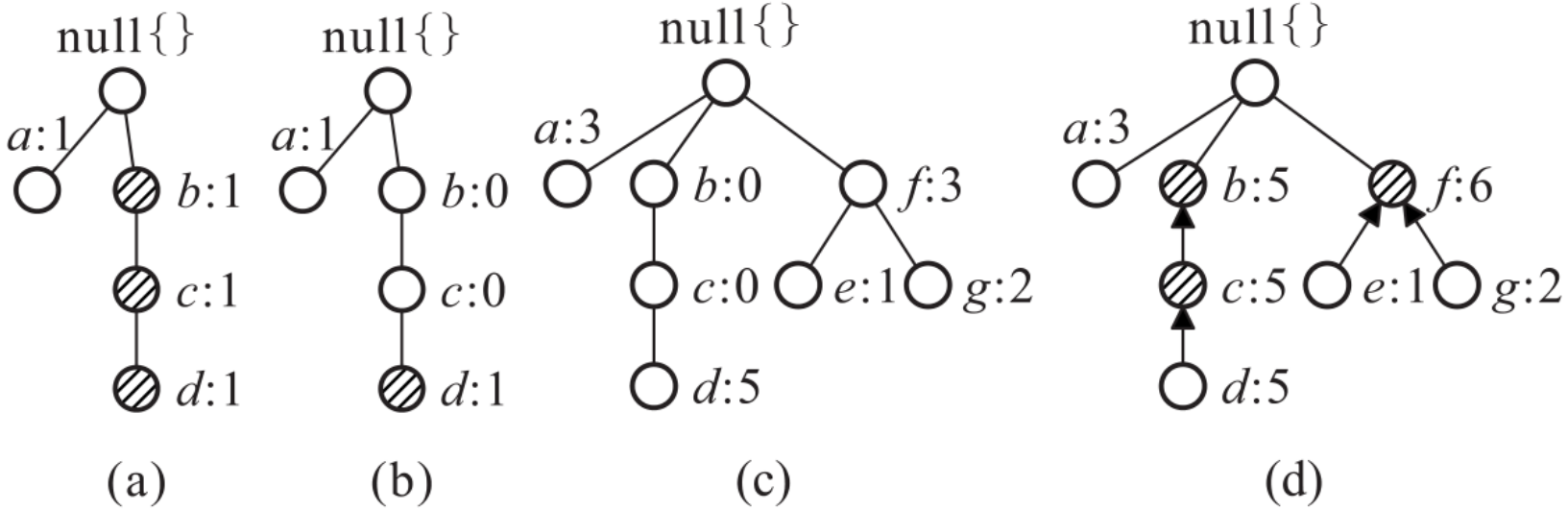
Case Study 18



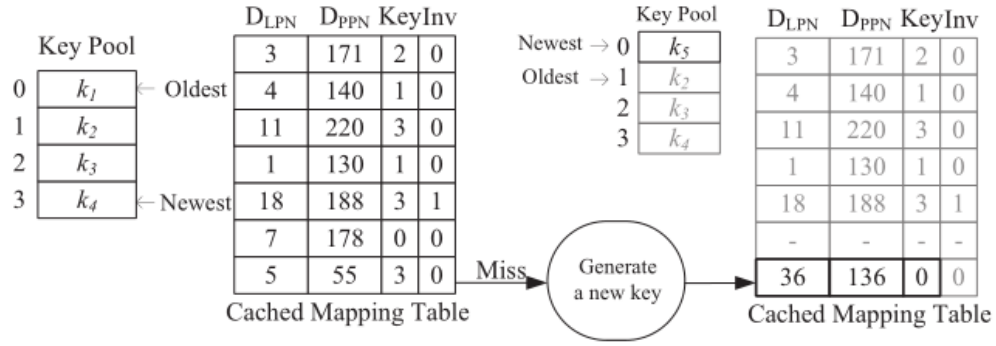
Case Study 18

TID	Data items
1	<*a*>
2	<*b, c, d*>
3	<*e, f*>
4—5	<*a*>
6—7	<*f, g*>
8—10	<*f*>
11—14	<*b, c, d*>

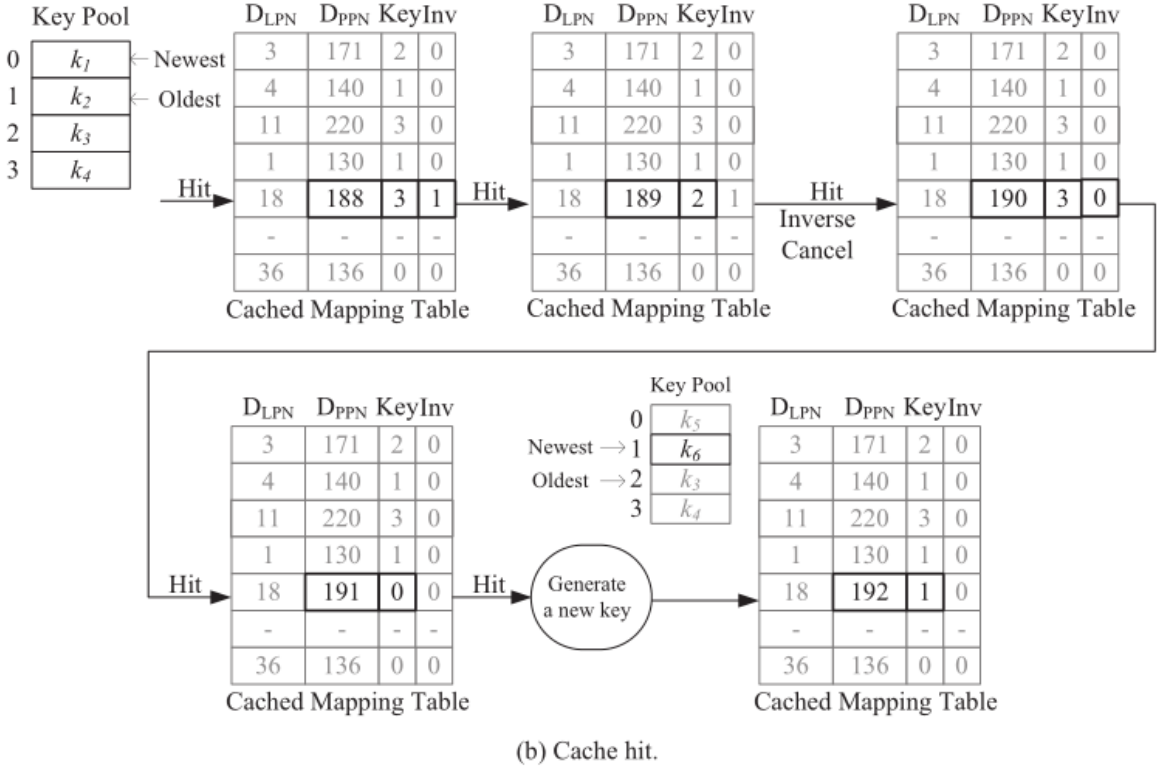




Case Study 18

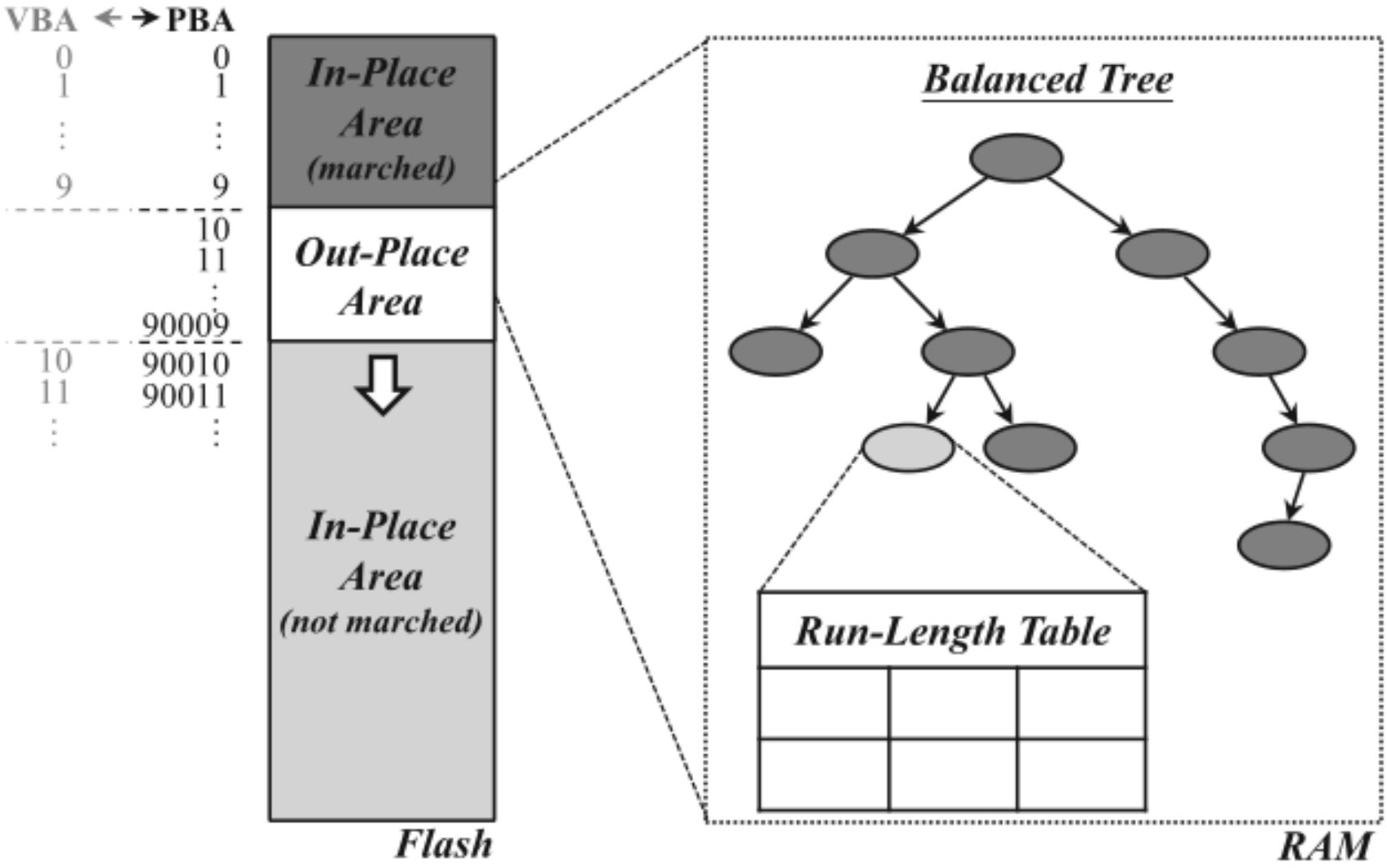


(a) Cache miss.

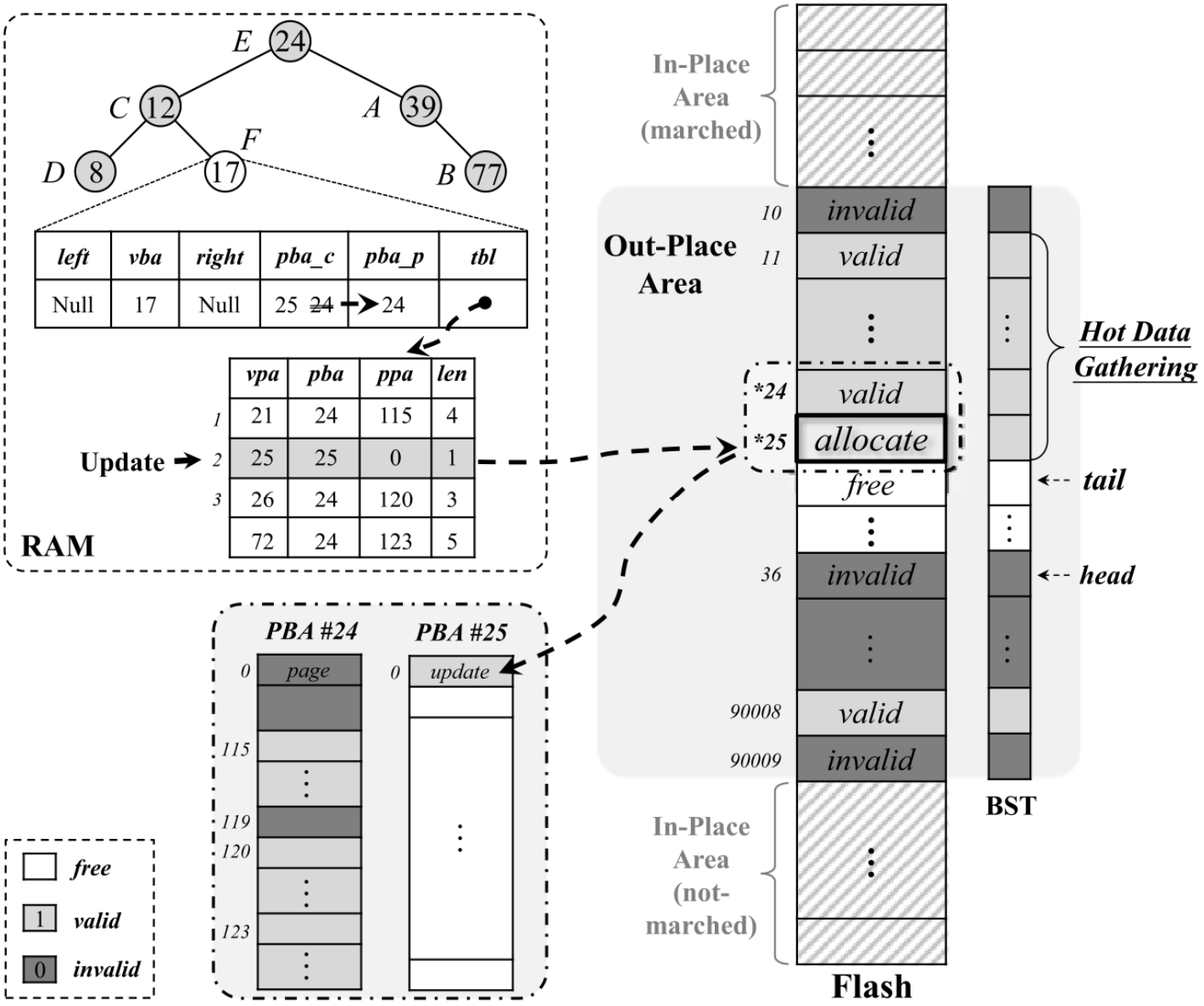


(b) Cache hit.

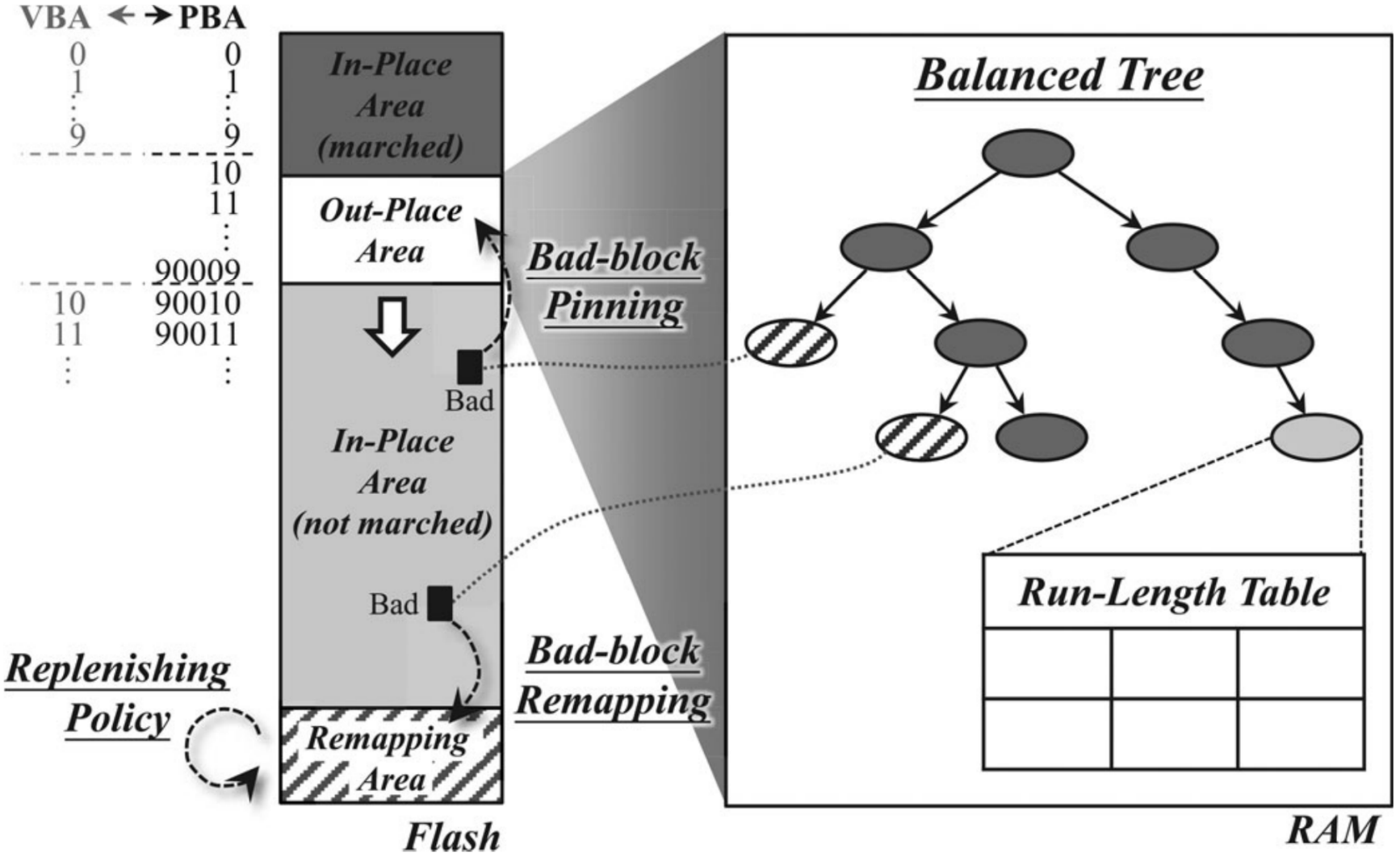
Case Study 19



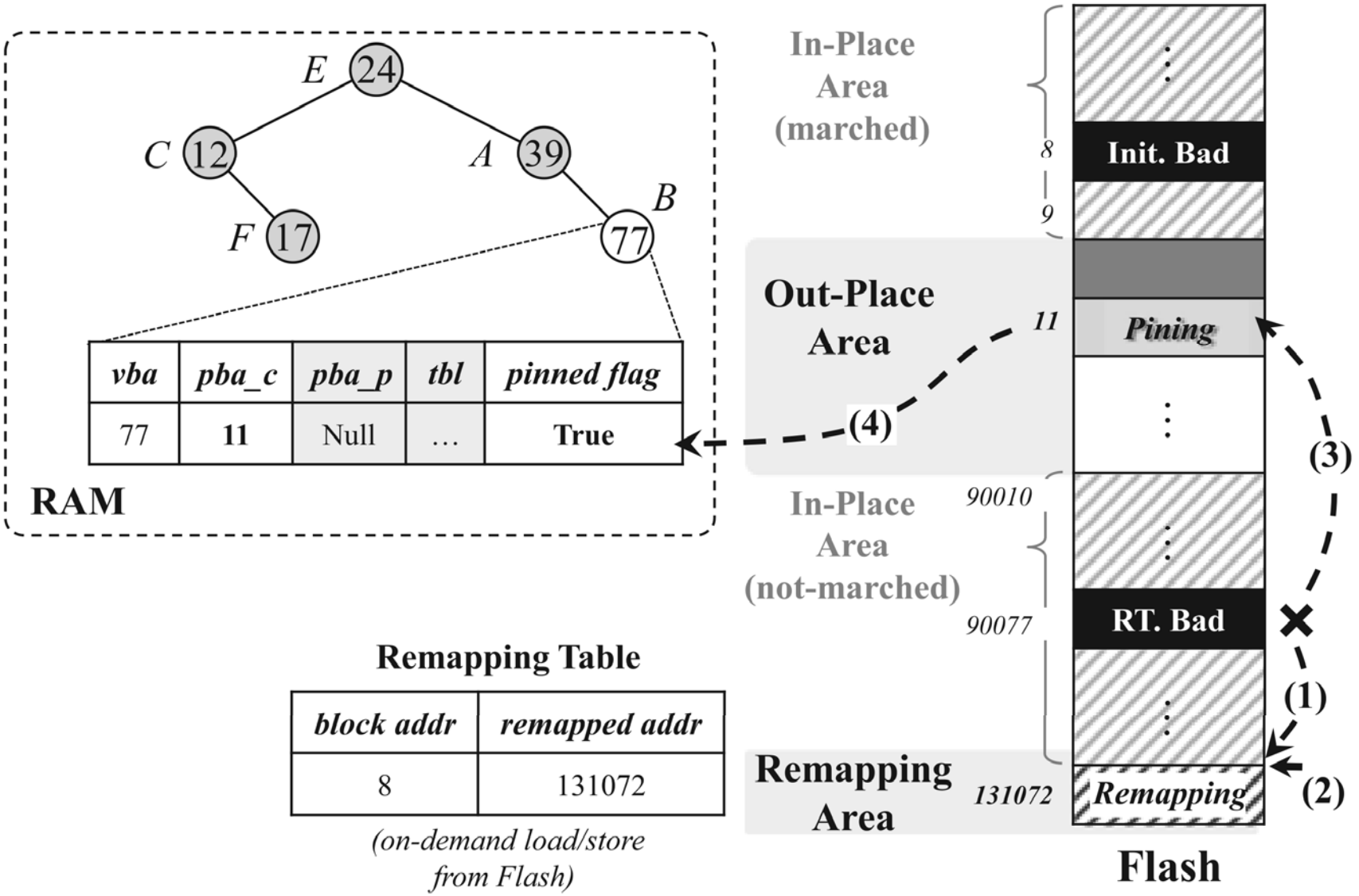
Case Study 20



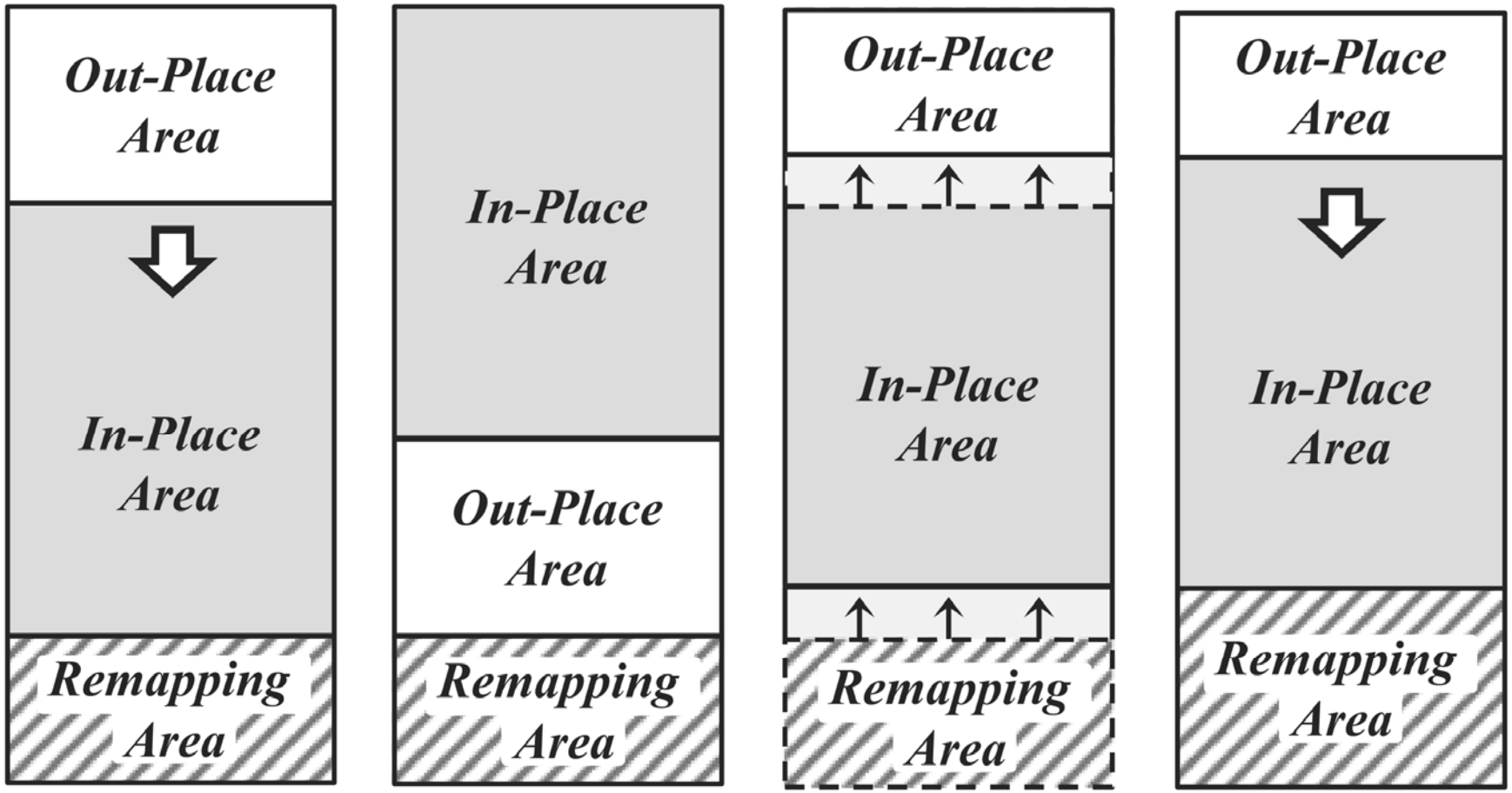
Case Study 21



Case Study 22

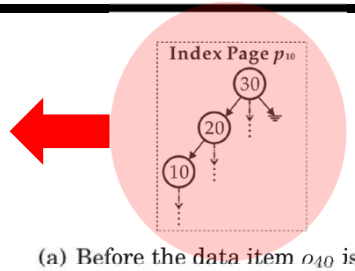


Case Study 23

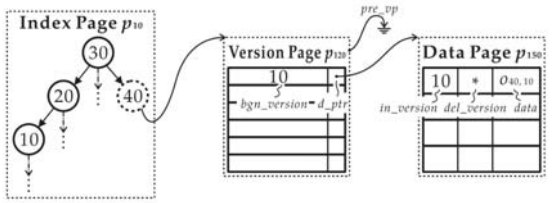


(a) Initially (b) Before (c) Replenish (d) After

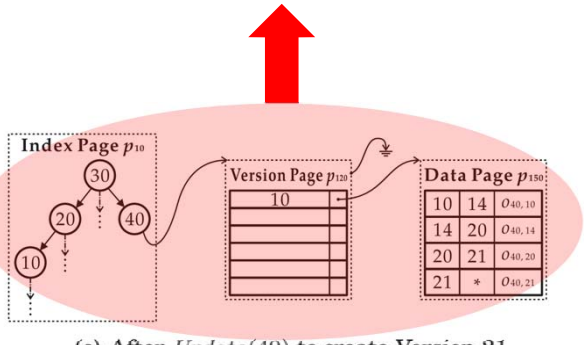
Case Study 24



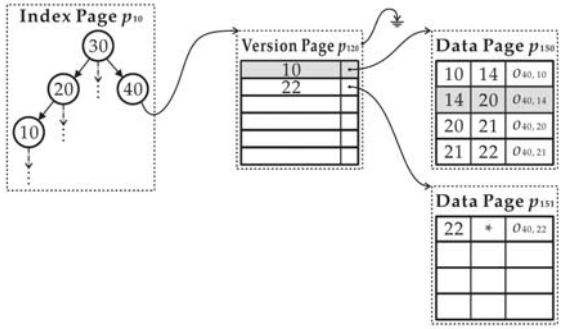
(a) Before the data item o_{40} is created



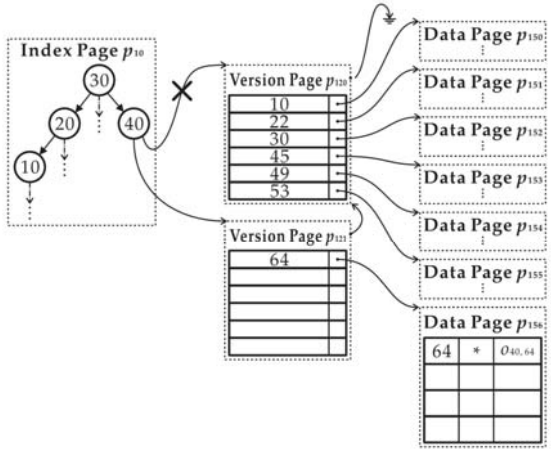
(b) After $Update(40)$ to create Version 10



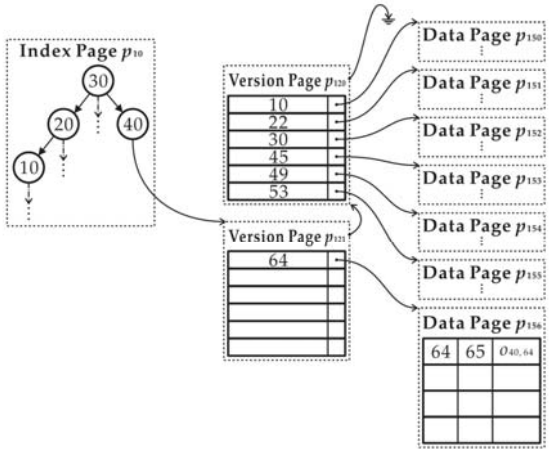
(c) After $Update(40)$ to create Version 21



(d) After $Update(40)$ to create Version 22

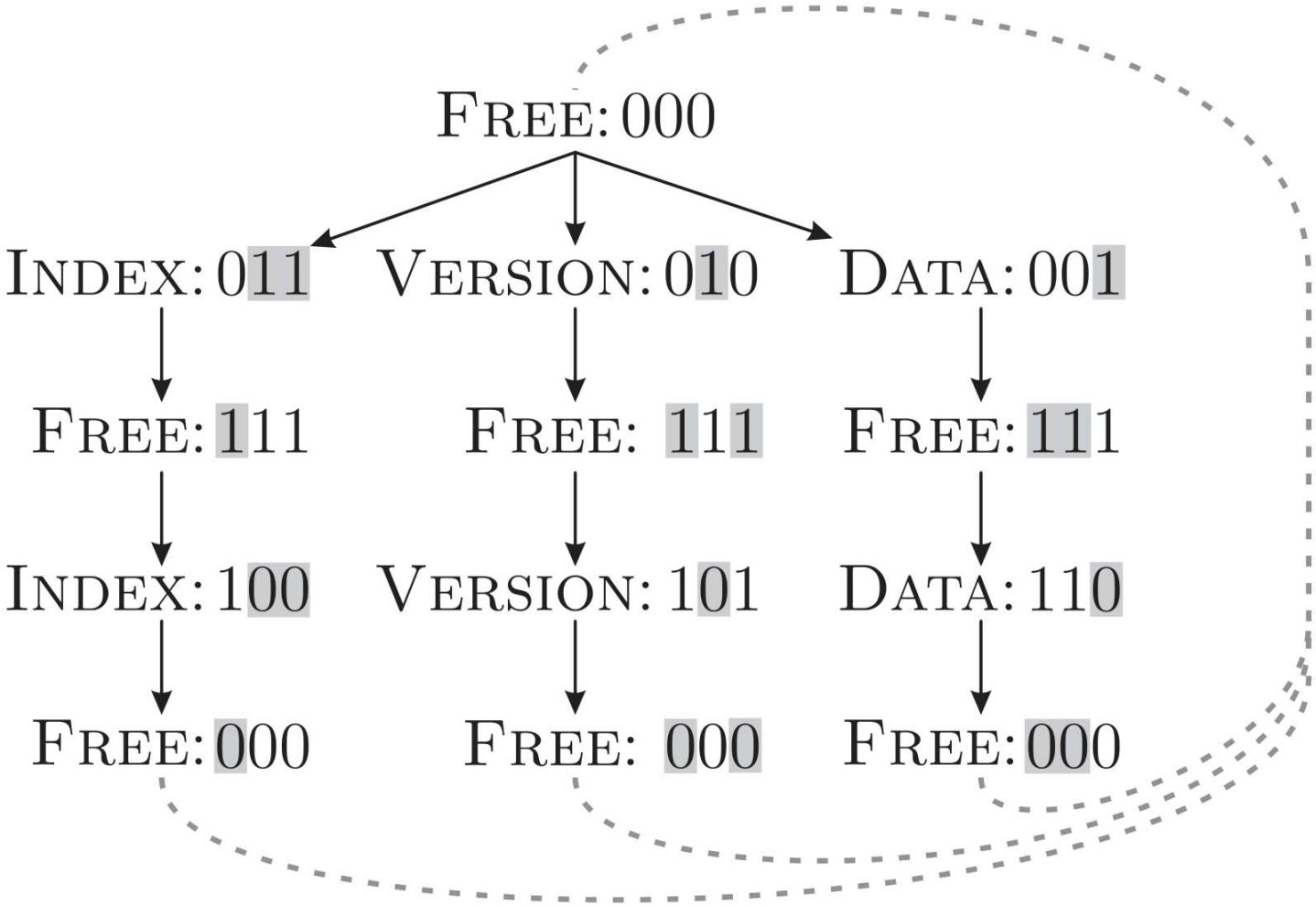


(e) After $Update(40)$ to create Version 64

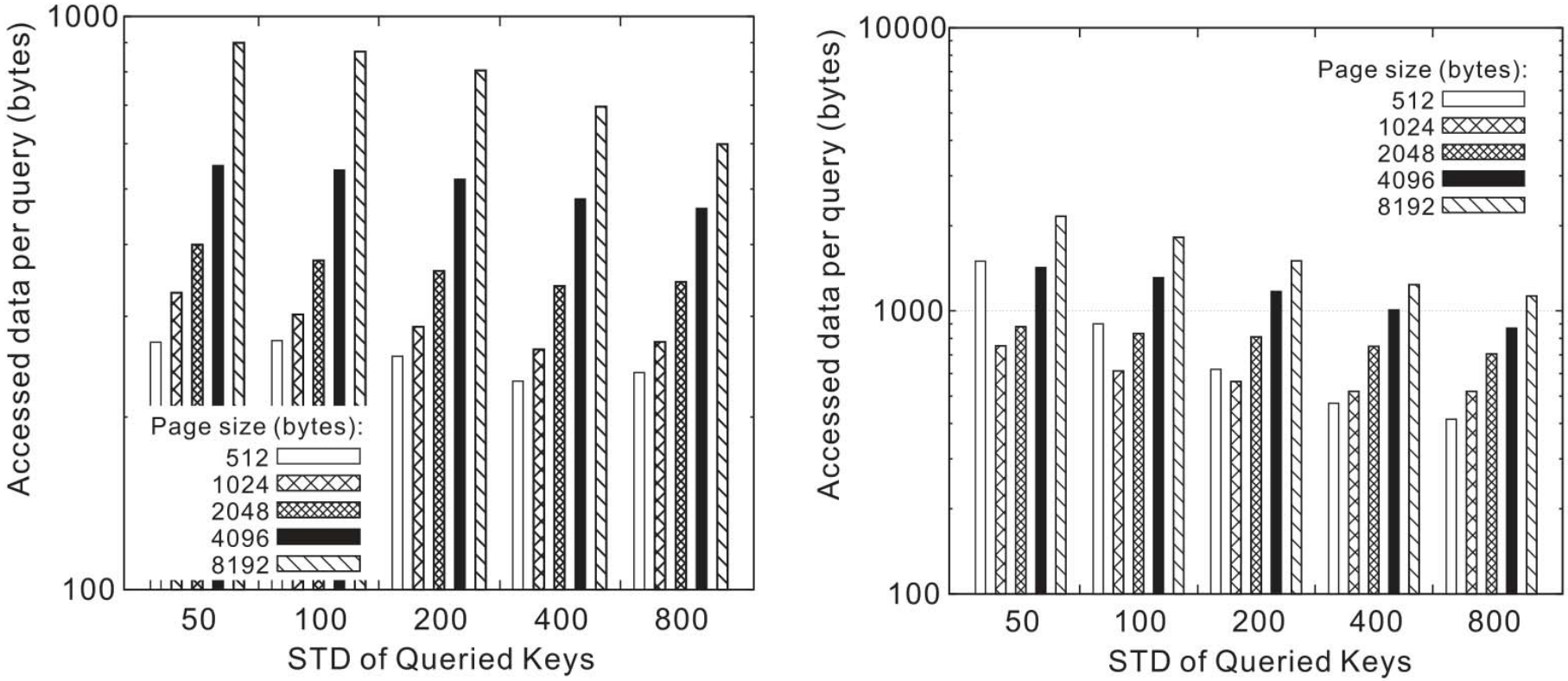


(f) After $Delete(40)$ to create Version 65

Case Study 25



Case Study 26

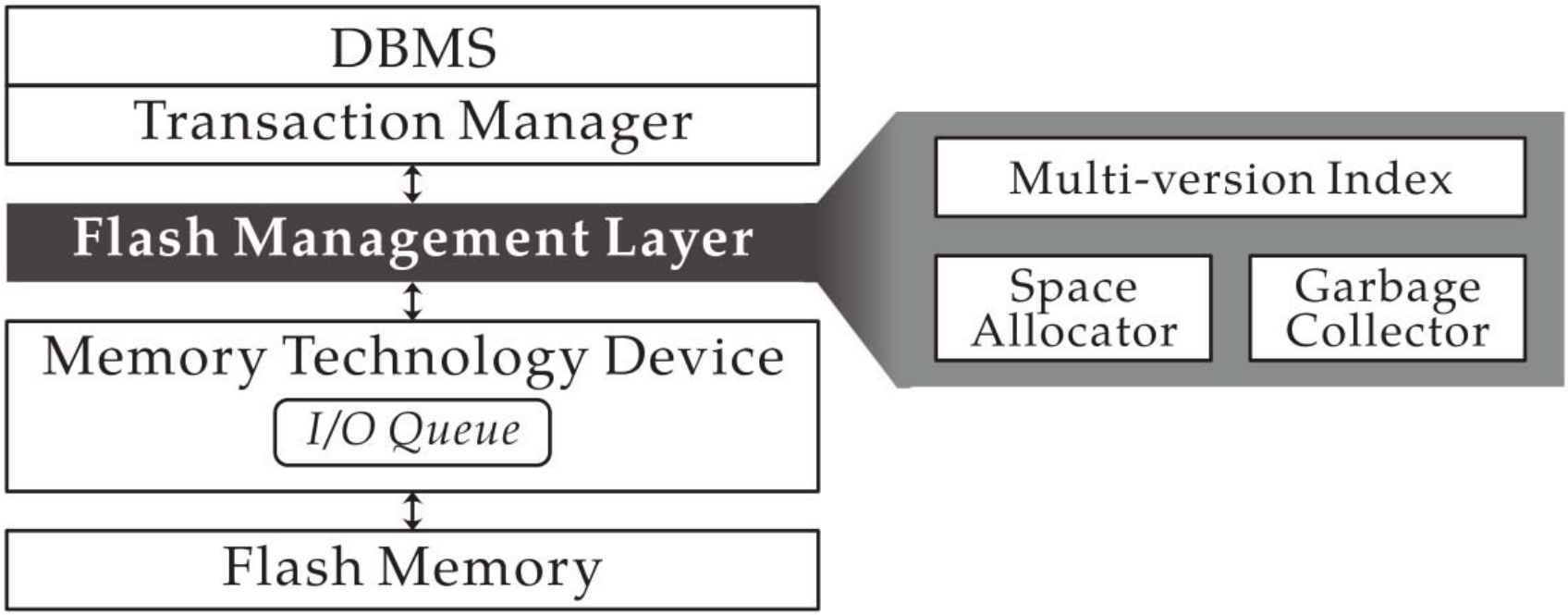


(a) Size of test case: 10^6 operations

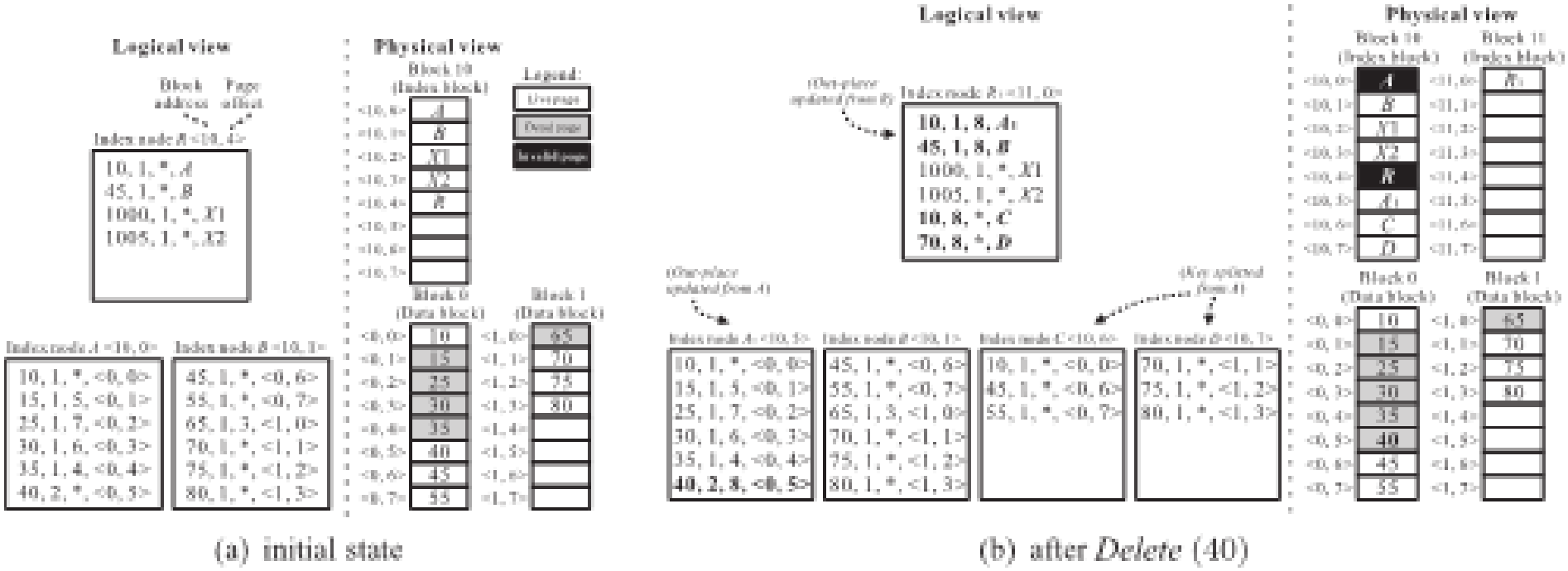
(b) Size of test case: 10^8 operations

Fig. 13. The cost for exact-match queries in SEMI.

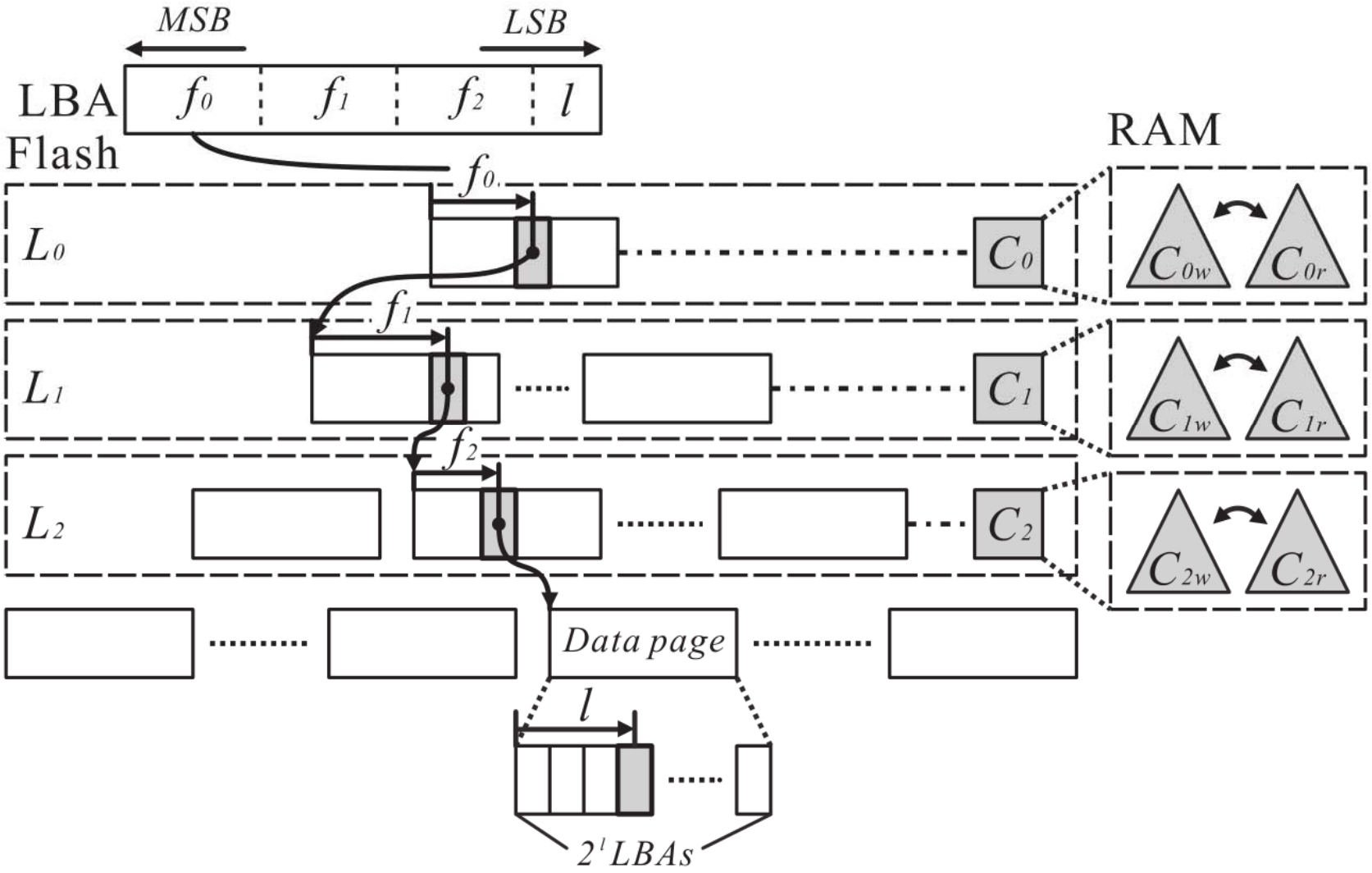
Case Study 27



Case Study 28



Case Study 29



Thank You Very Much!

Q & A ?