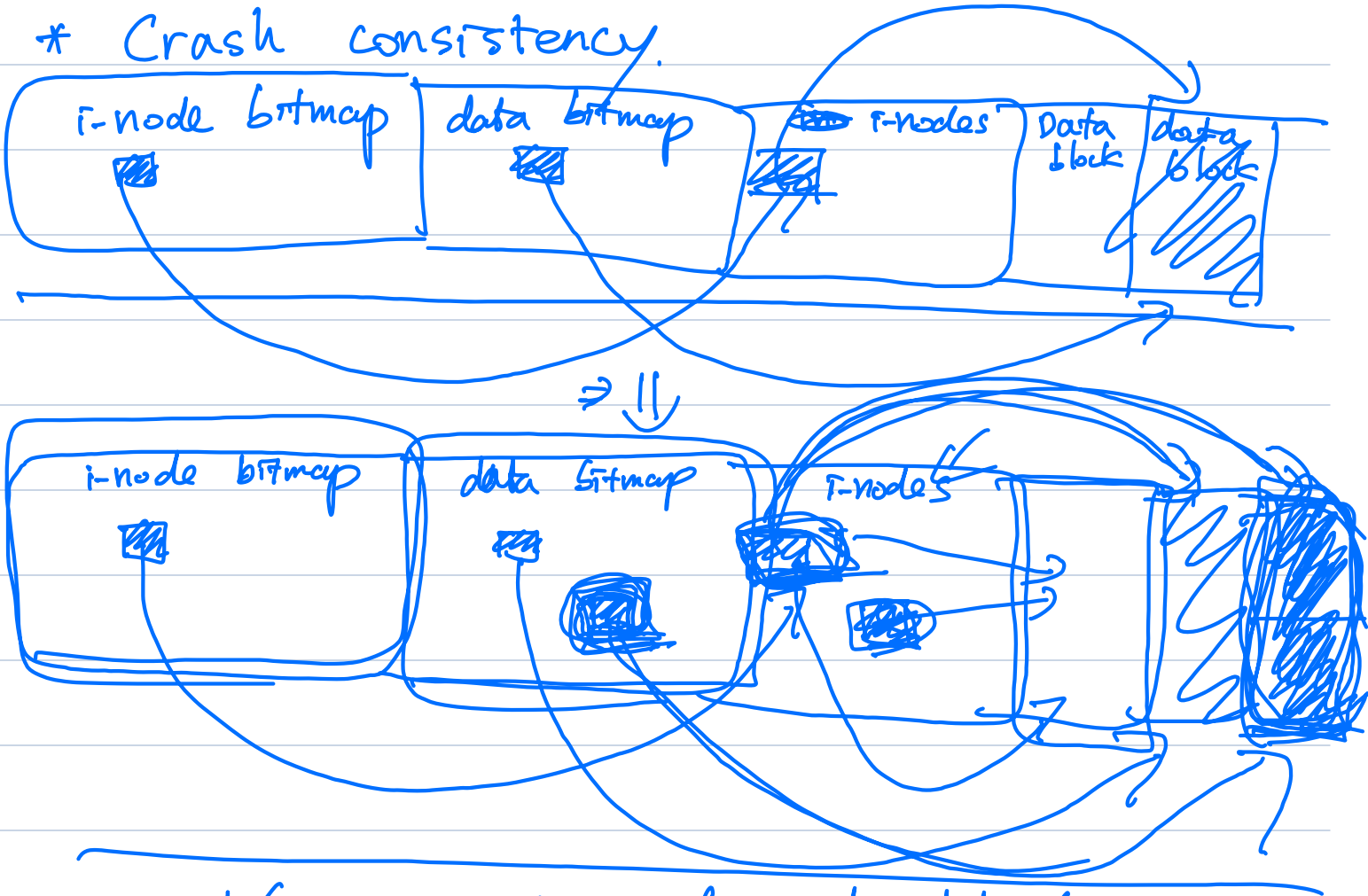
* Crash consistency.



read/write operations of i-node. data bitmap. data block

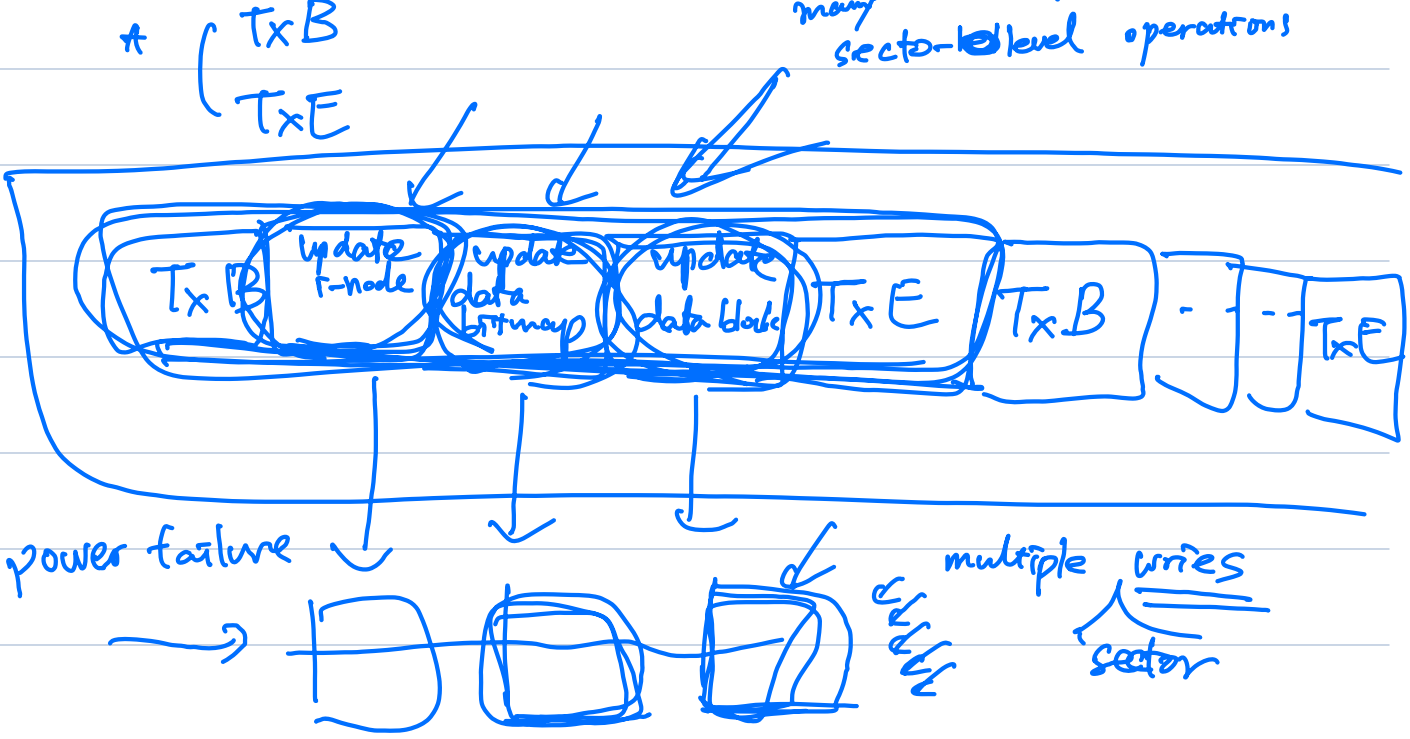
→ power failure

⇒ inconsistency problem.

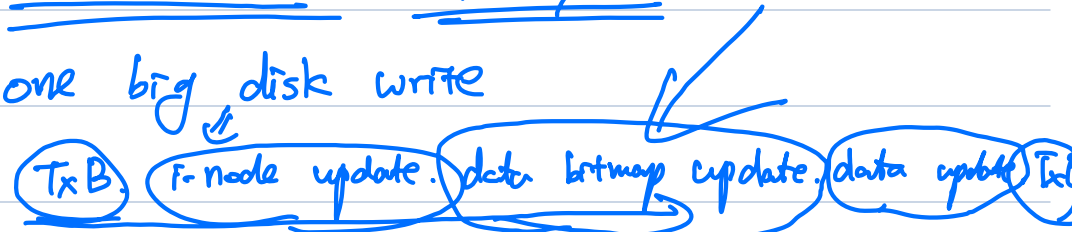
* fsck
  + Detection mechanism for file system inconsistency problems.
  * Super block. : file system size
  * i-node : duplication ⇒ if two i-nodes point to the same data block.
  * ~~i-node bitmap~~ data bitmap. : compares to the i-node pointers.

* Journaling

* Journal block.
  * Write all the operations that you need to perform in the journal block. then once that's done. you actually perform the operations
* Privides a solution for consistency.
* Transaction.
  * A group of operations that should occur atomically.
    either everything or nothing.
* Journaling uses logs
  A log should record all disk operations that should occur atomically.
* TxB
  TxE



* Atomic write for a disk
  ⇒ sector level. 512 bytes
* Sending one big disk write



⇒ These are multiple sector writes ⇒ no guarantee.

* TxB. | i-node update. (i) | data bitmap update (d) | data update. (du)

one write ⇒

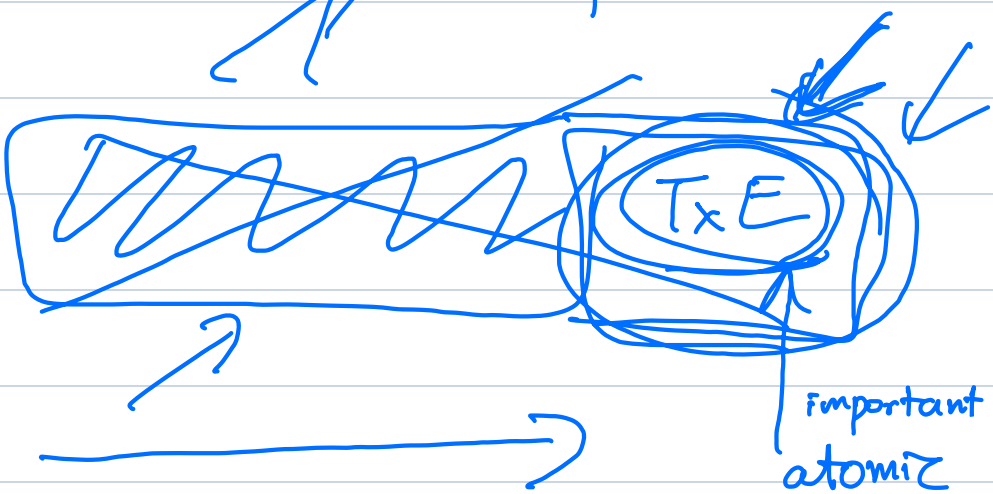
another write ⇒ TxE ← sector size.

power failure.

request to the disk
⇒ write the whole thing.

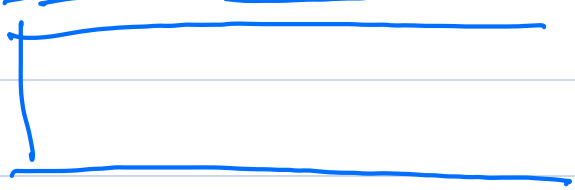
Journal.

i-node blocks
data blocks.



important to make this atomic (one sector maximum)

case 1: crash happens before writing to the journal

* 

(e.g., user buffer. kernel buffer.)

⇒ losing the write.

* Case 2: crash happens after writing the journal

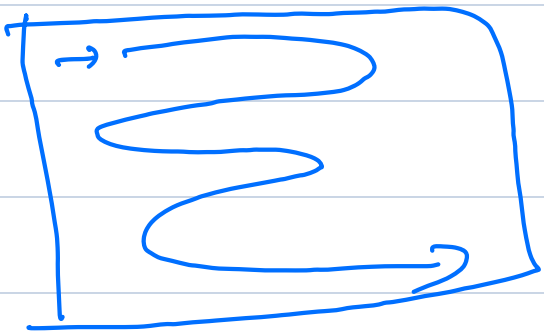
but before commit.

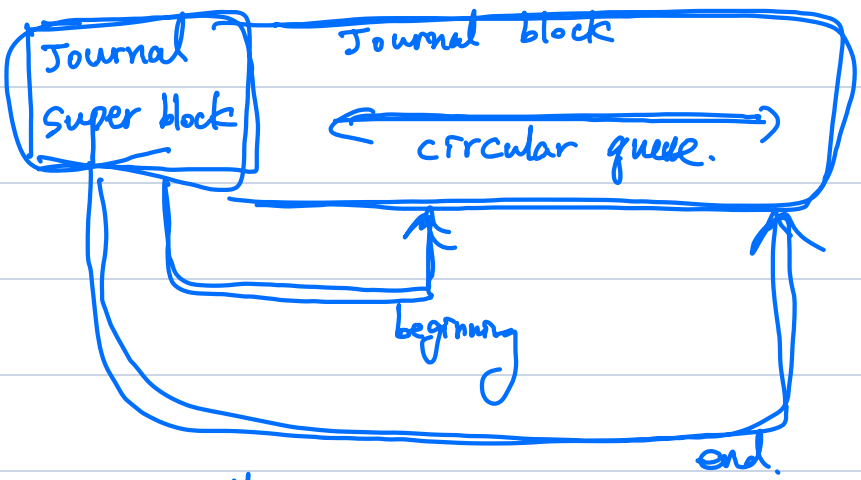
⇒ replay the logs. ⇒ restore the data.

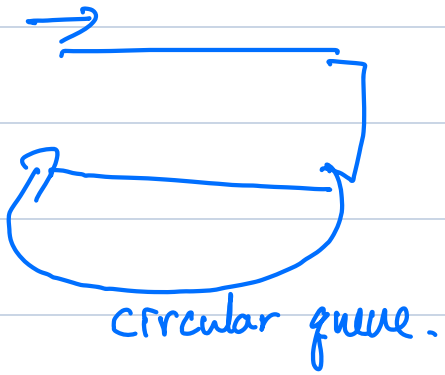
* Case 3: crash happens during commit.

⇒ replay the logs ⇒ overwrite what's done.

* Journal block. ⇒ finite







* Overall operations.
  * Write to the journal (transaction)
  * Commit
  * Remove from the journal.
  * Adjust the journal block queue pointers.
* Metadata journaling vs. Data journaling.
  - Just log metadata in the journal.
  - write everything to the journal including data

⇒ Attempt 1:

① ~~A~~ log the metadata.
- Tx B. i-node update. data bitmap update.
- Tx E.

power failure.

② Write the data. (to the actual data block)

⇒ Writing the data before committing the transaction.

~~①~~ ① data write

~~②~~ ② log metadata.

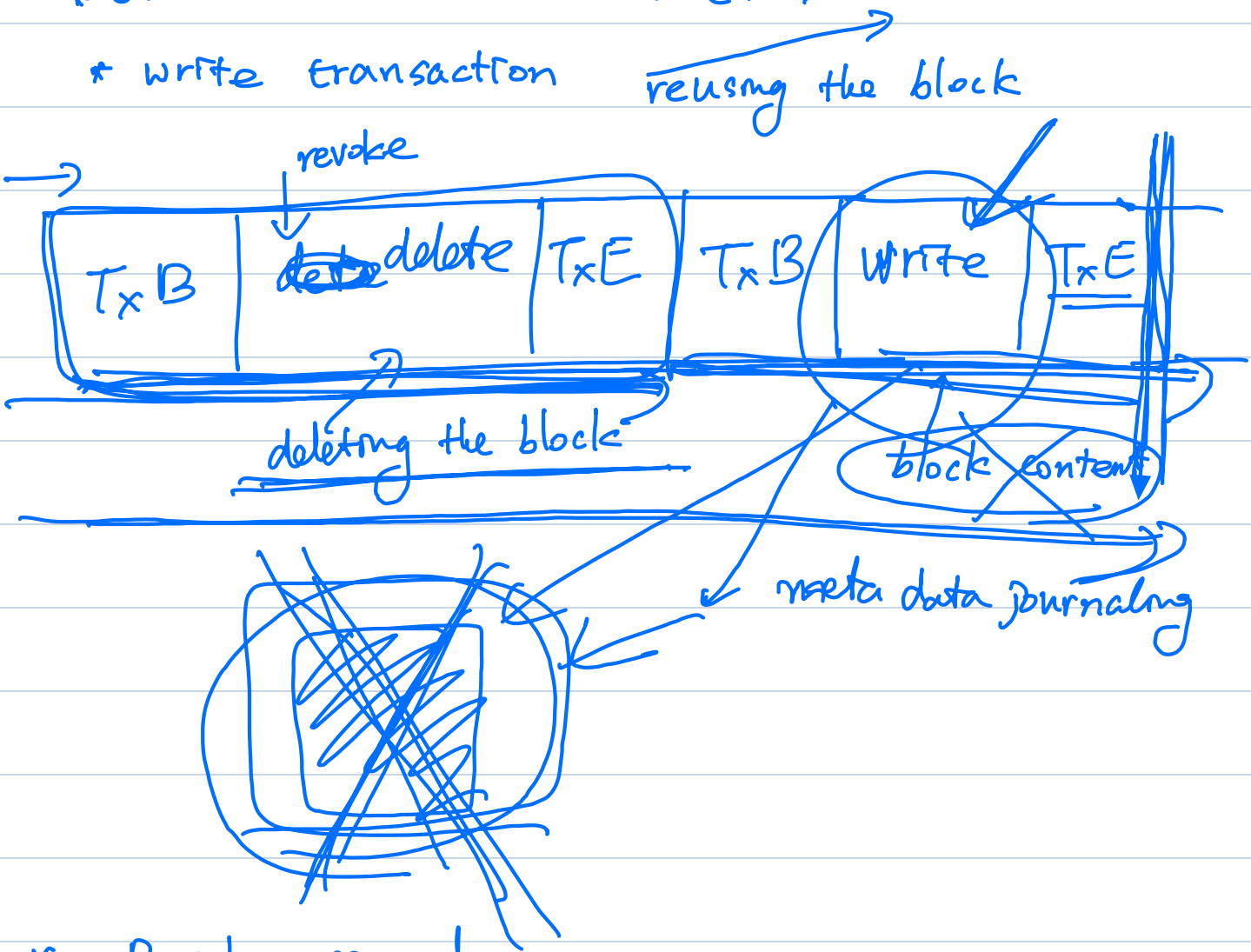
* Data Journaling.

	Journal			File system	
	Tx B	operations	Tx E	meta data	data
①	✓	✓			
②			✓		
③				✓	✓

* Meta data Journaling

	Journal			File system	
	Tx B	operations	Tx E	meta data	data
①					✓
②	✓	✓			
③			✓		
④				✓	✓

* Corner Case for meta data journaling
  * one transaction is delete. => block deletion
  * write transaction reusing the block



* Revoke record.

Additional resource:
https://pages.cs.wisc.edu/~remzi/OSTEP/file-journaling.pdf