

Utreexo - What is it?

What it is and how it works

Utreexo Agenda for today

What is it? How does it work?

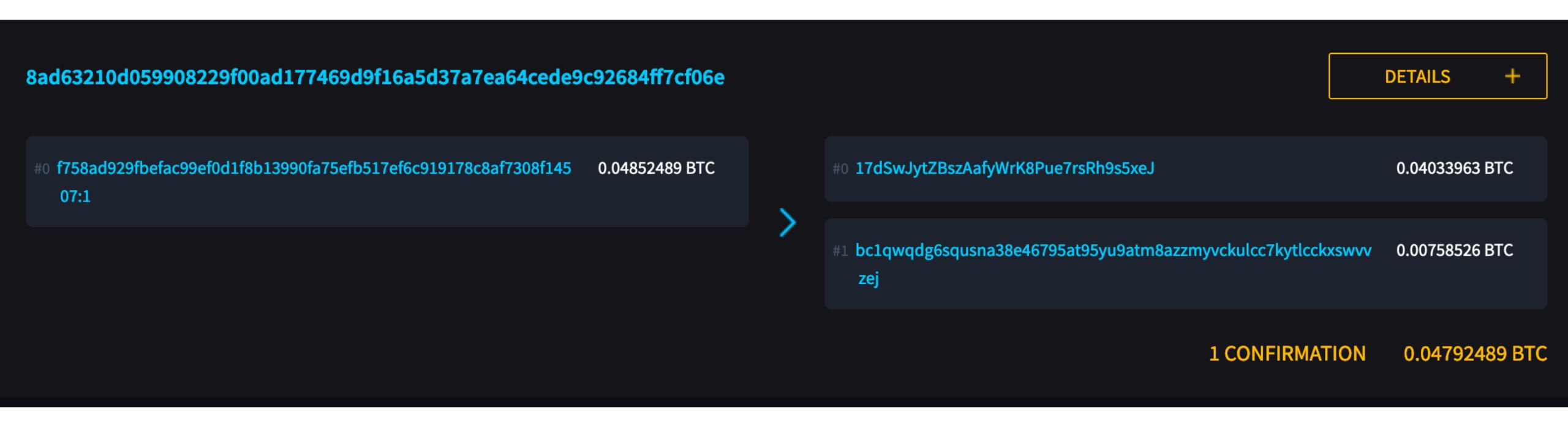
What is it?

Turn the UTXO set into a Merkle tree

Quick overview

•TXO - An Output of a TX (transaction)

Typical TX



Outputs of a TX



Quick overview

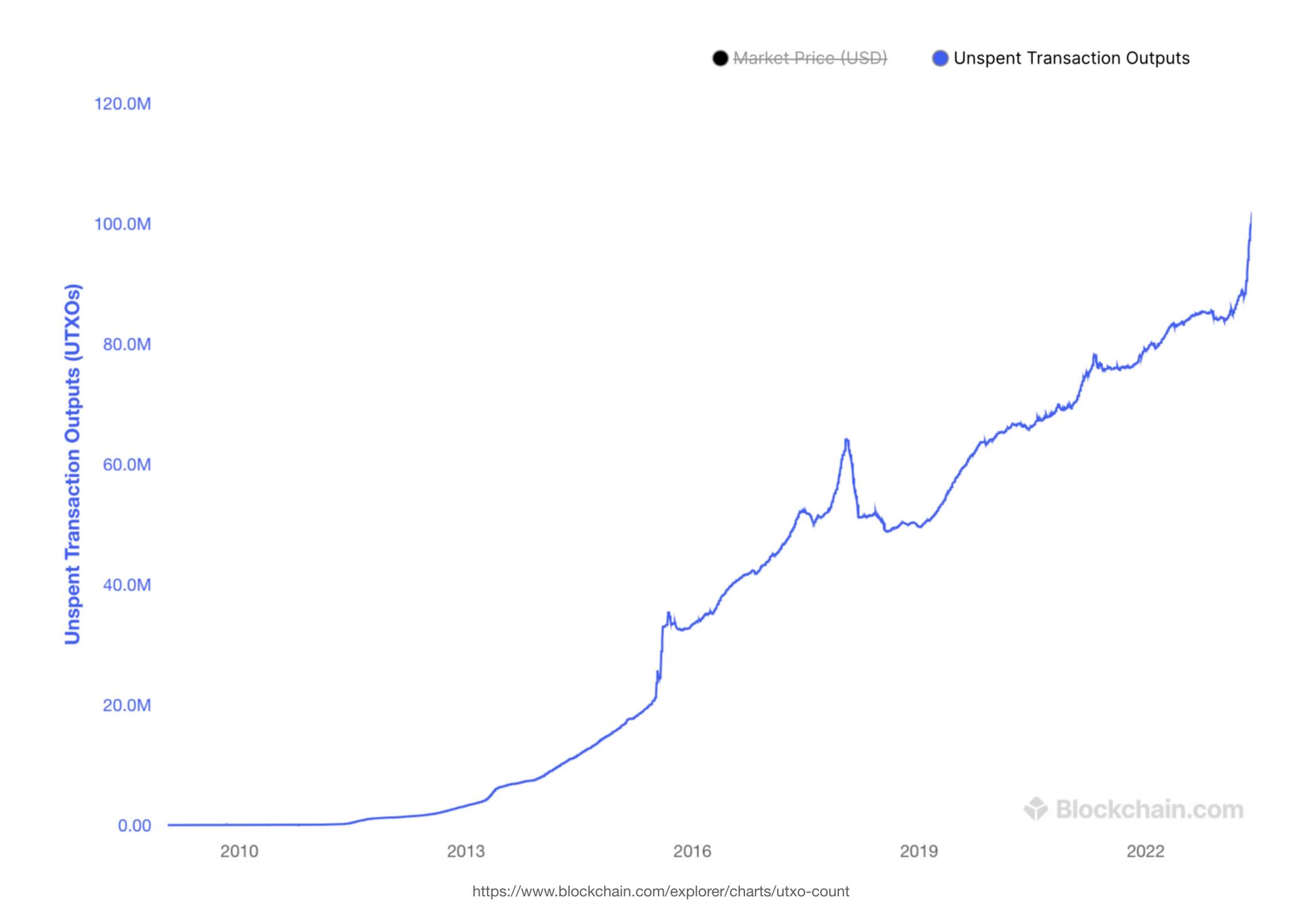
- •TXO An Output of a TX (transaction)
- •UTXO Unspent TXO

Quick overview

- •TXO An Output of a TX (transaction)
- •UTXO Unspent TXO
- •UTXO set Set of all UTXOs

Why merklize it?

Puts a bound to the UTXO set growth



O(N) -> O(log N)

N=8 billion UTXOs, at 59B per entry

472GB->608B

Why merklize it?

- Puts a bound to the UTXO set growth
- Allows for tiny nodes

5.8G chainstate/

384B chainstate/

Immediate node bootstrap

AssumeUTXO

Quick way to sync up

 Receive a UTXO set hash commitment with the binary

AssumeUTXO

Quick way to sync up

- Receive a UTXO set hash commitment with the binary
- Download the ~5GB of UTXO set from torrents

AssumeUTXO

Quick way to sync up

- Receive a UTXO set hash commitment with the binary
- Download the ~5GB of UTXO set from torrents
- Download blocks from peers

AssumeUtreexo

Quicker way to sync up

Receive the UTXO set with the binary

AssumeUtreexo

Quicker way to sync up

- Receive the UTXO set with the binary
- Download blocks/tx from peers



† Download Bitcoin Core

Why merklize it?

- Puts a bound to the UTXO set growth
- Allows for tiny nodes
- Faster block validation

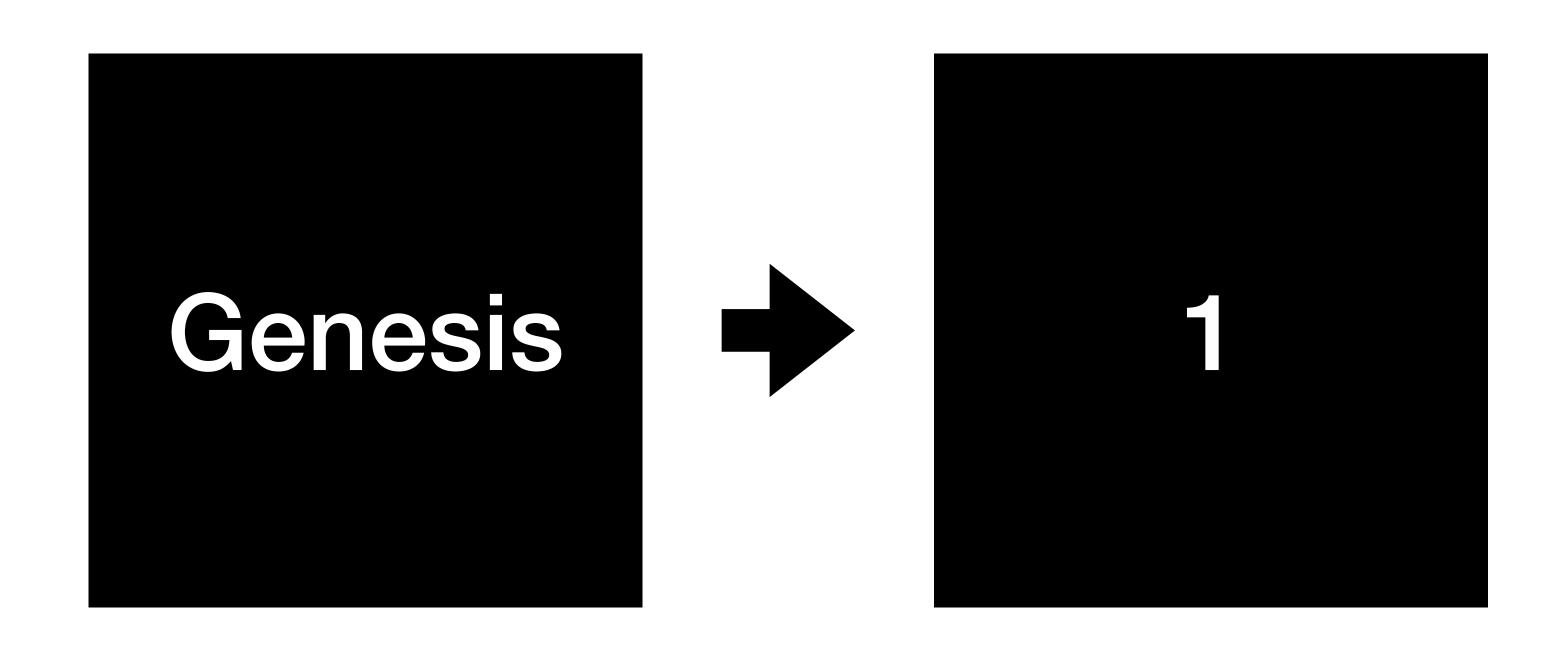
Current block validation

Sequential validation



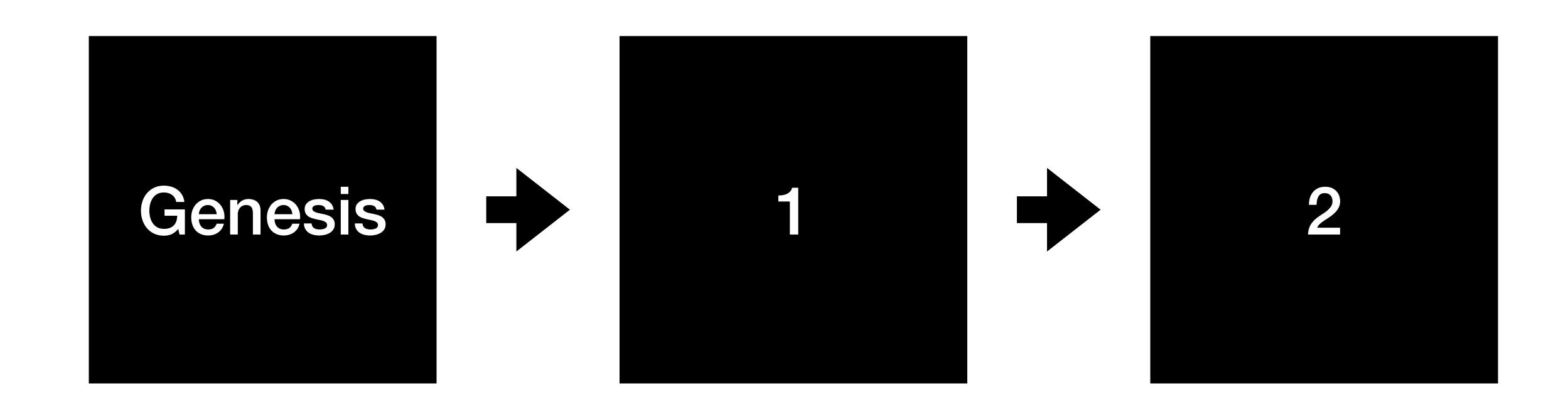
Current block validation

Sequential validation



Current block validation

Sequential validation



Parallel block validation

Replay blocks in any order

- Utreexo roots committed into the binary
- Process blocks starting from any of the roots that are committed
- https://github.com/mit-dci/utcd/blob/master/ chaincfg/mainnetroots.go

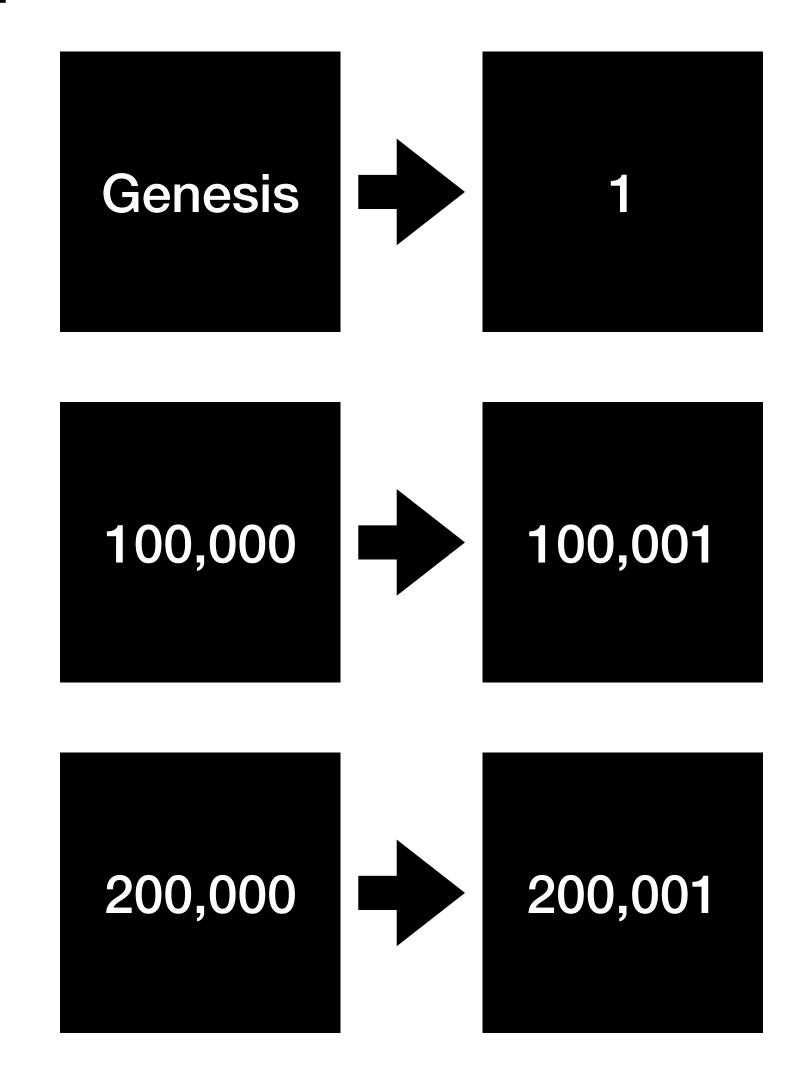
Efficient parallel validation

Genesis

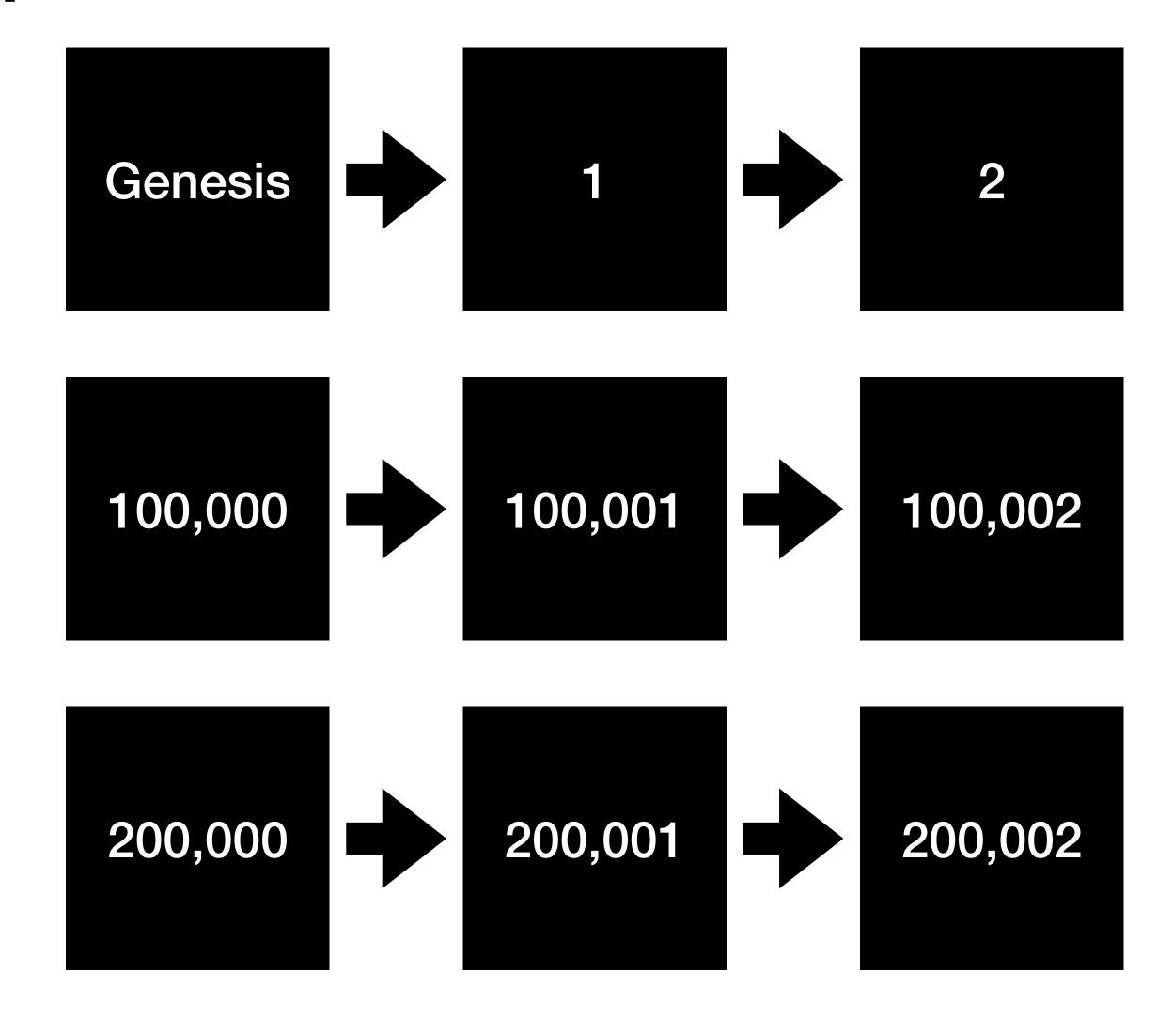
100,000

200,000

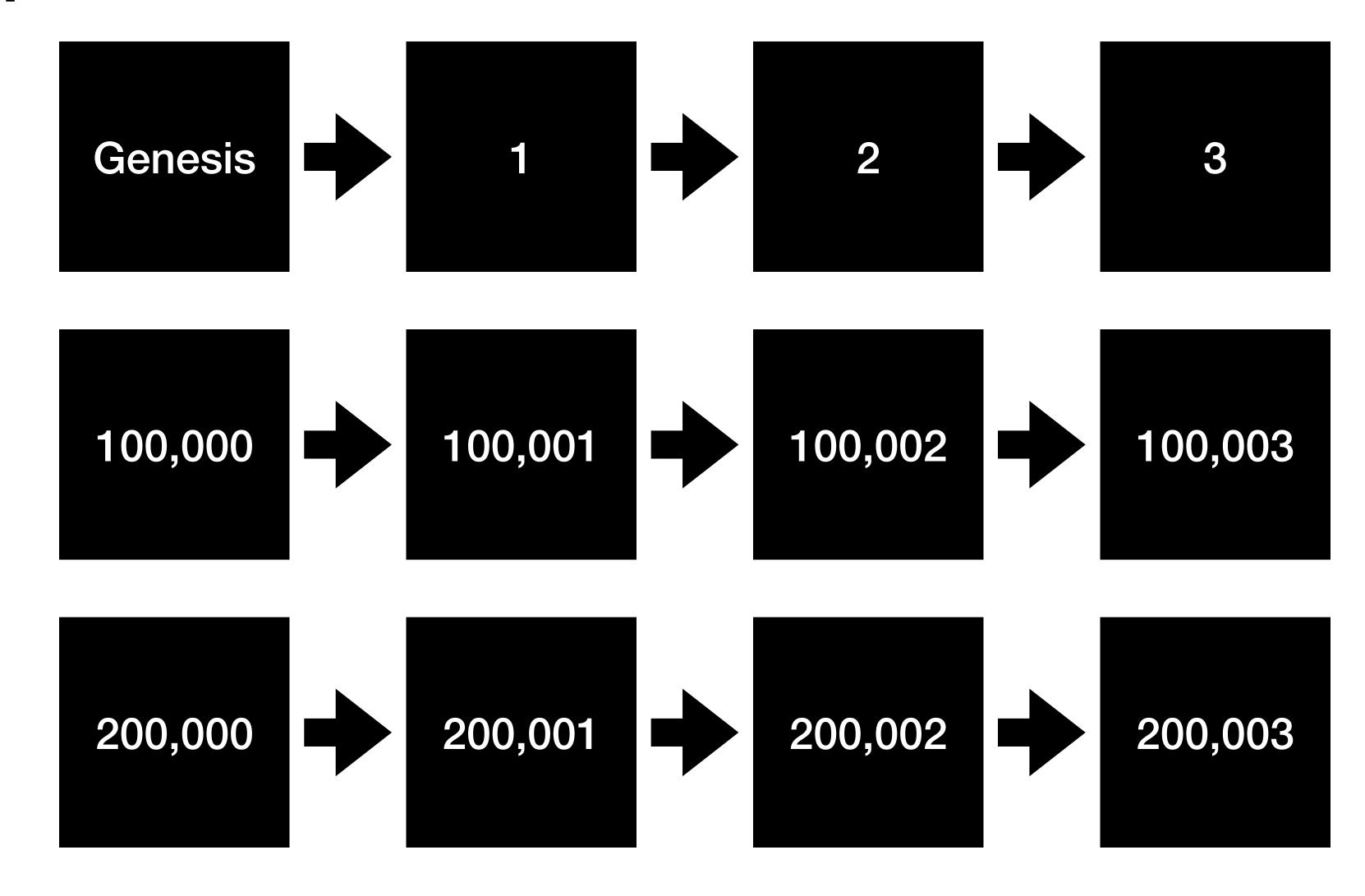
Efficient parallel validation



Efficient parallel validation



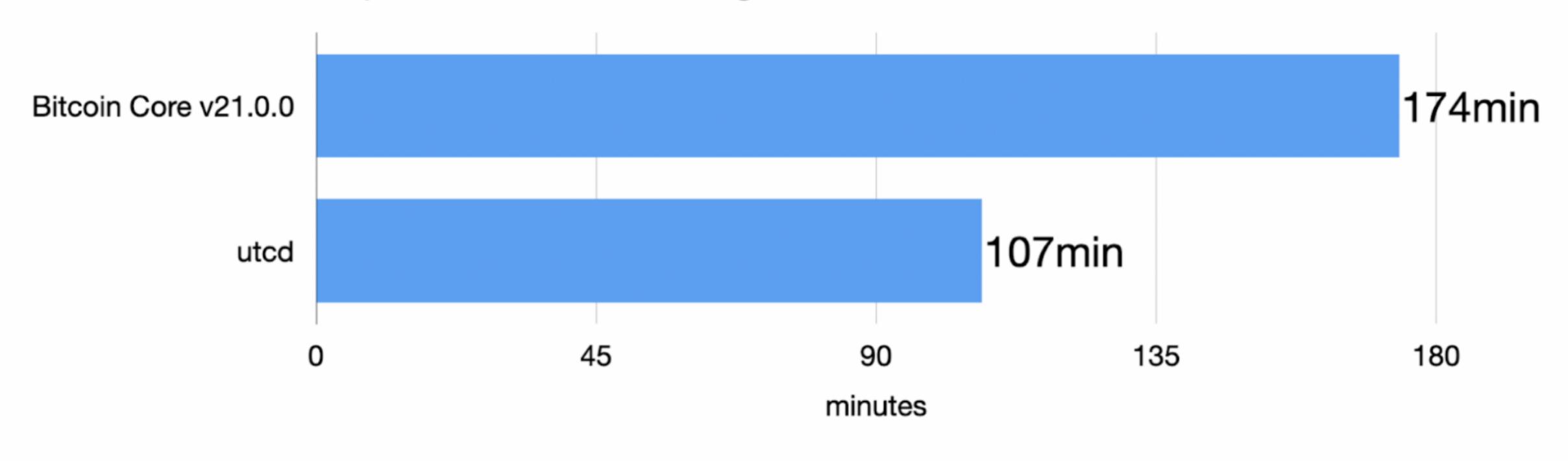
Efficient parallel validation



Efficient Parallel Validation

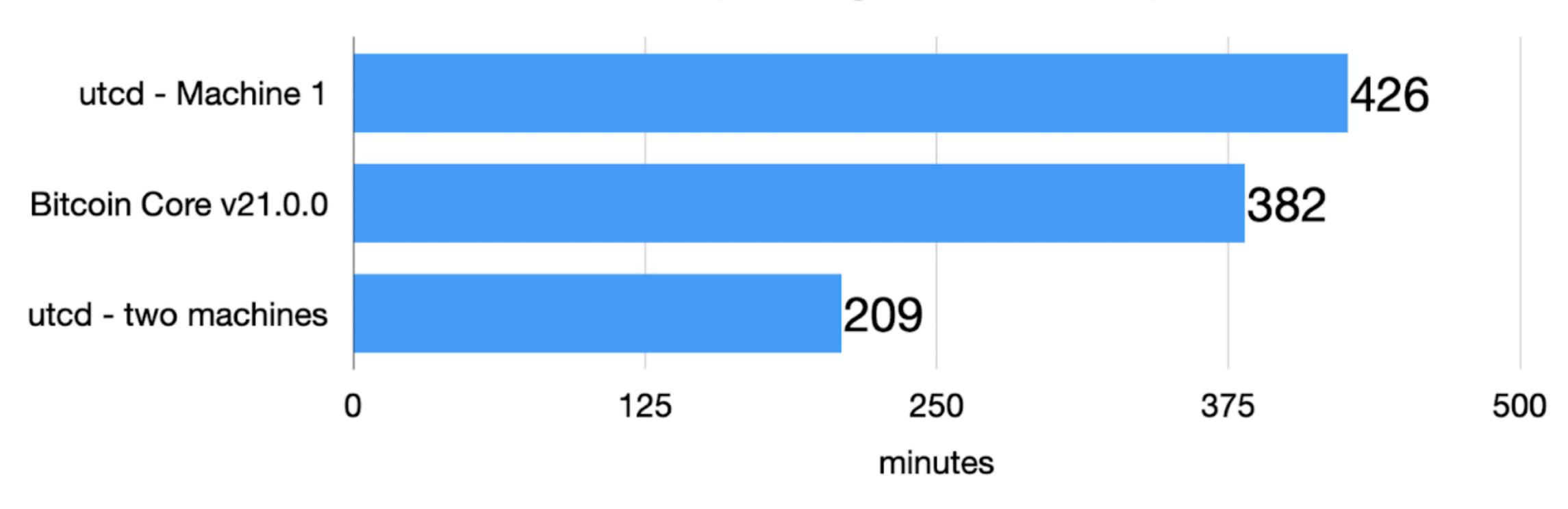
SHA256 > Map access >

Initial Block Download Speeds - local nodes (default mode, no signature checks until block 654,683



Coordinator/worker worker worker worker

Initial Block Download Speeds to block 671,000 (Full signature check)



The UTXO set

Why merklize it?

- Puts a bound to the UTXO set growth
- Allows for tiny nodes
- Faster block validation
- Define consensus without leveldb

Basic steps

1. Check Header (PoW)

Basic steps

- 1. Check Header (PoW)
- 2. Fetch inputs from all TXs (LevelDB)

Basic steps

- 1. Check Header (PoW)
- 2. Fetch inputs from all TXs (LevelDB)
- 3. Verify signatures

Basic steps

1. Check Header (PoW)

Basic steps

- 1. Check Header (PoW)
- 2. Verify Utreexo proof

Basic steps

- 1. Check Header (PoW)
- 2. Verify Utreexo proof
- 3. Verify signatures

How does it work?

Replace levelDB

What does levelDB do for bitcoin?

Role of levelDB

It let's you

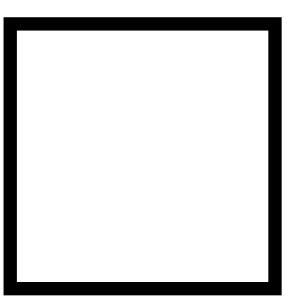
- 1. Add a UTXO
- 2. Delete a UTXO
- 3. Tell you the existence of a UTXO
- 4. Provide the data for verification

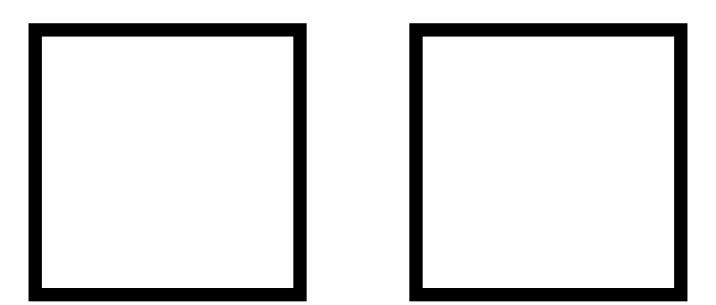
Role of levelDB

It let's you

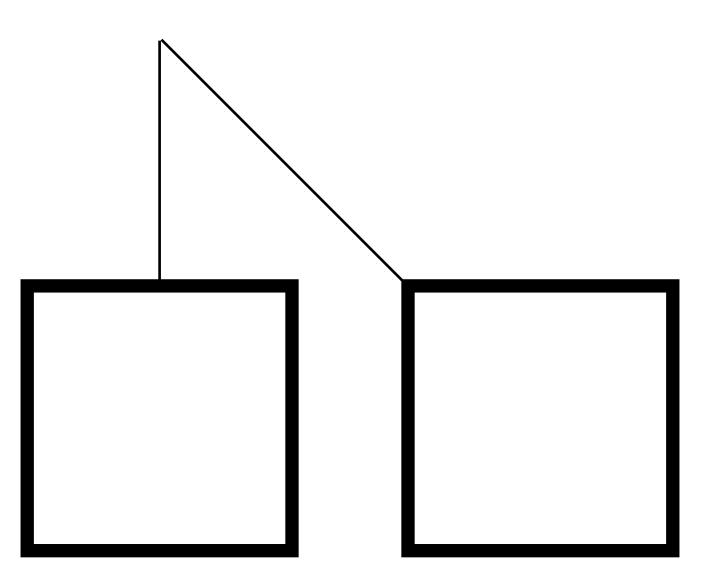
- 1. Add a UTXO
- 2. Delete a UTXO
- 3. Tell you the existence of a UTXO
- 4. Provide the data for verification

1 UTXO Single root



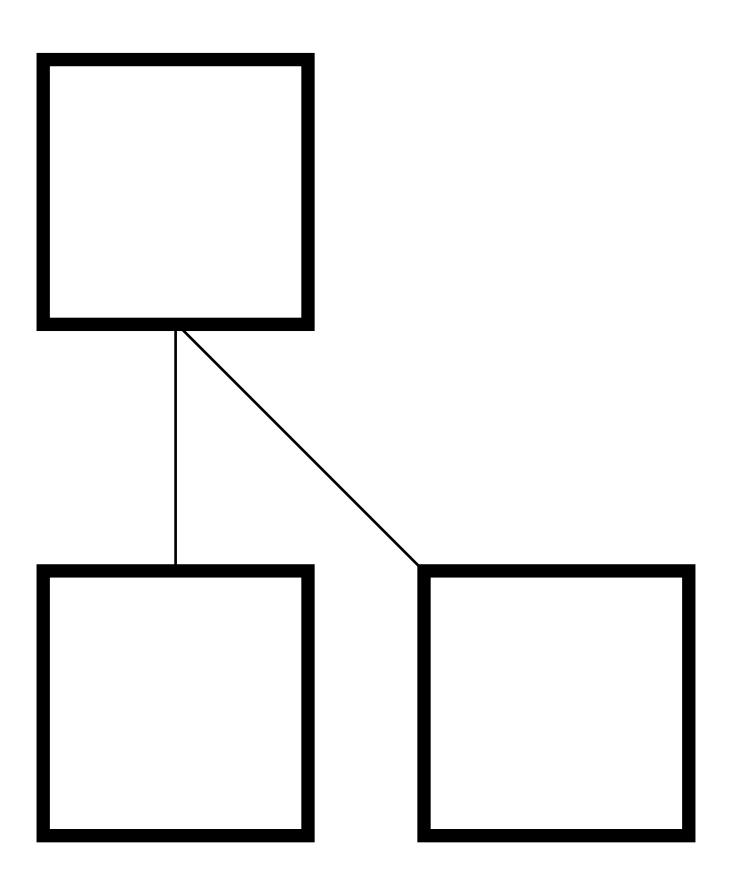


2 UTXOs Concatenate



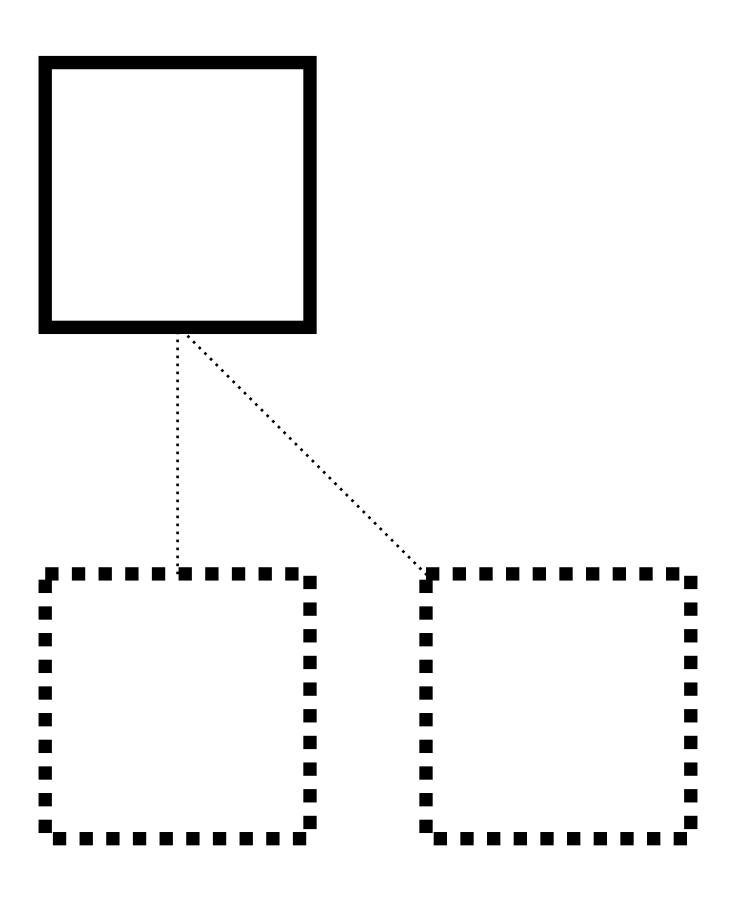
2 UTXOs

Hash to create a new root

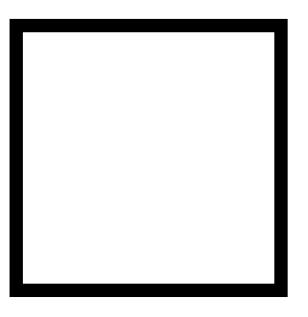


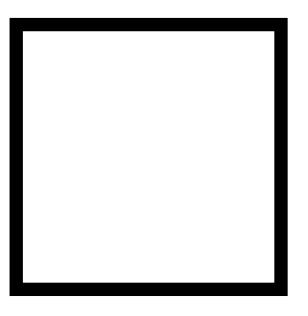
2 UTXOs

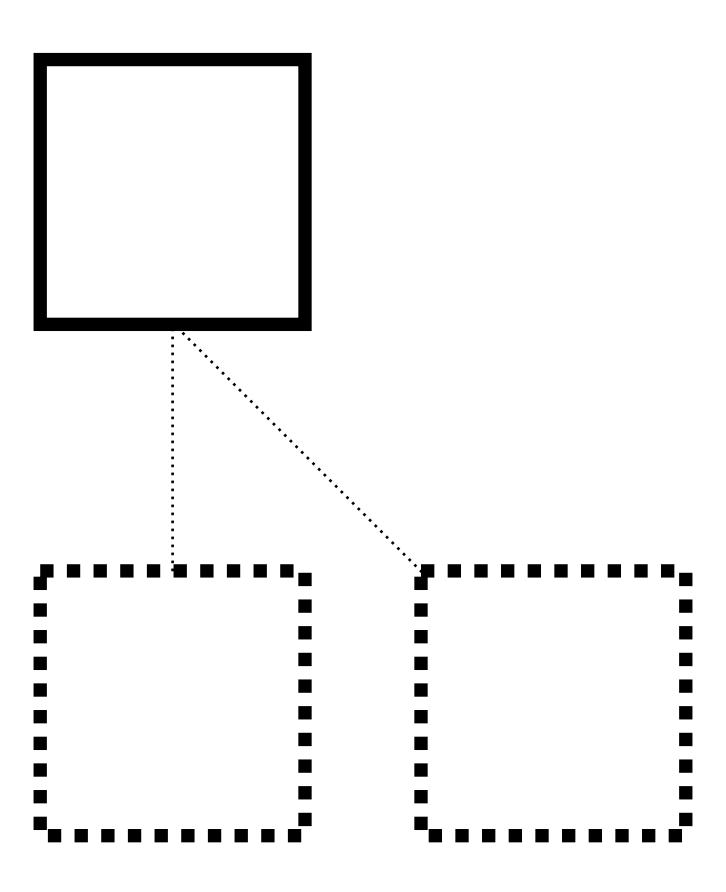
Can now delete the leaves

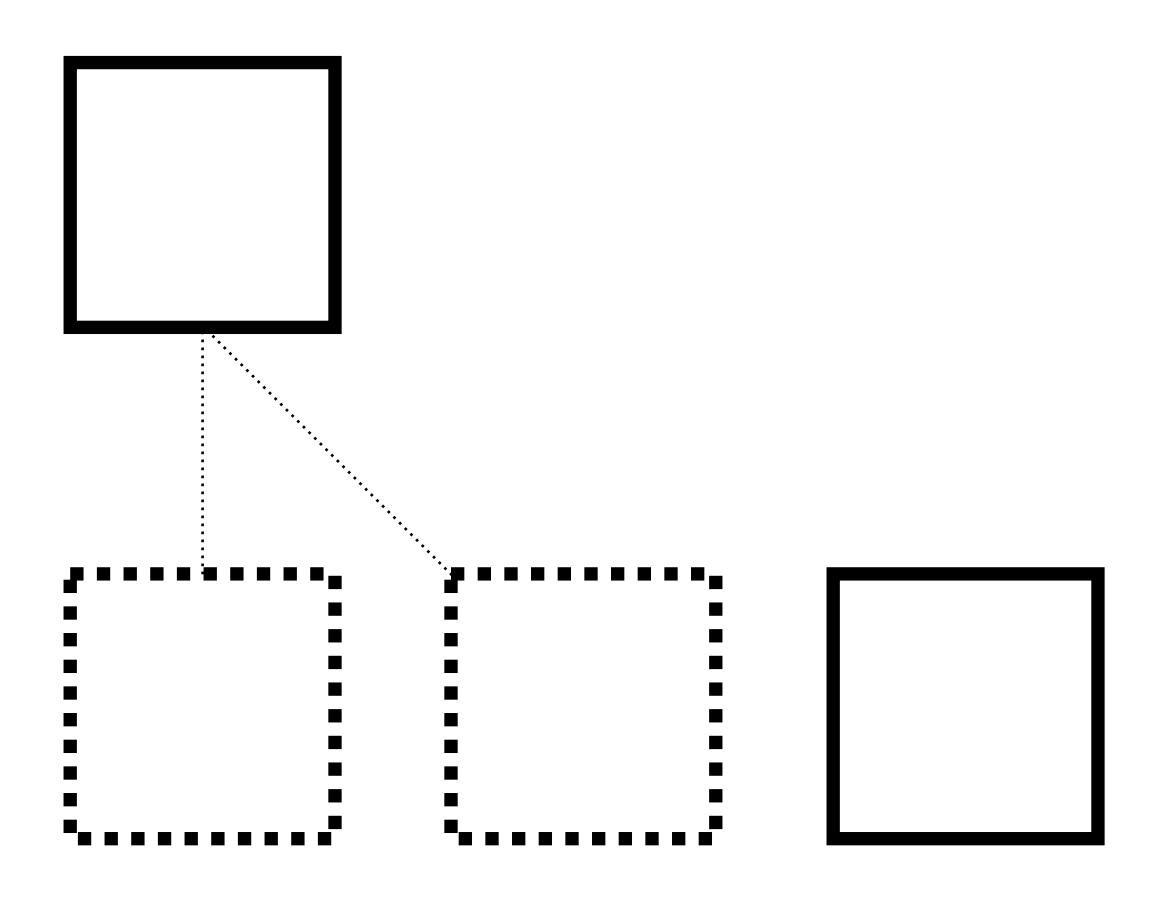


2 UTXOs Only keep the roots



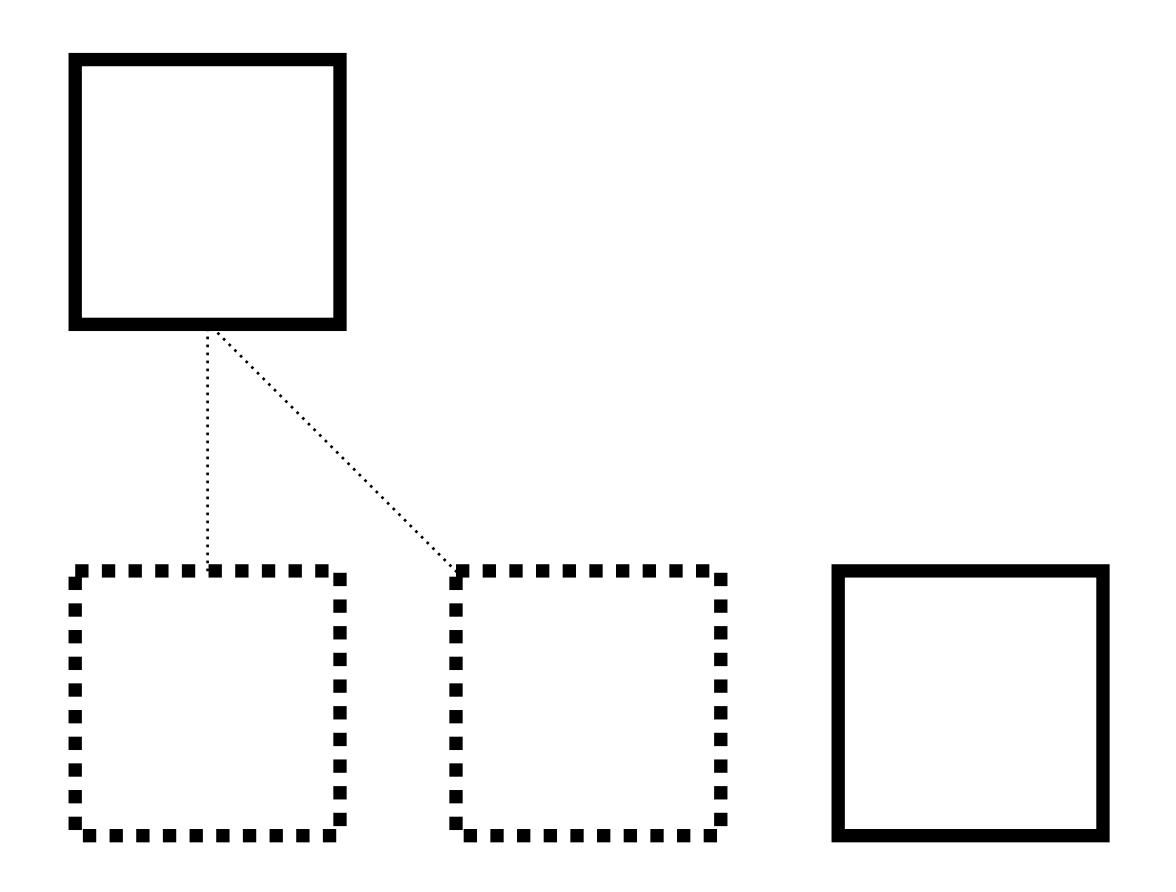




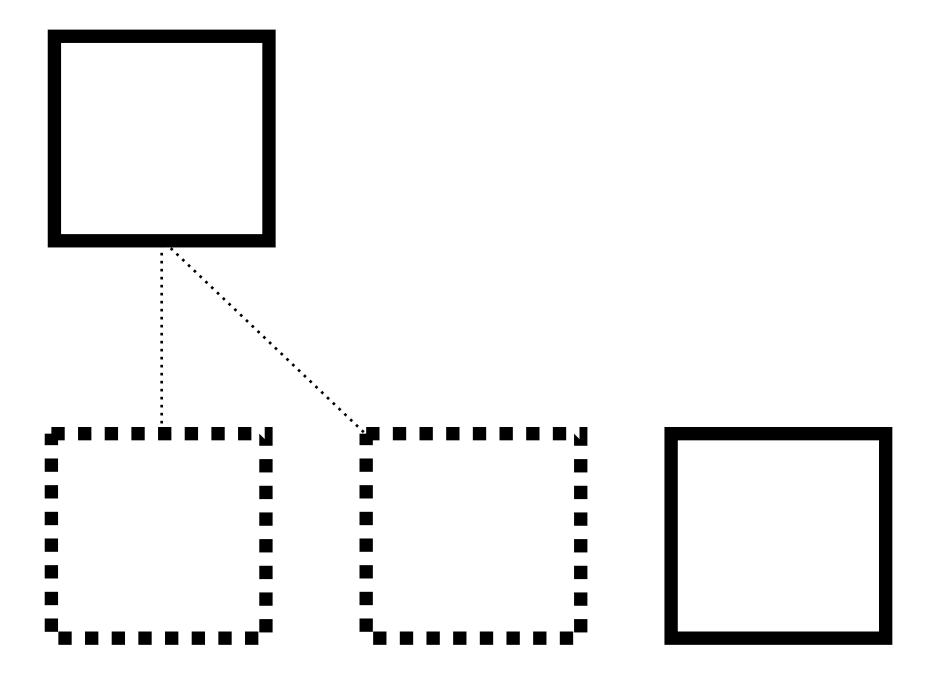


3 UTXOs

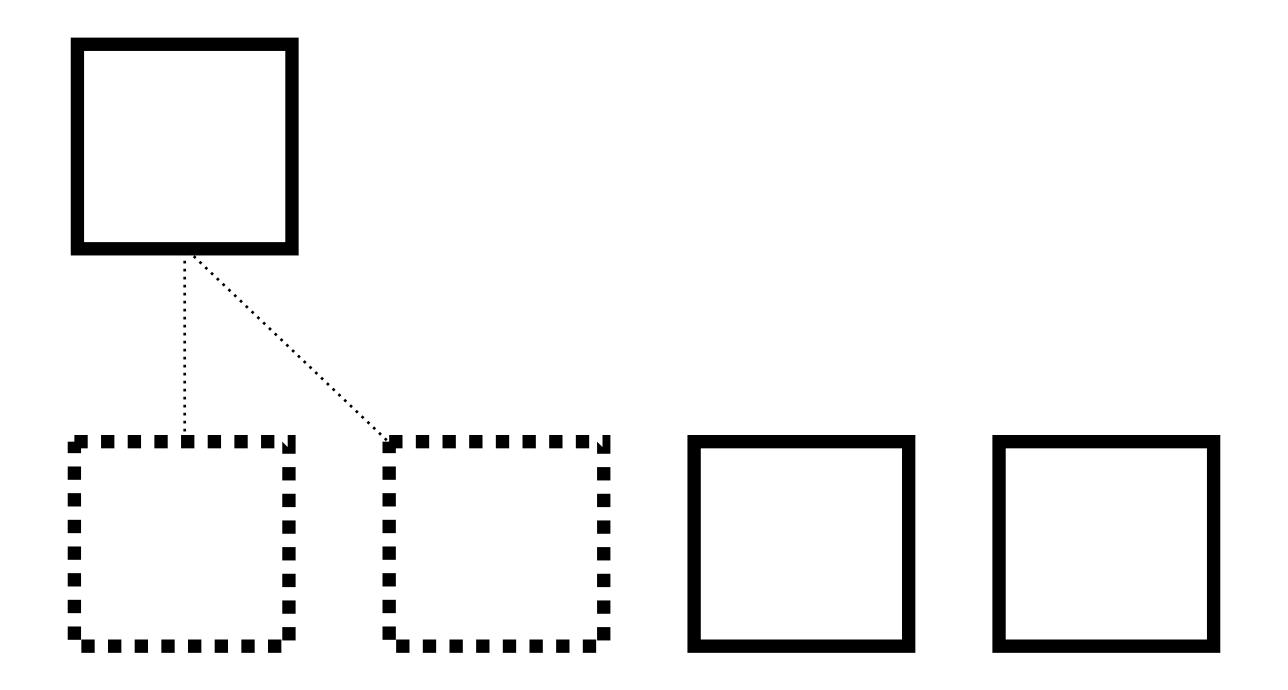
Nothing to hash with. Becomes a root



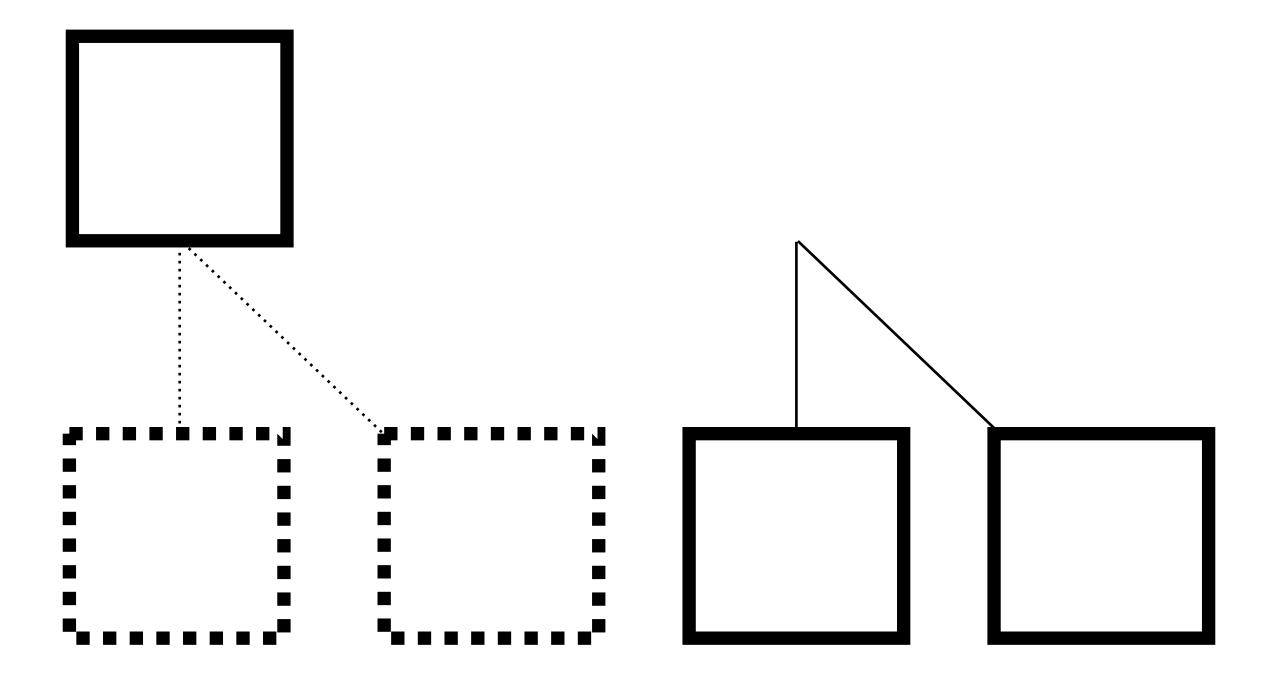
4 UTXOs Another one



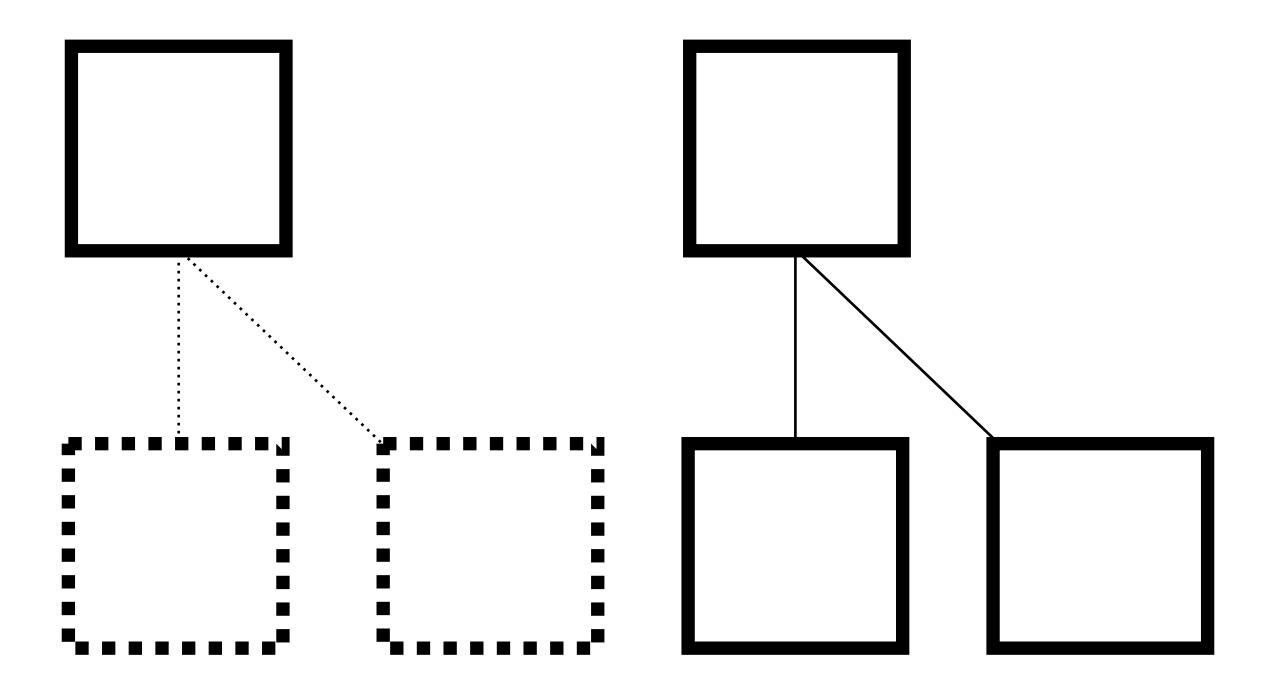
4 UTXOs Another one



4 UTXOs Concatenate

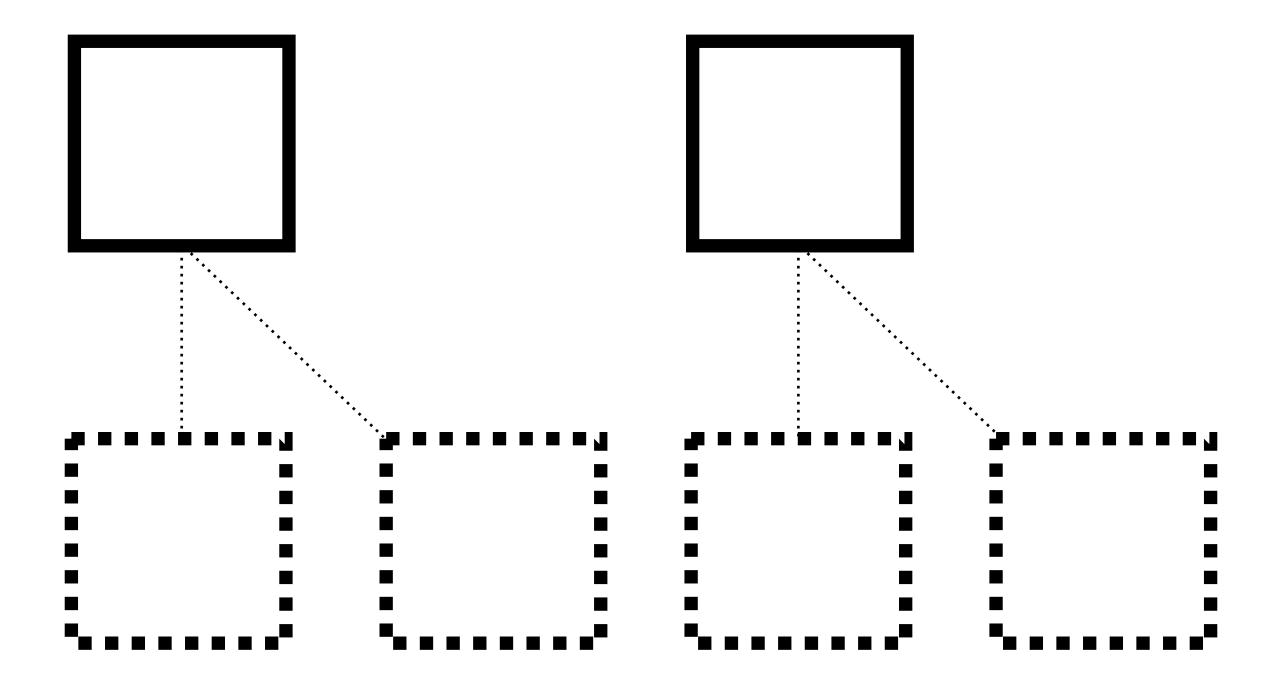


4 UTXOs Hash

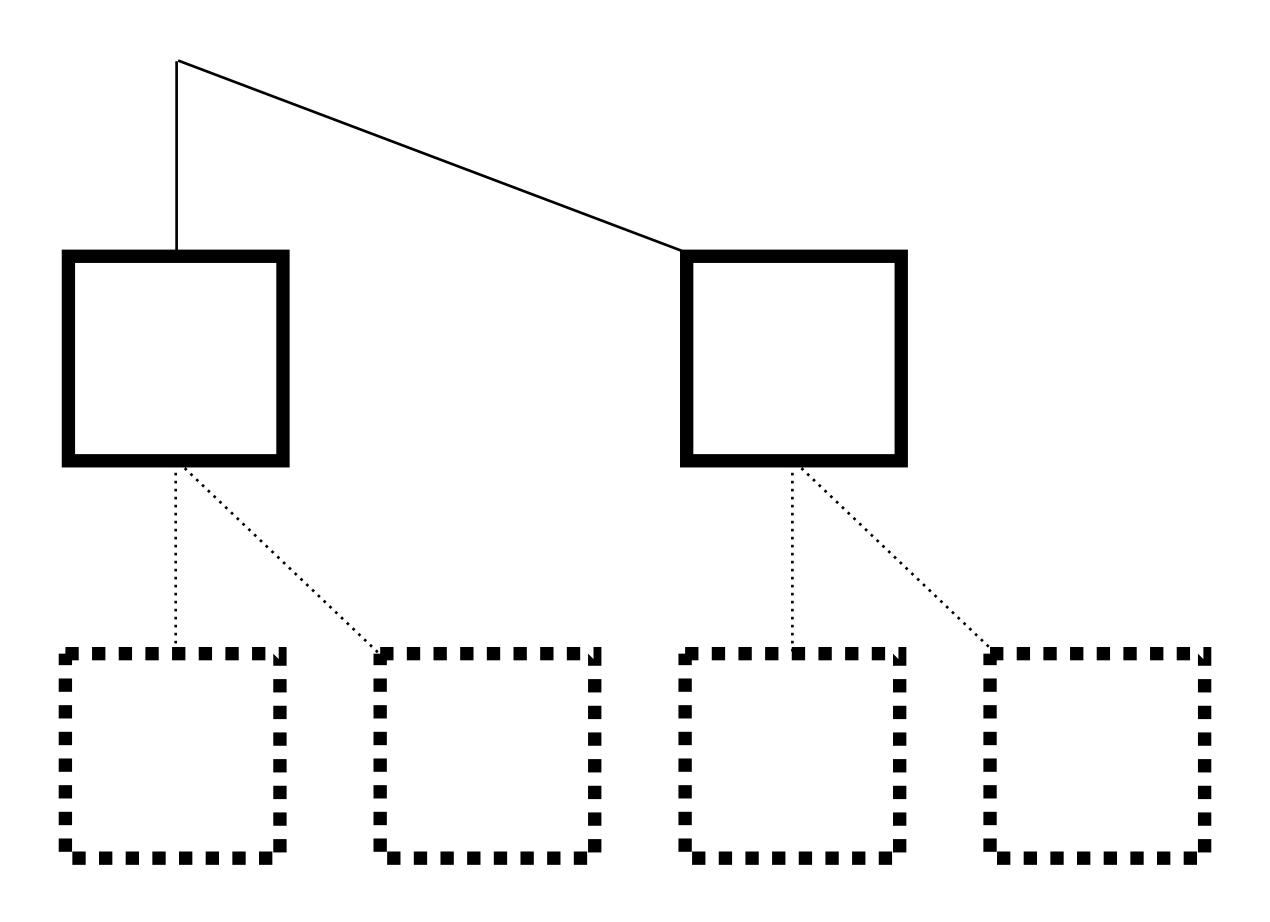


4 UTXOs

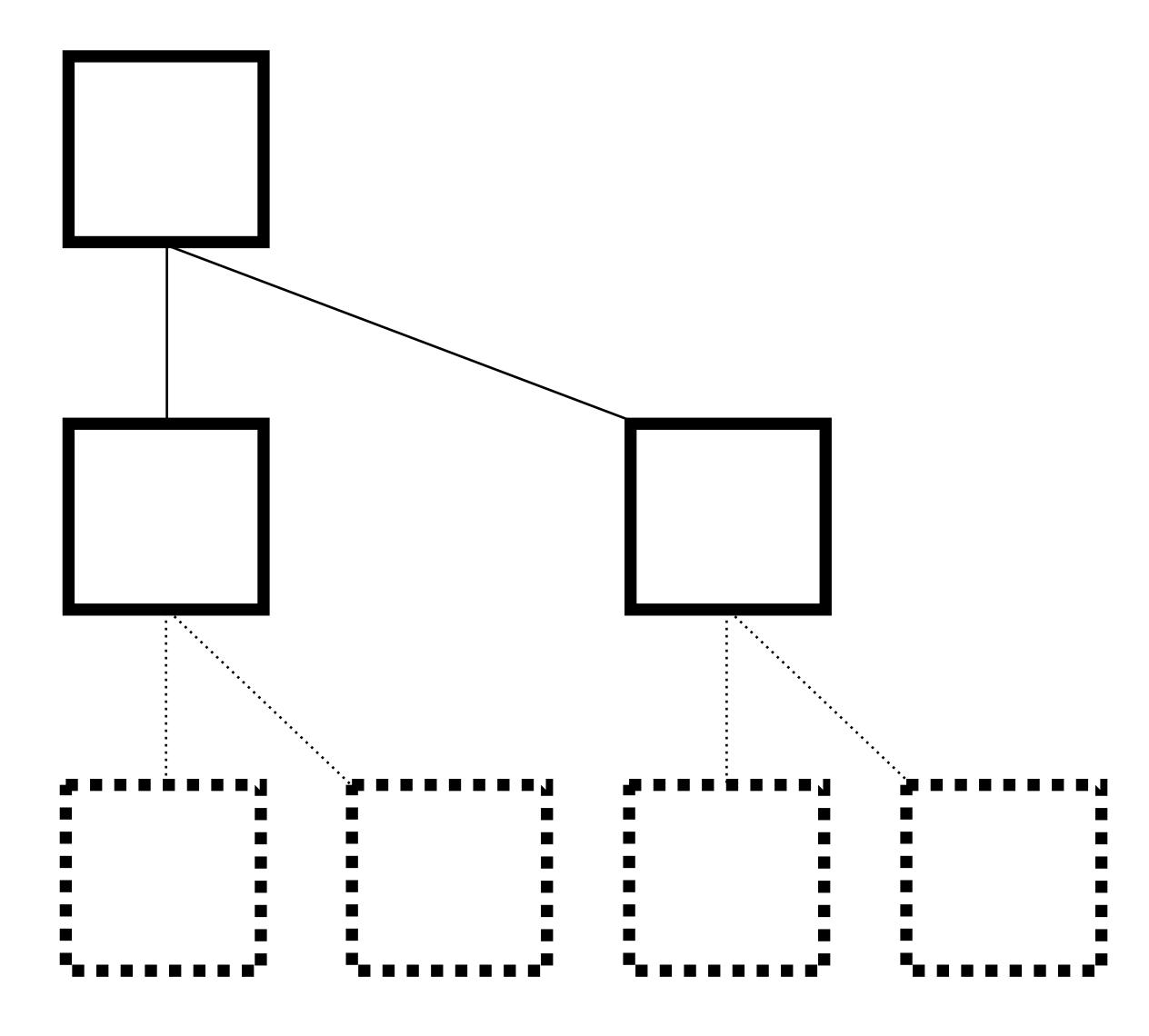
Can now throw away the leaves



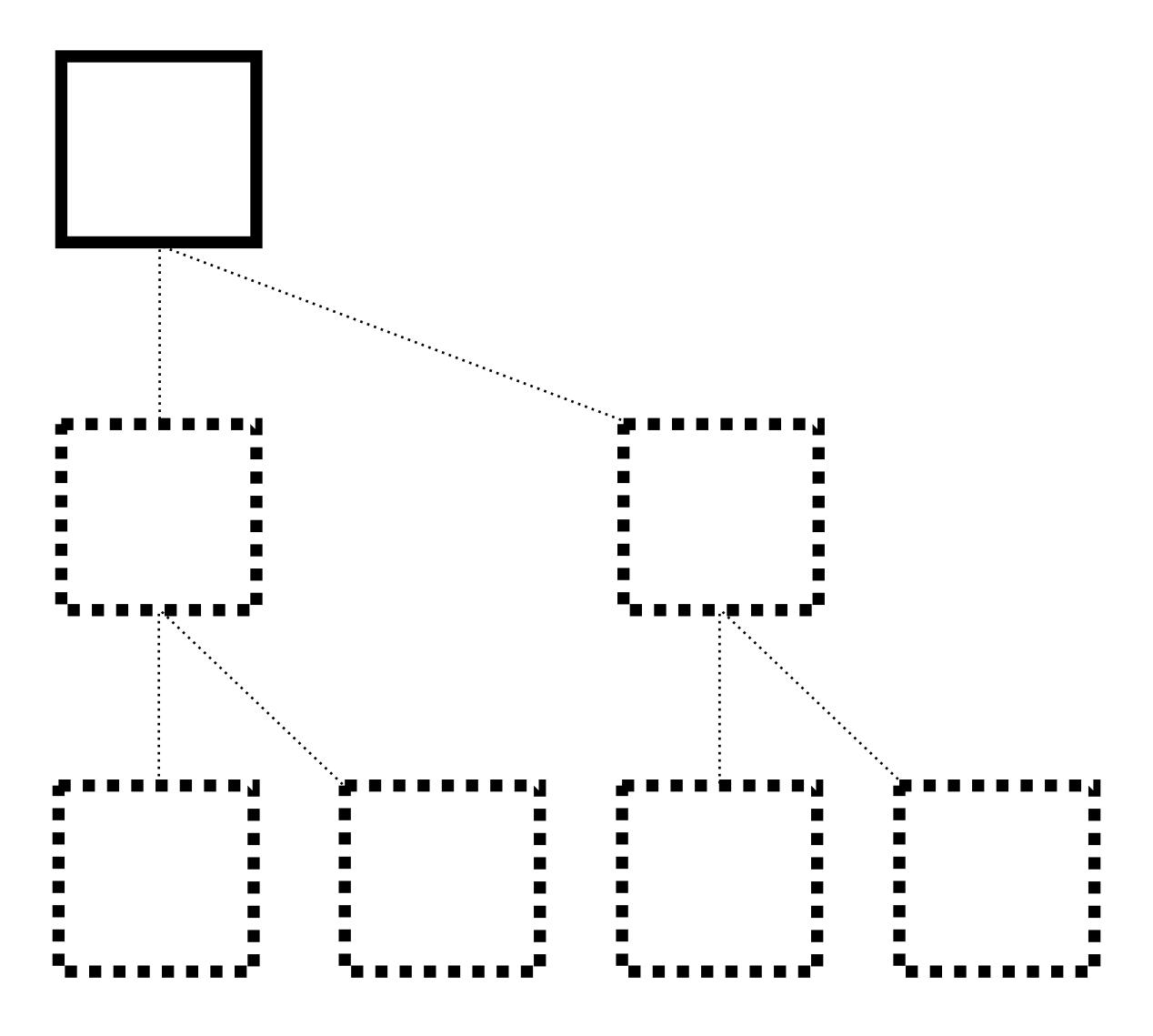
4 UTXOs Concatenate



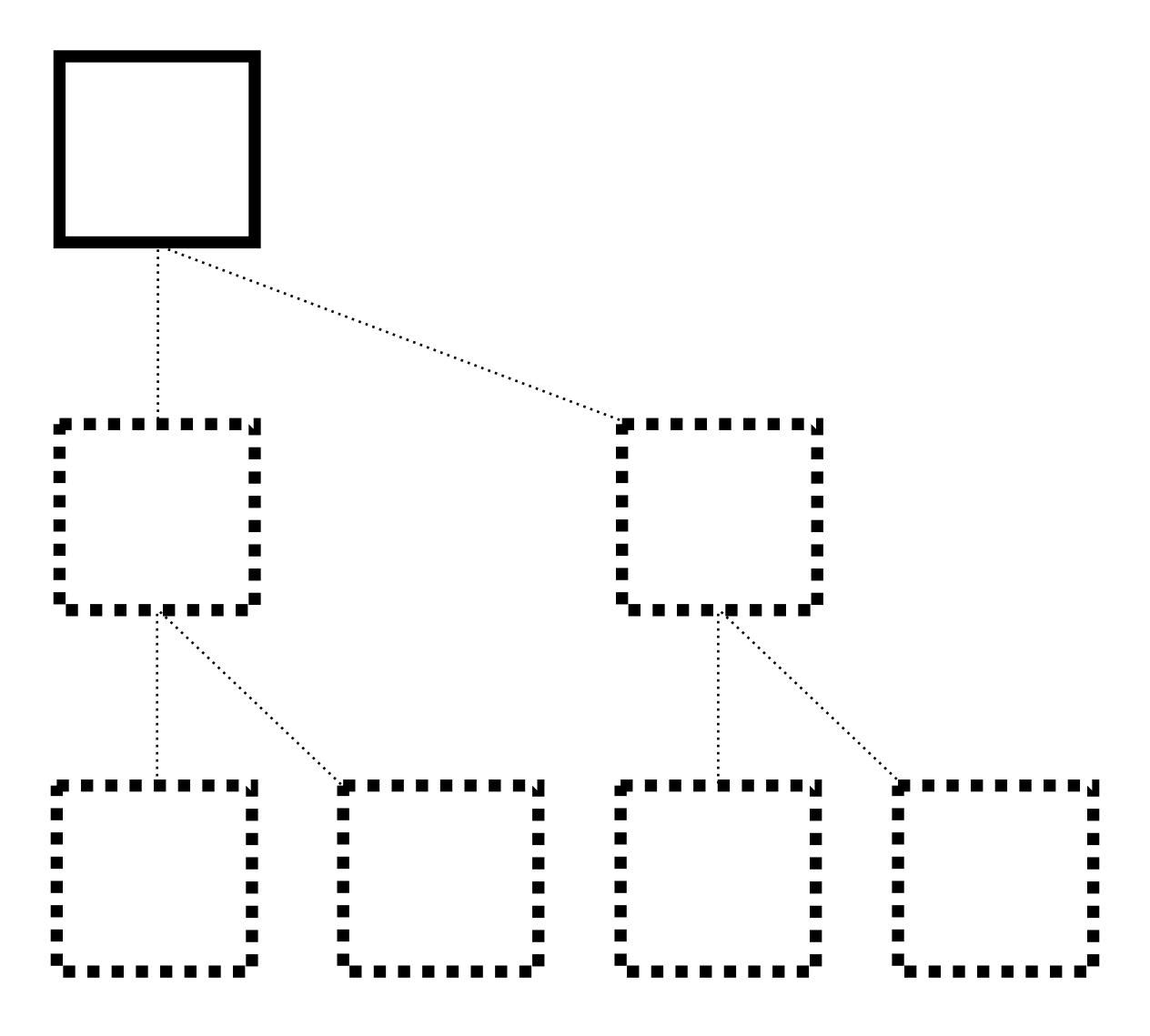
4 UTXOs Hash



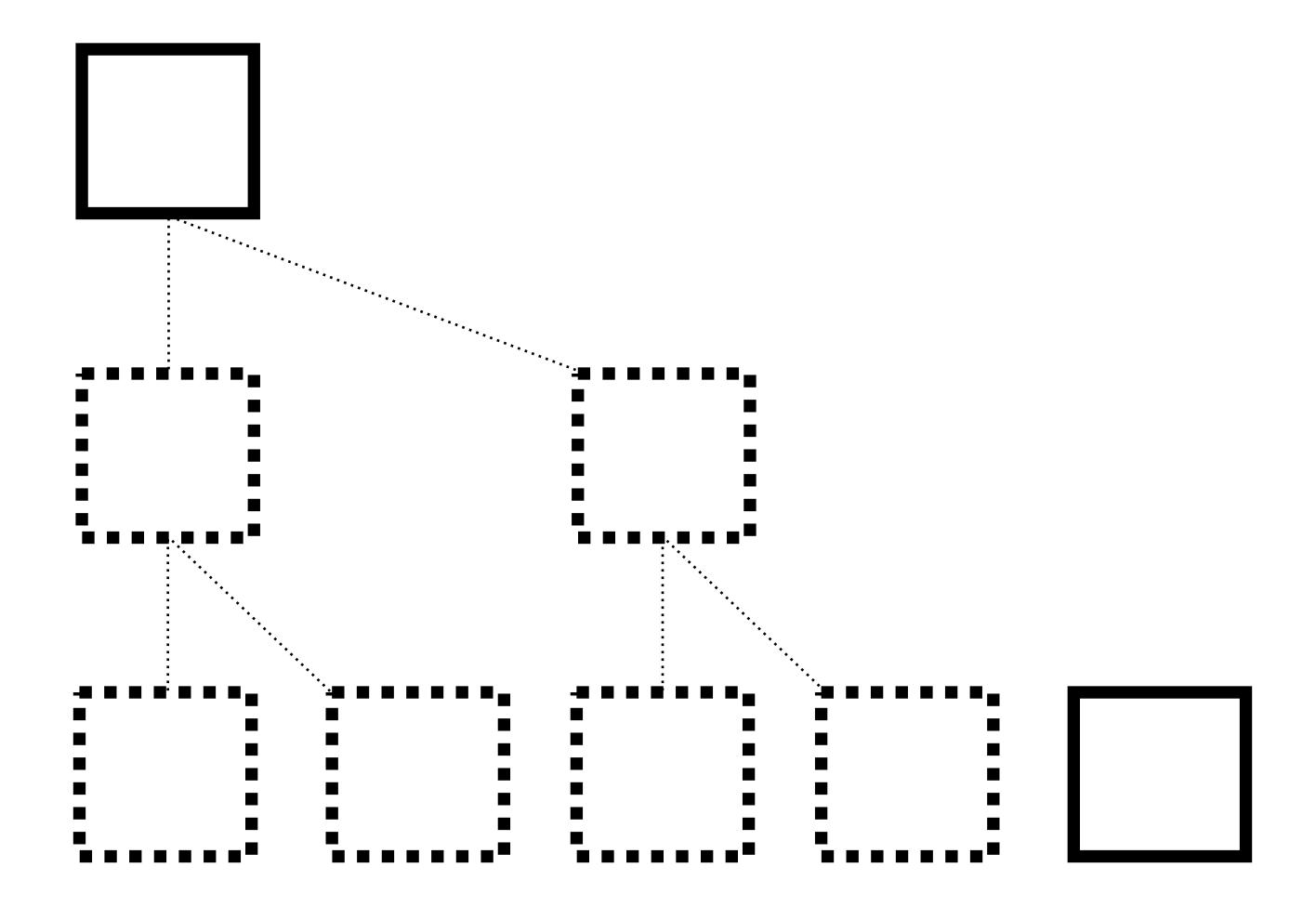
4 UTXOs Throw away



4 UTXOs Finished tree

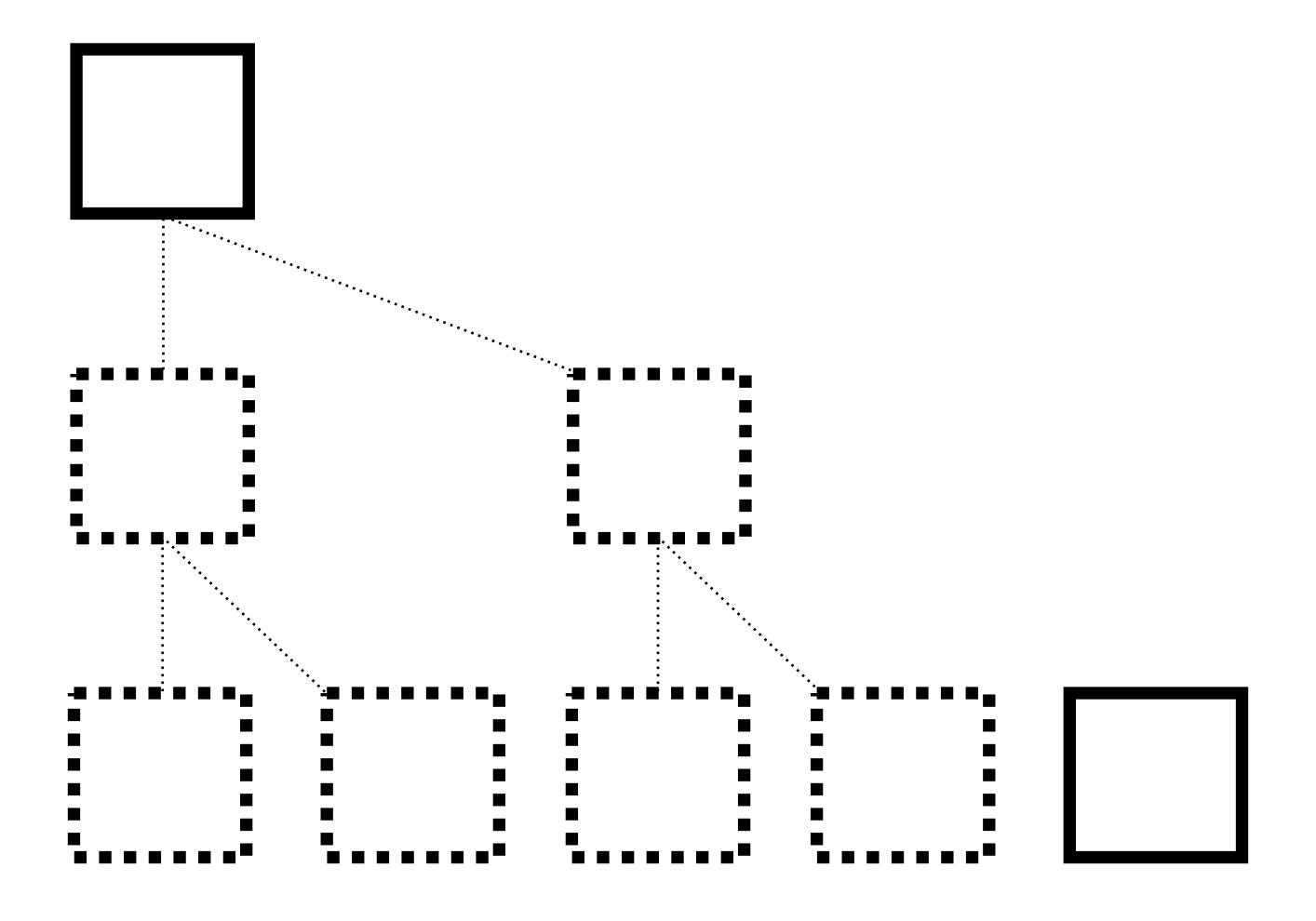


5 UTXOs Add one more



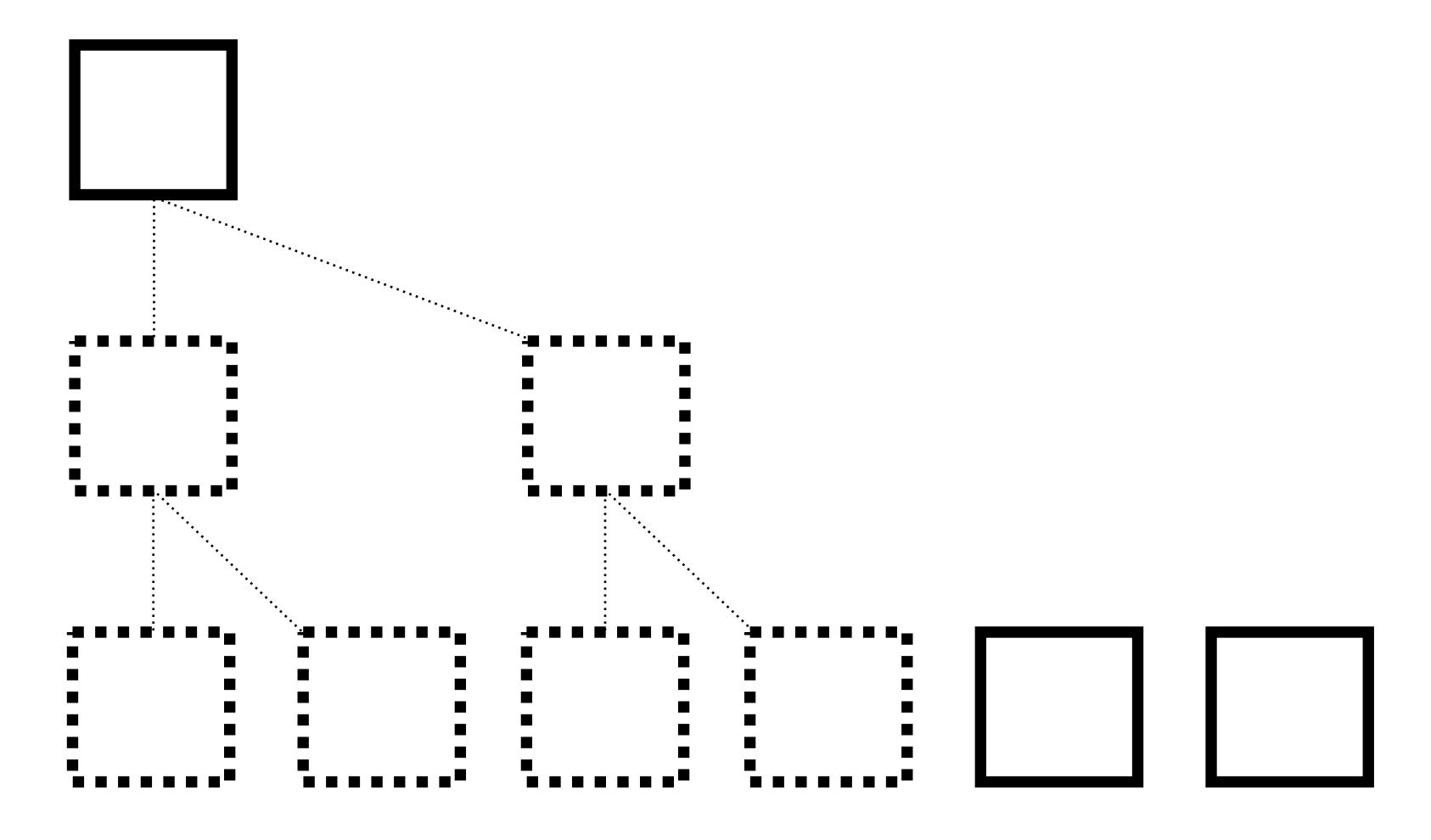
5 UTXOs

Nothing to hash with so it becomes a root

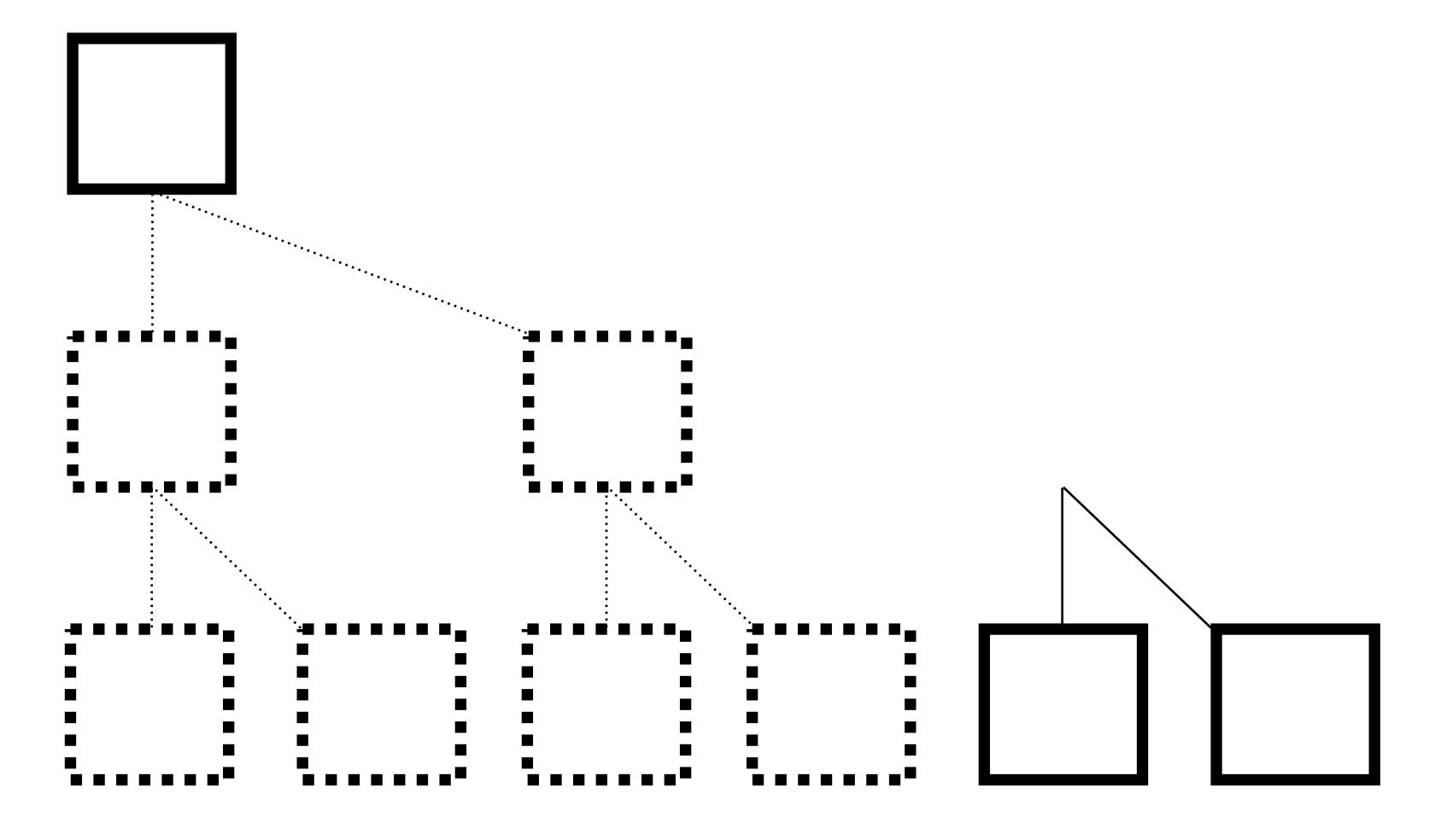


6 UTXOs

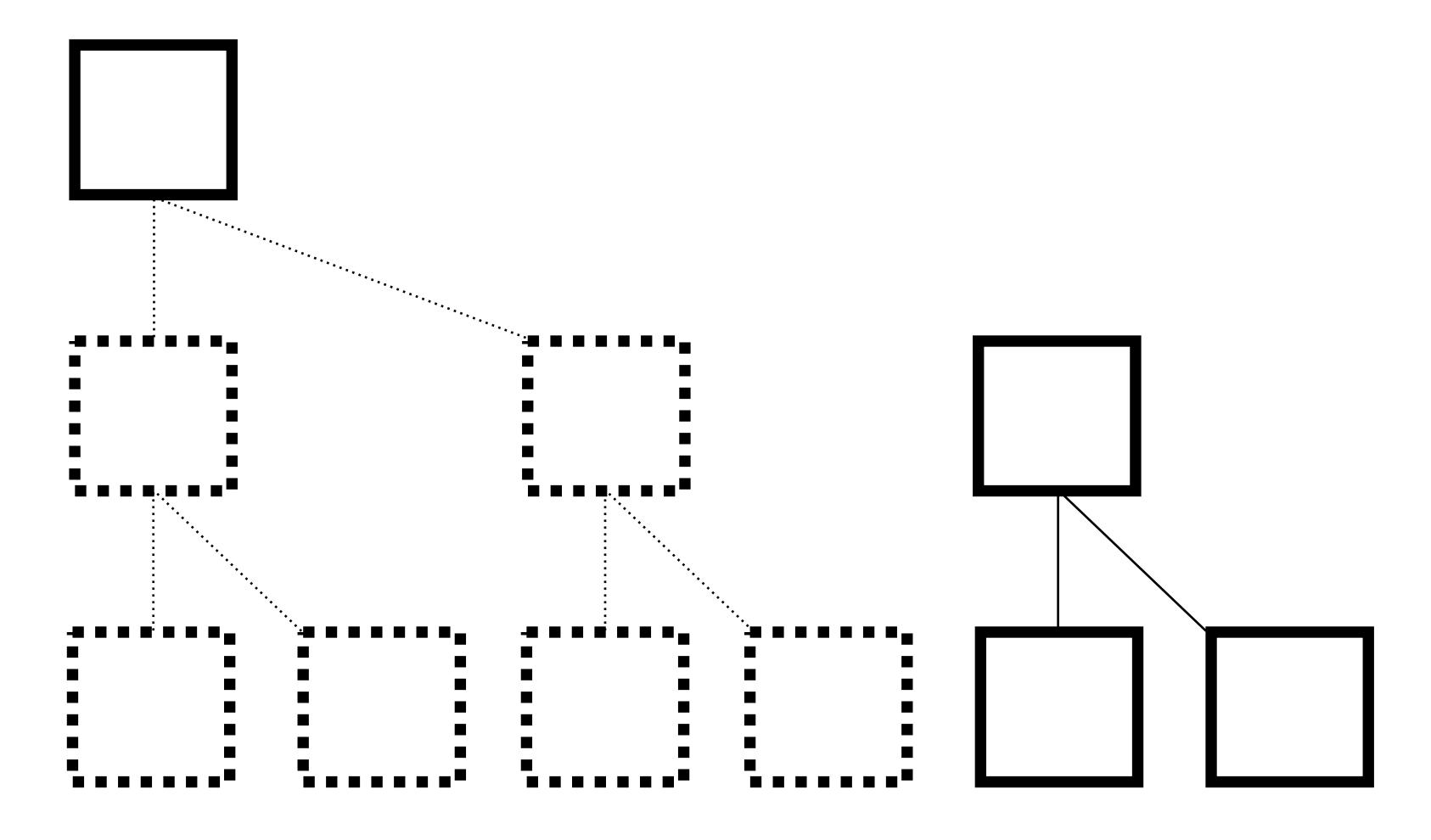
1 more



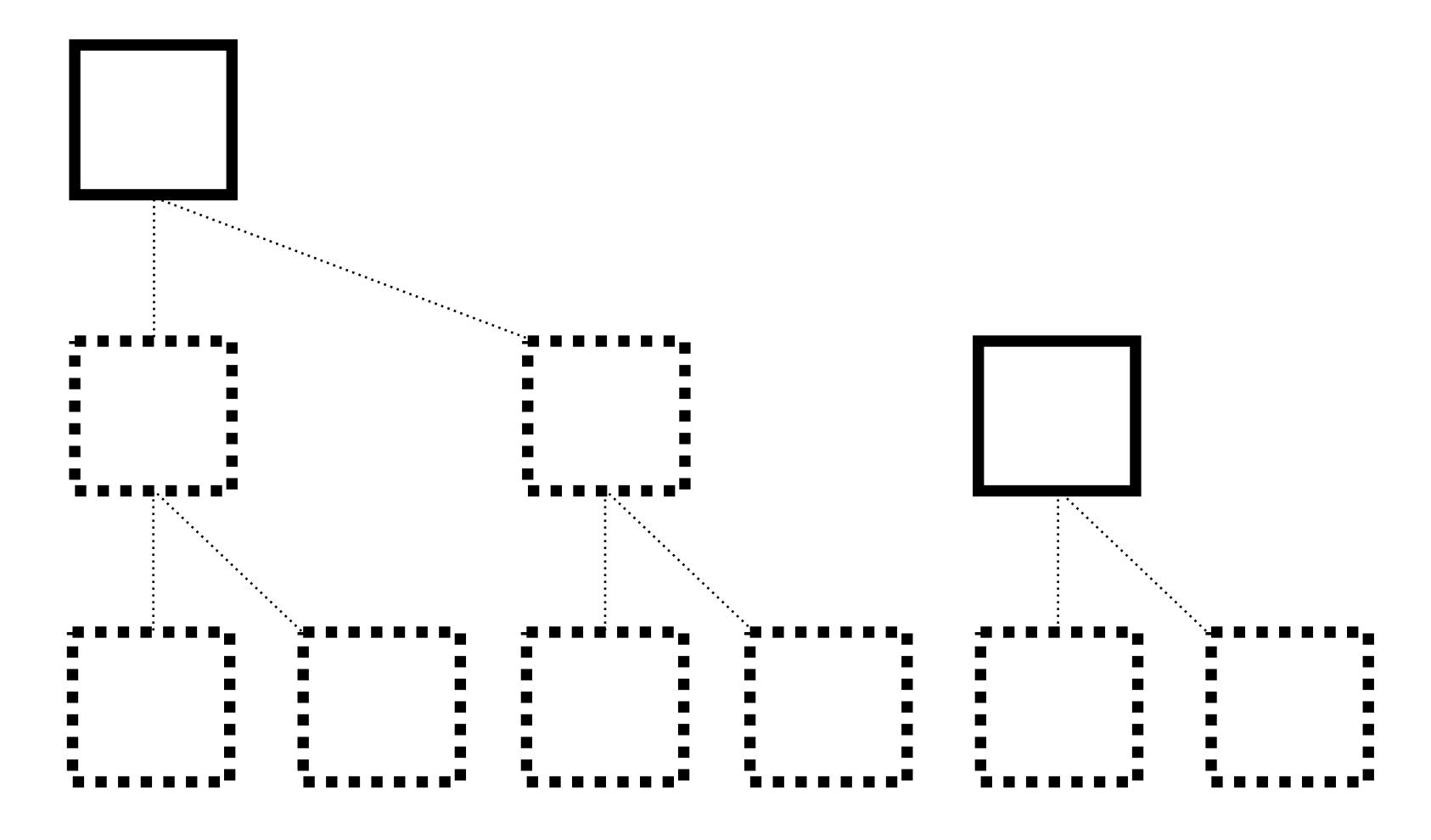
6 UTXOs Concatenate



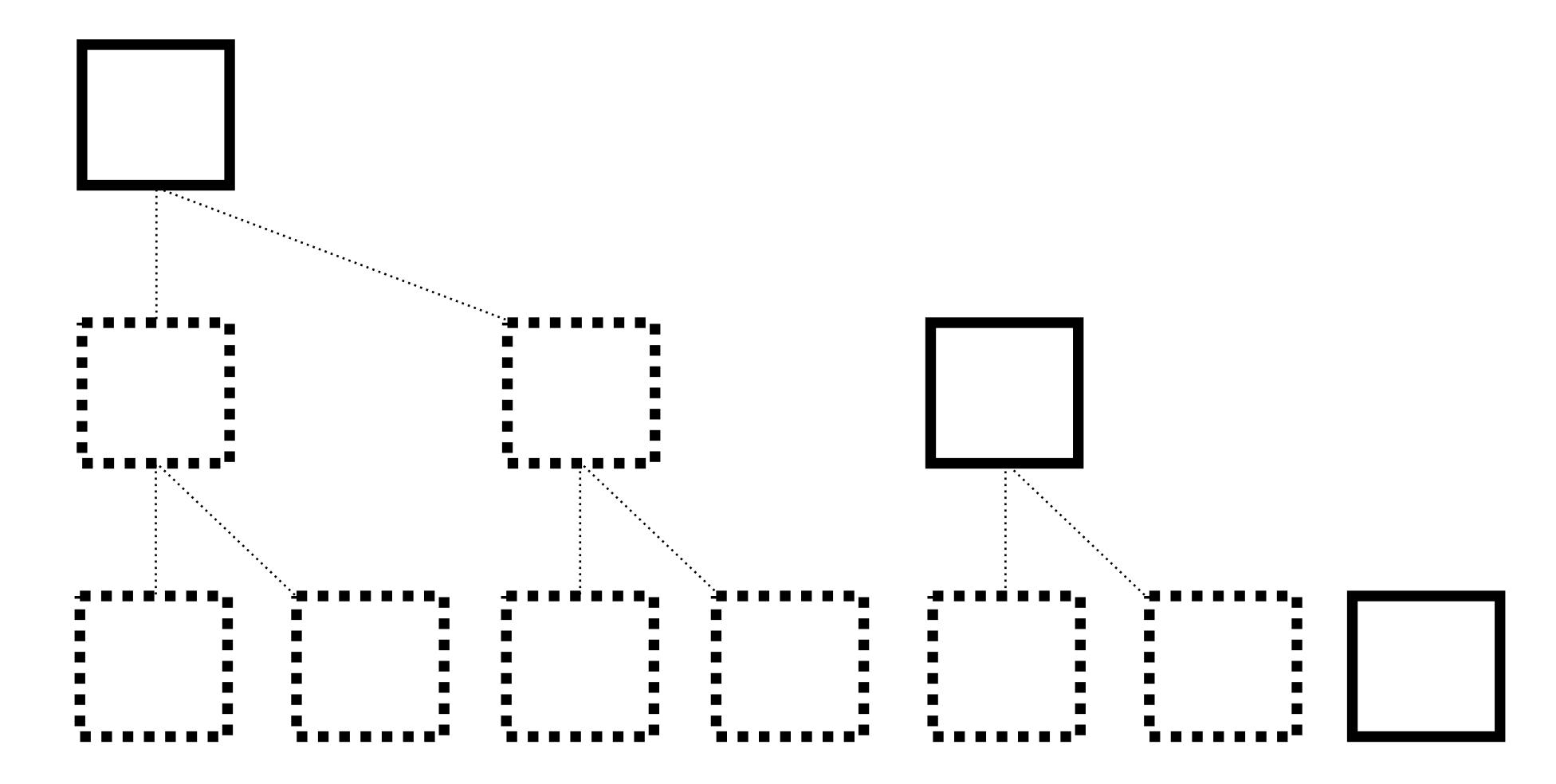
Hash



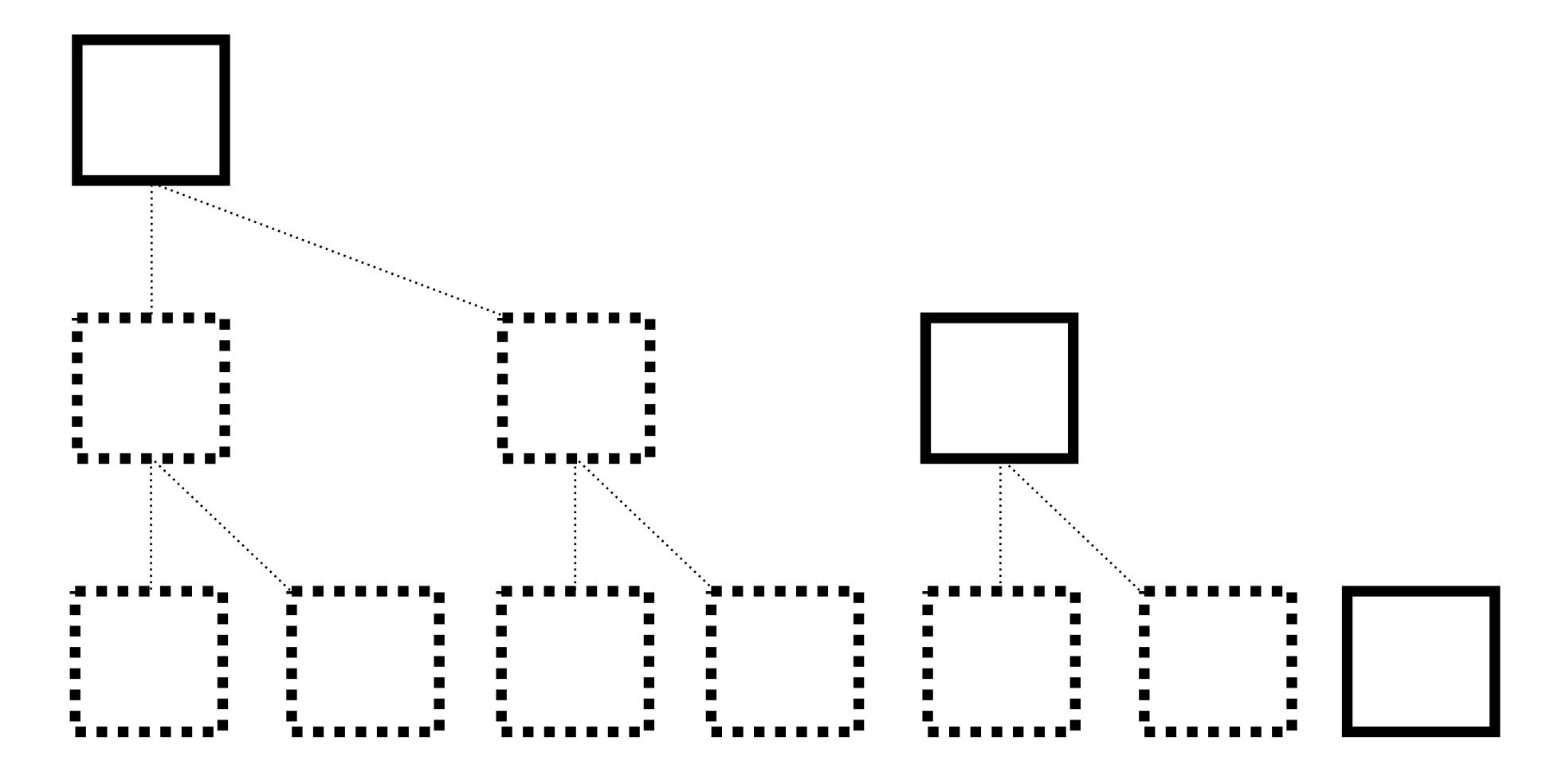
Throw away



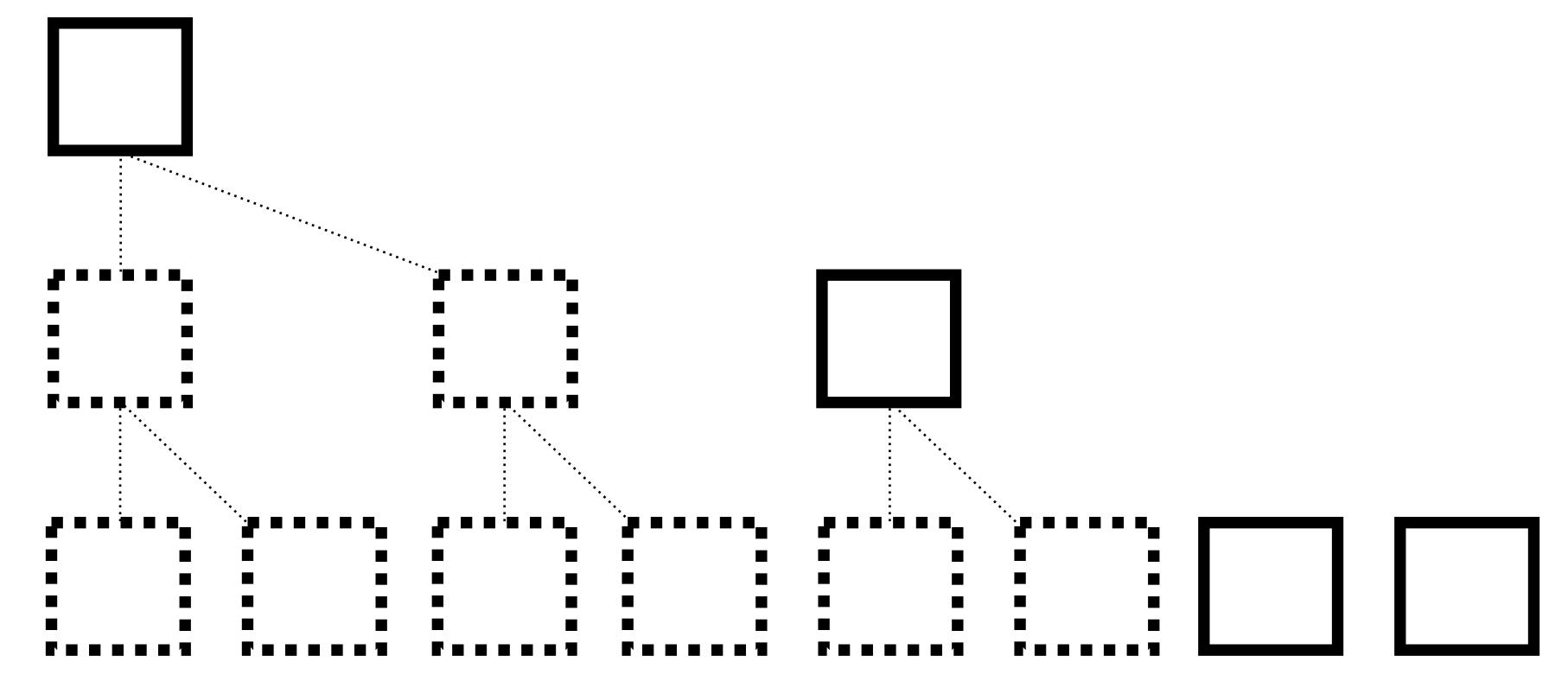
Another one



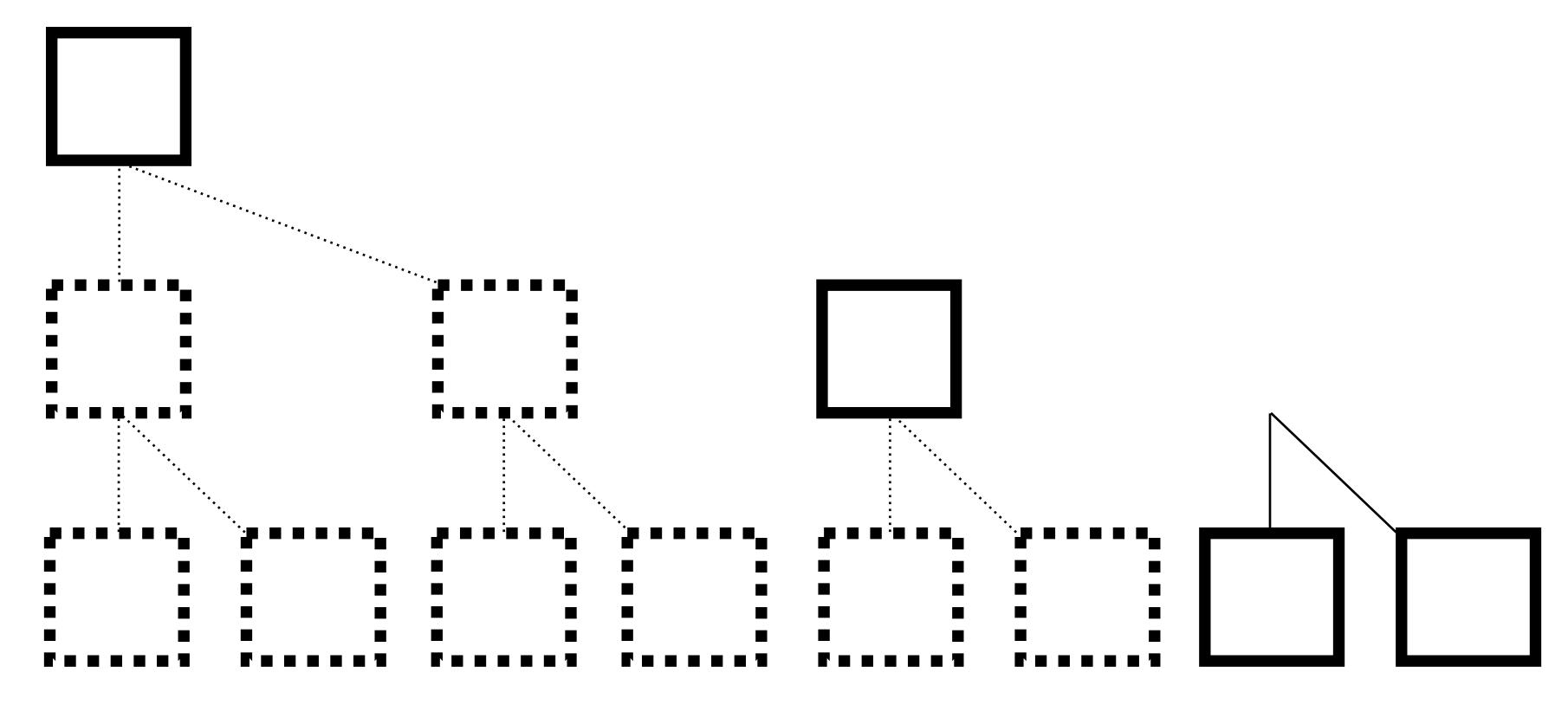
Nothing to hash with so it becomes a root



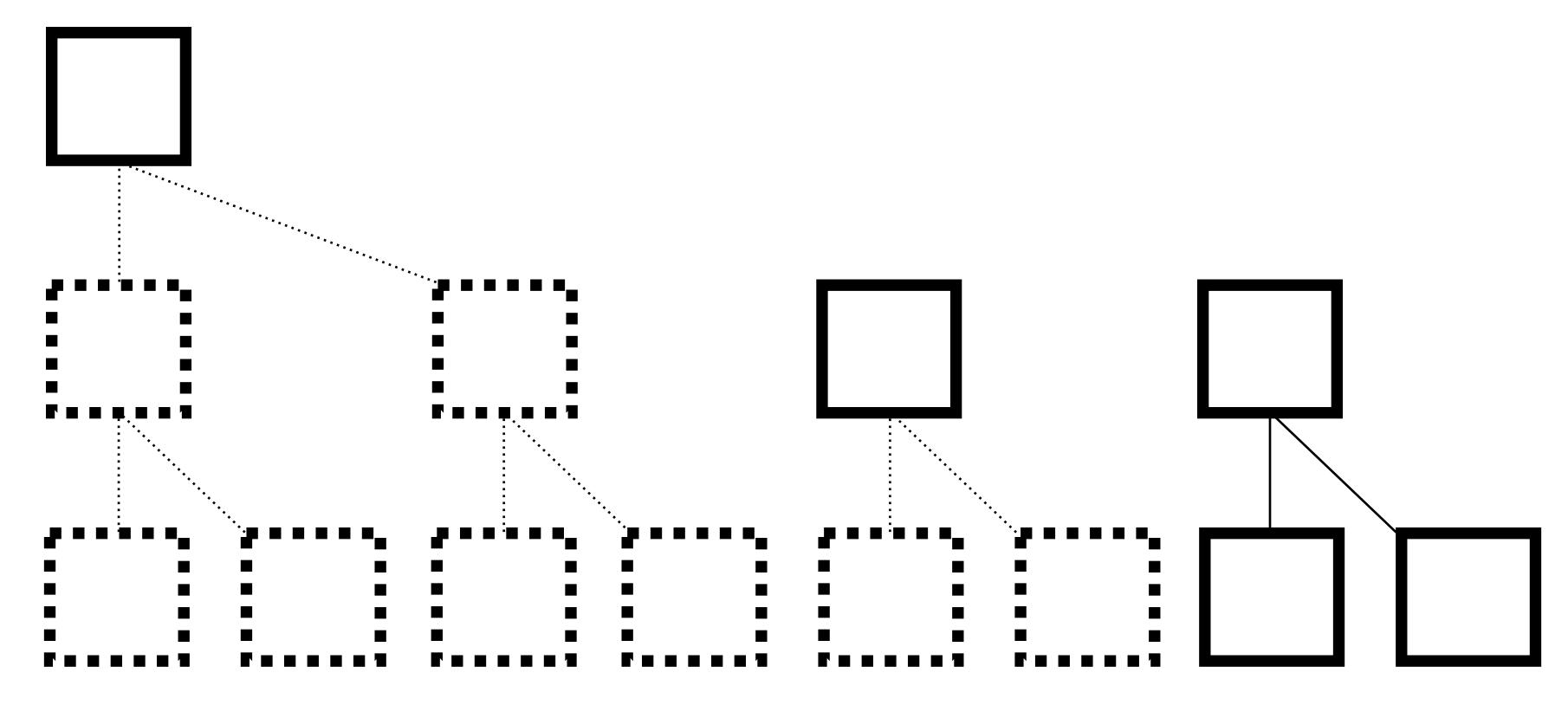
Another one



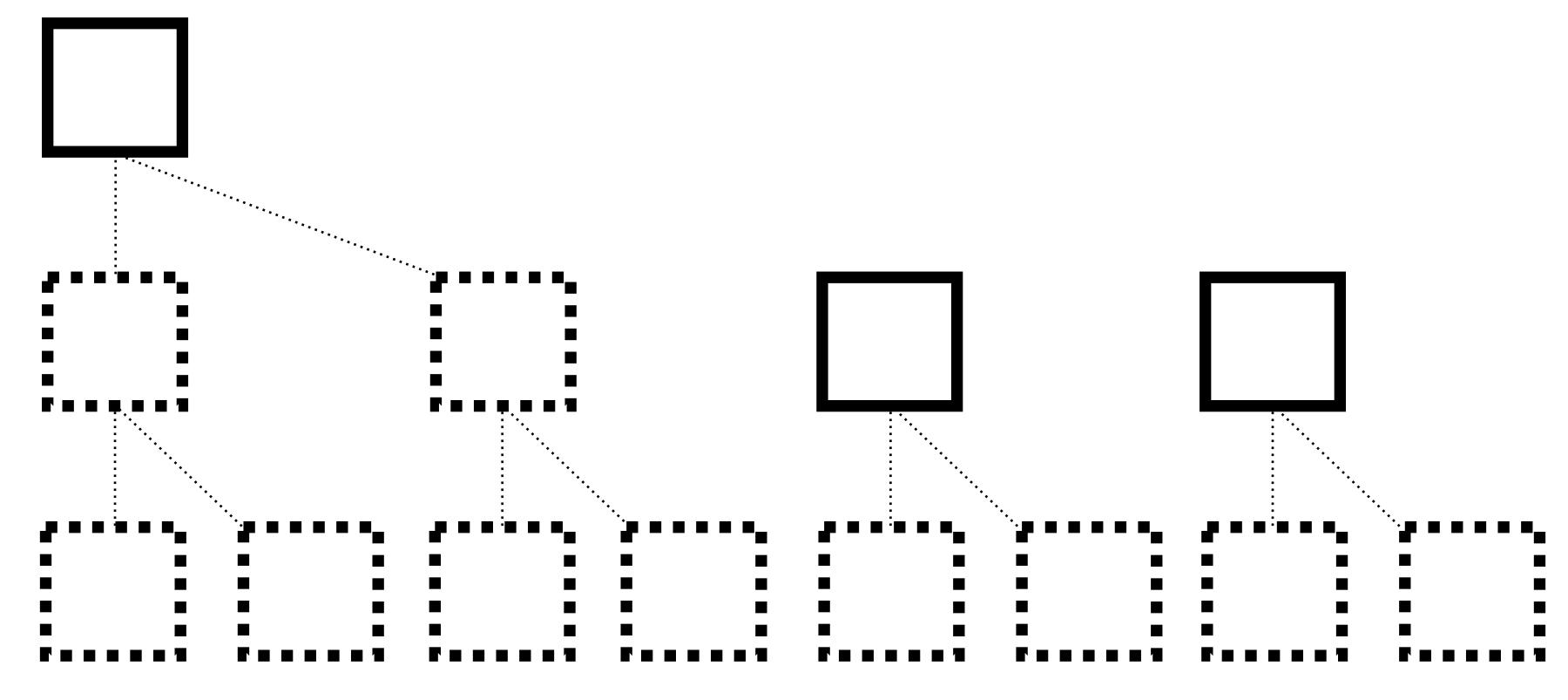
8 UTXOs Concatenate



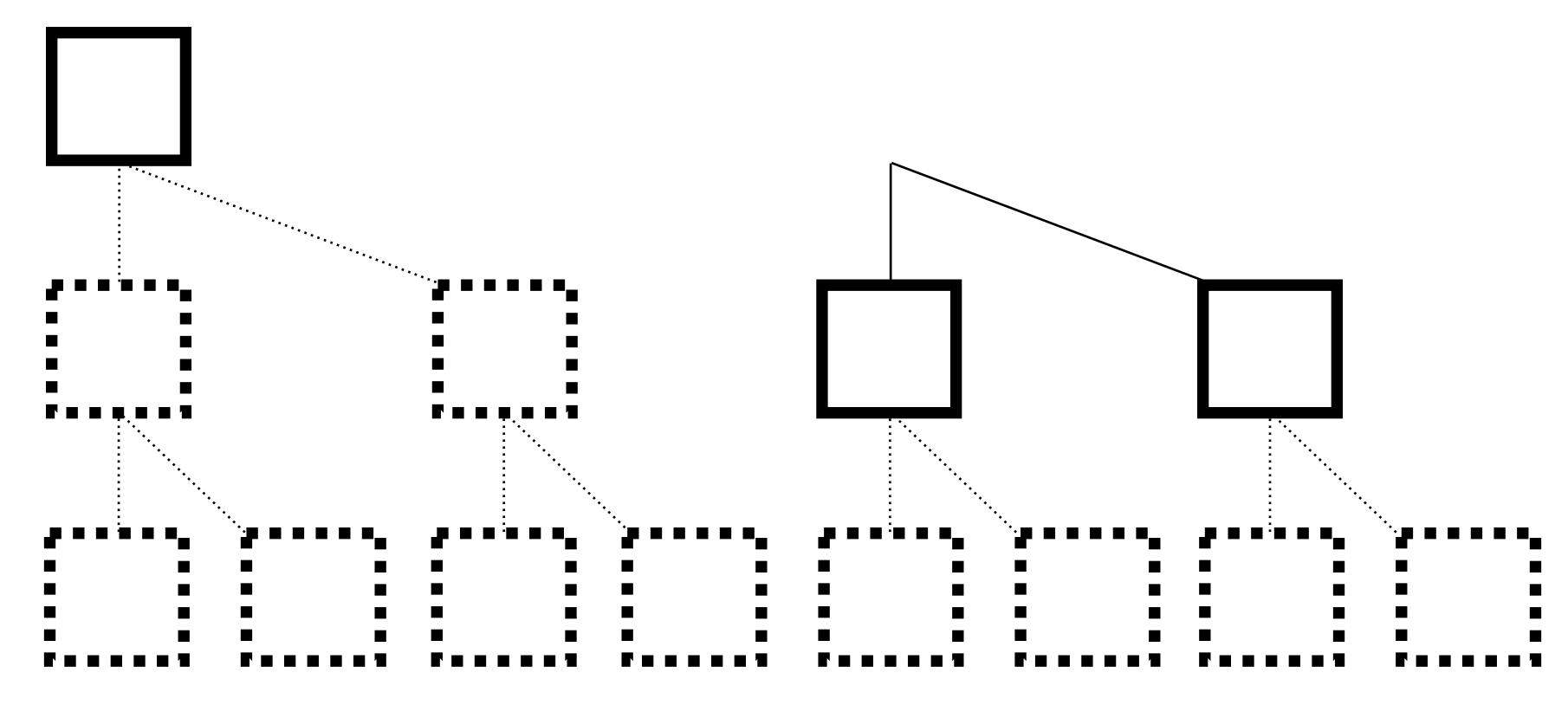
8 UTXOs Hash



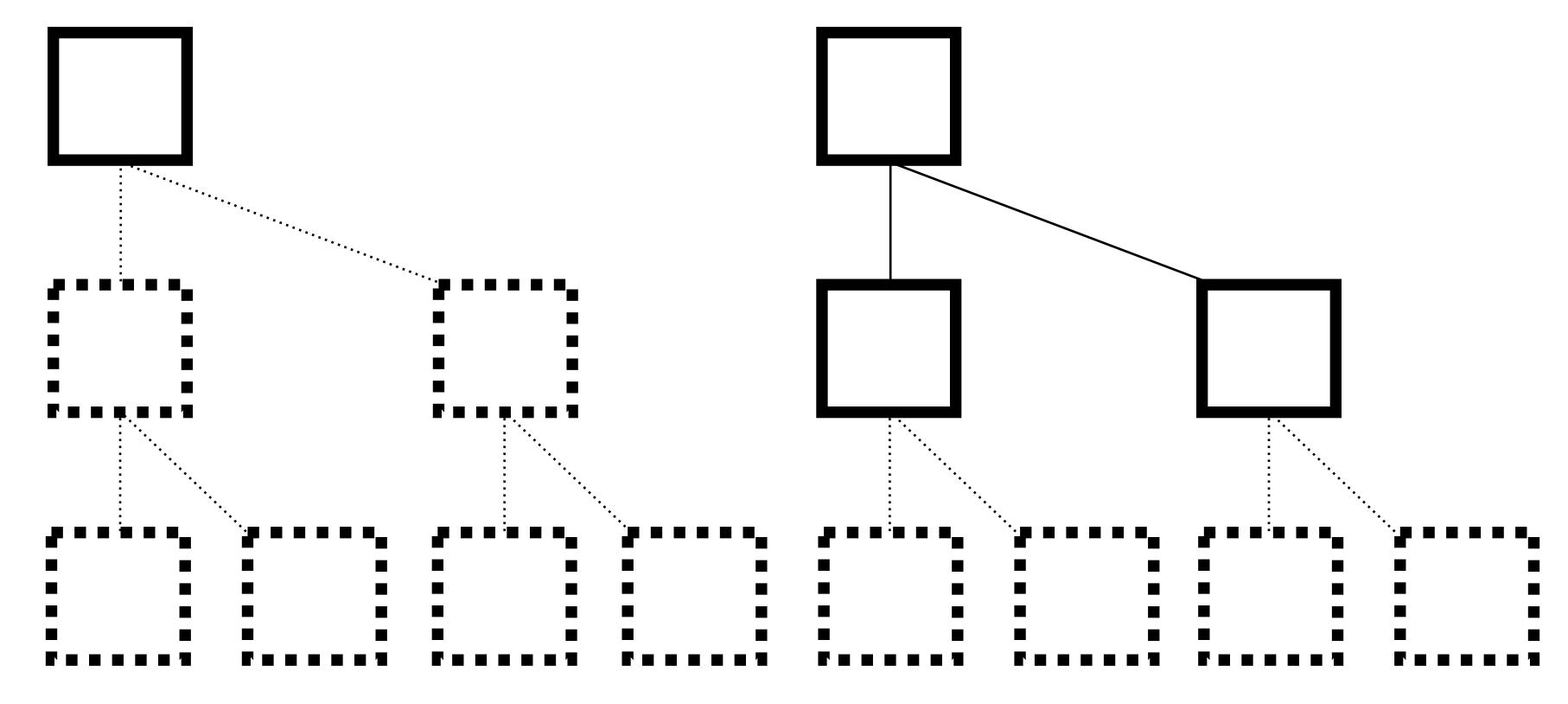
8 UTXOs Throw away



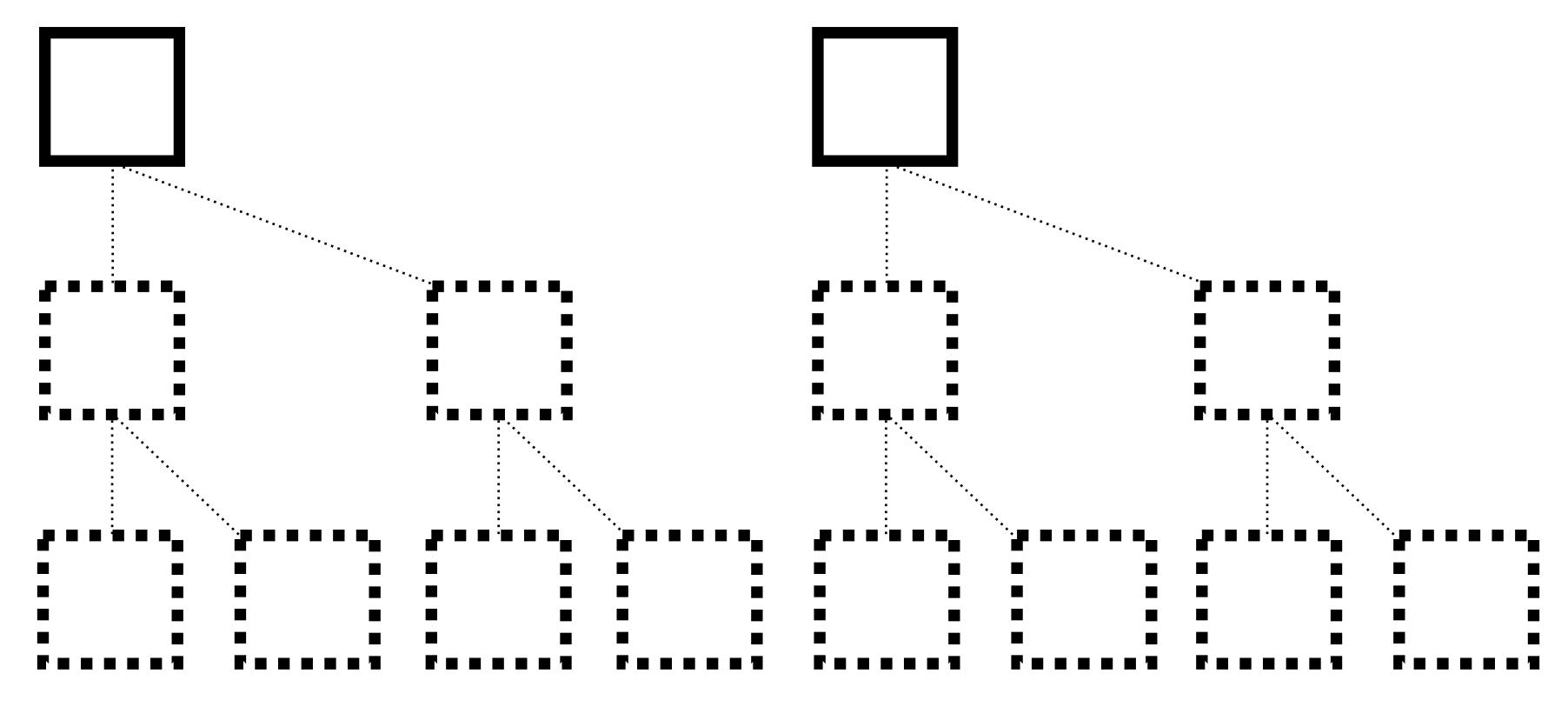
8 UTXOs Concatenate



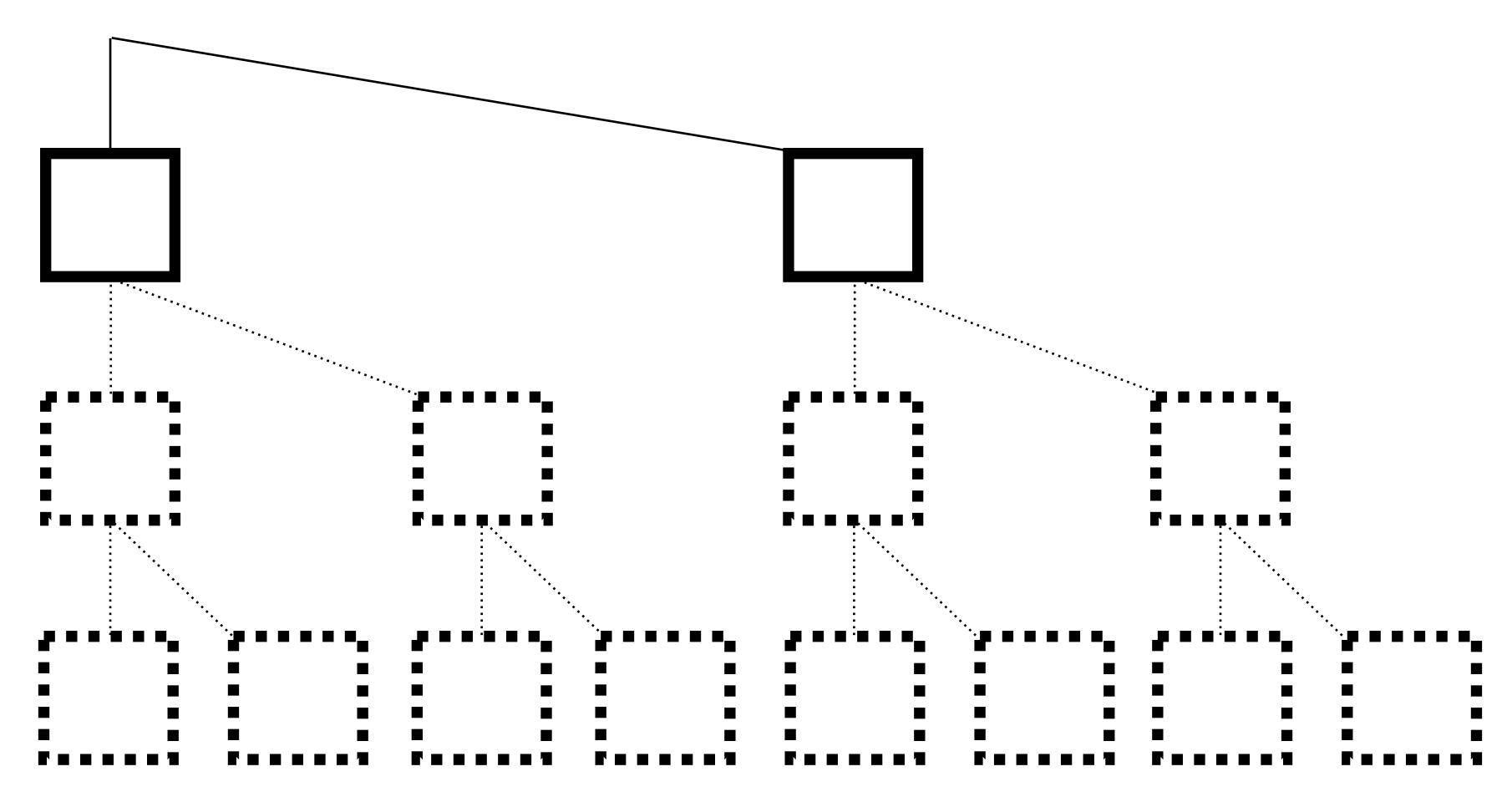
8 UTXOs Hash

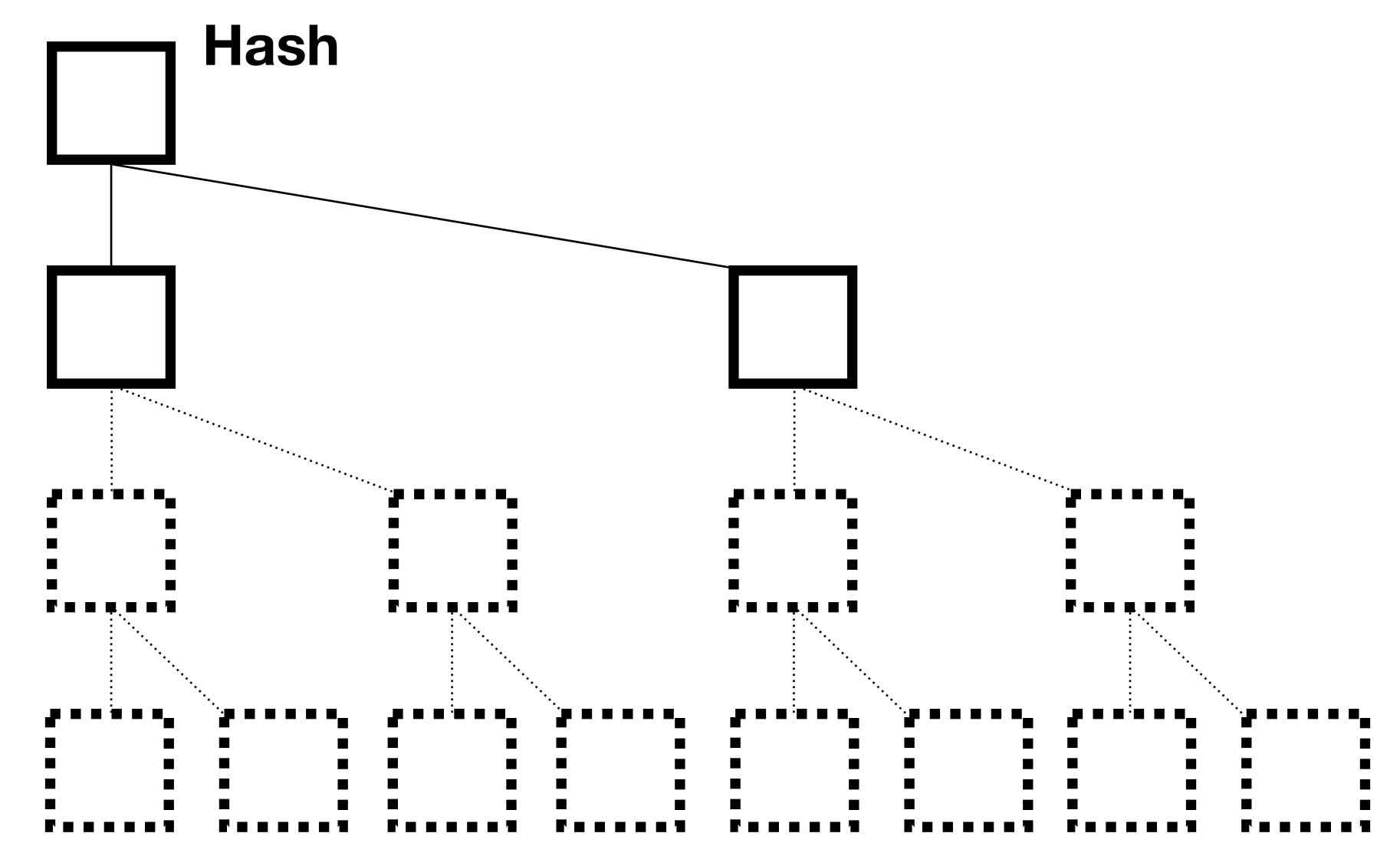


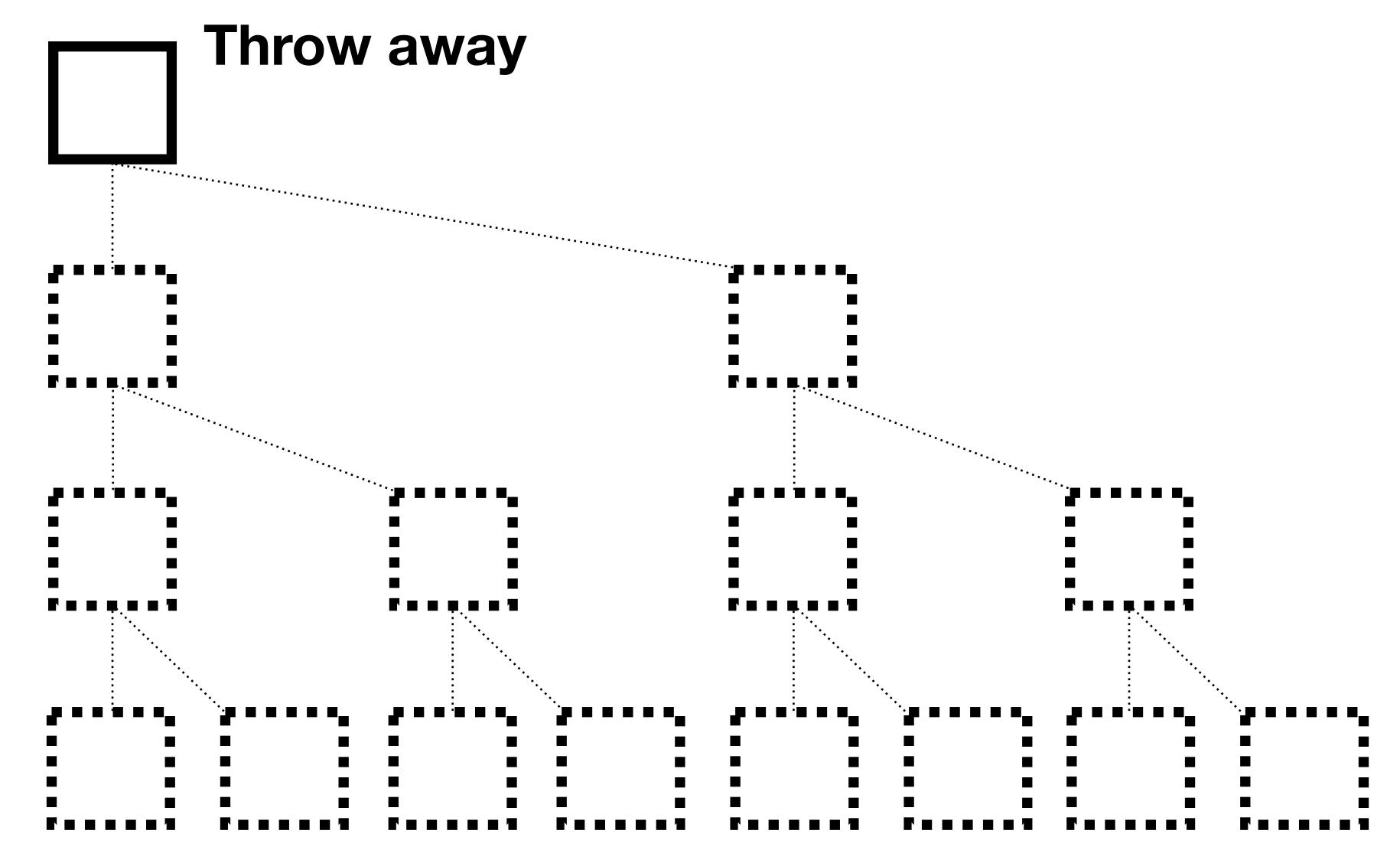
8 UTXOs Throw away



Concatenate







5.1.2 Chaining Value Stack

To help picture the role of the CV stack, Figure 6 shows a growing tree as chunk CVs are added incrementally. As just discussed above, chunk CVs are added to this tree only after the caller has supplied at least 1 byte for the following chunk, so we know that none of these chunks or parent nodes is the root of the tree, and we do not need to worry about the ROOT flag yet.

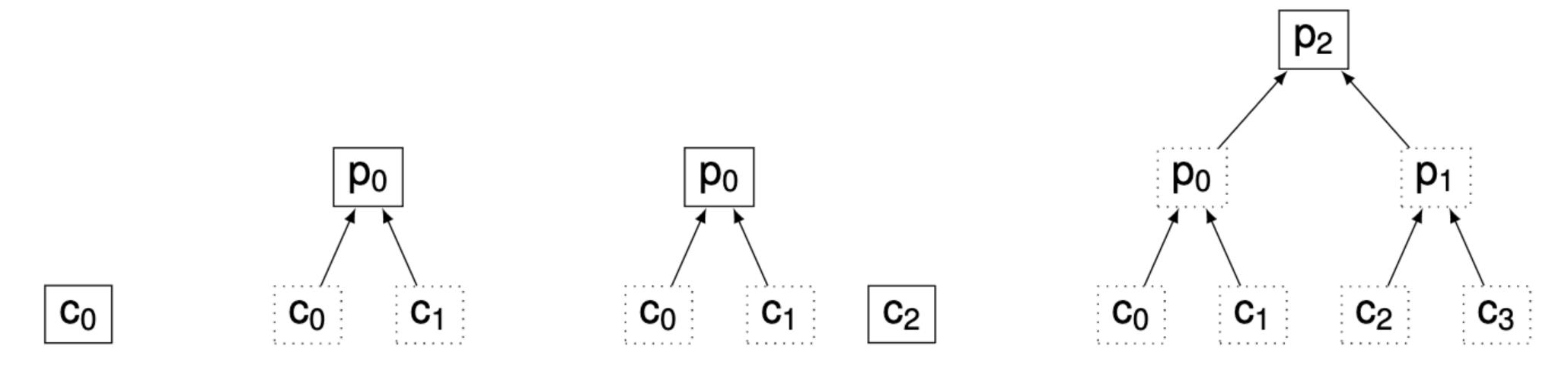


Figure 6: An incomplete tree growing incrementally from 1 to 4 chunks. Dotted boxes represent CVs that no longer need to be stored.

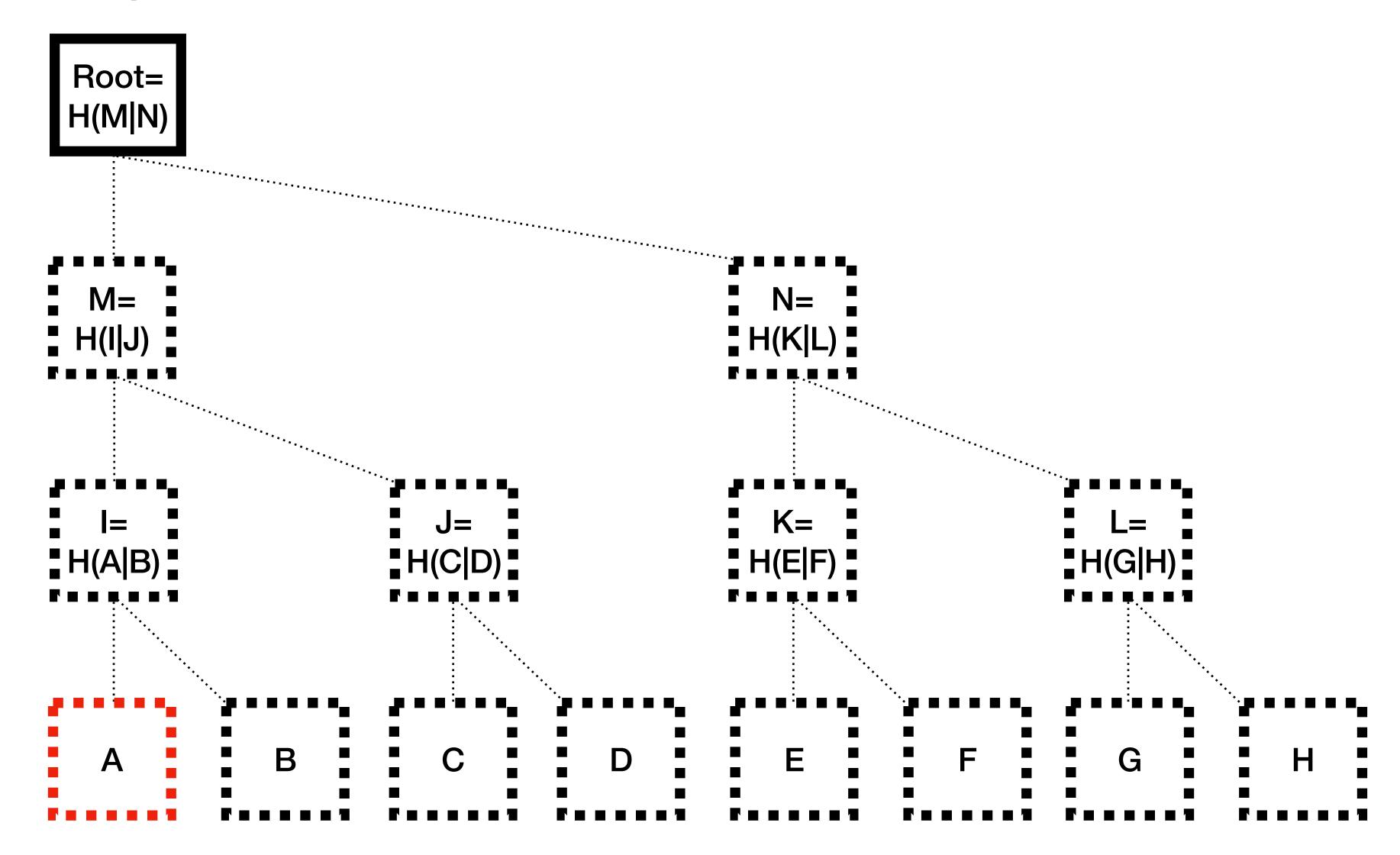
Role of levelDB

It let's you

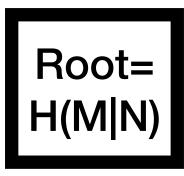
- 1. Add a UTXO
- 2. Delete a UTXO
- 3. Tell you the existence of a UTXO
- 4. Provide the data for verification

Deletion happens with proving existence

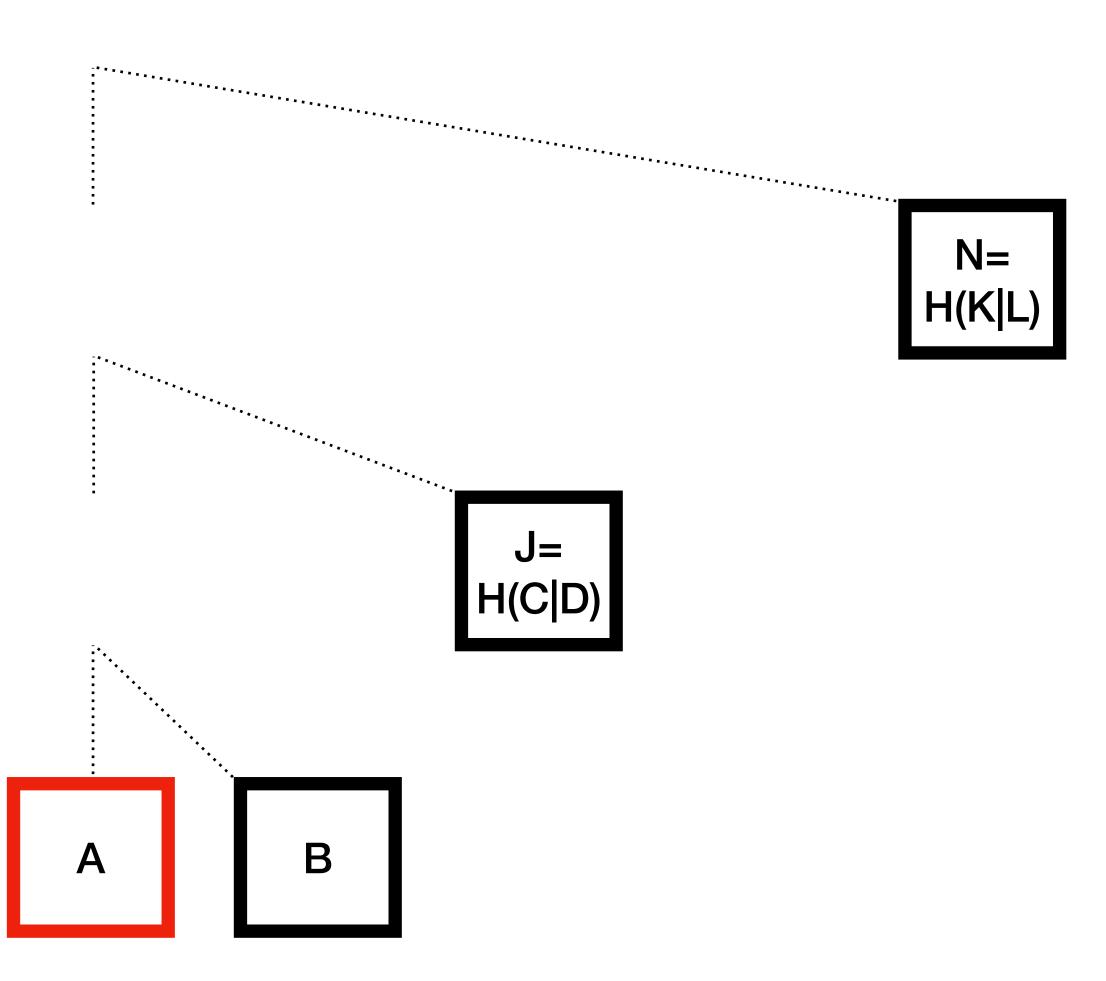
Deleting A



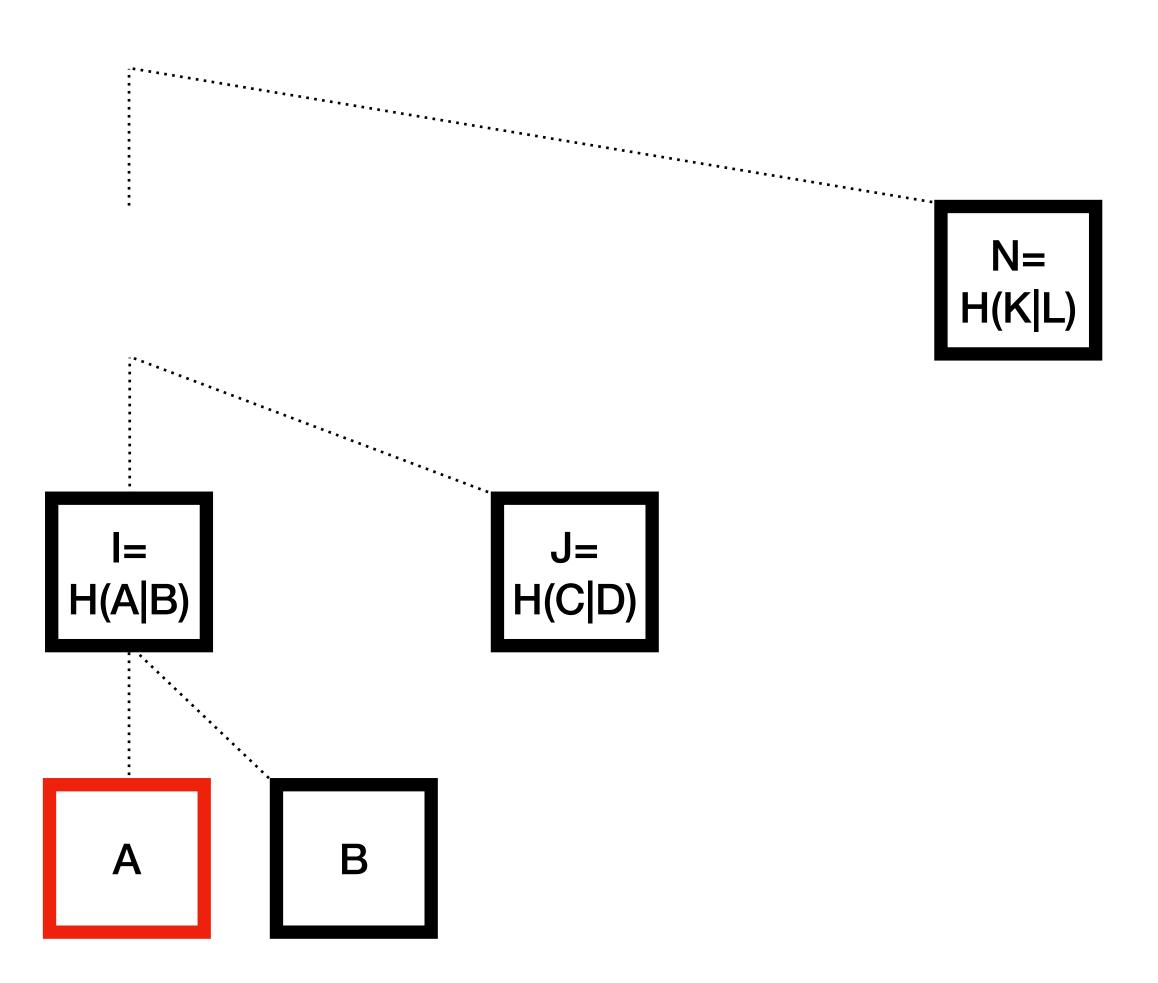
The only requirement is the roots



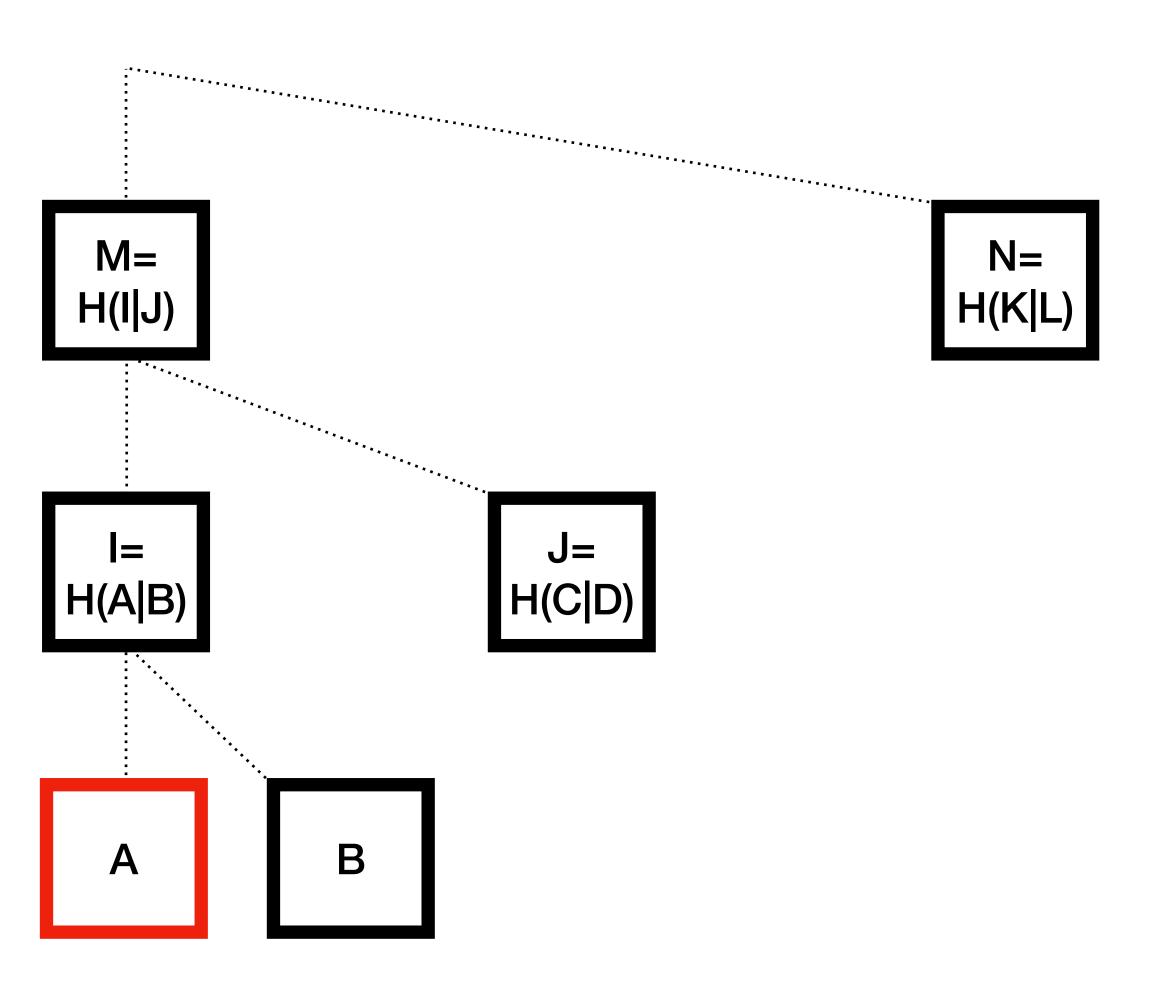
Receive the proof for A



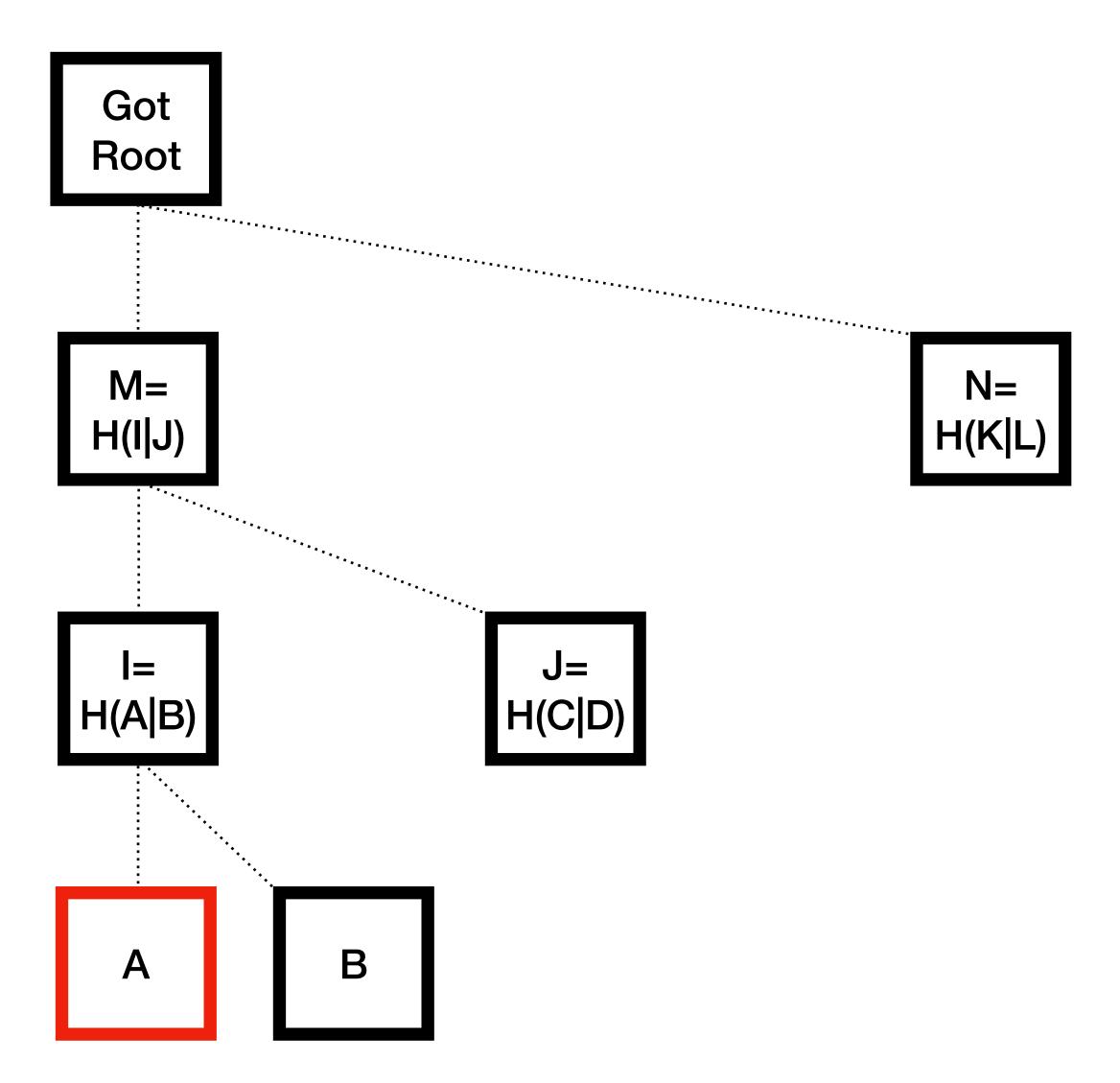
Calculate the hash for I



Calculate the hash for M

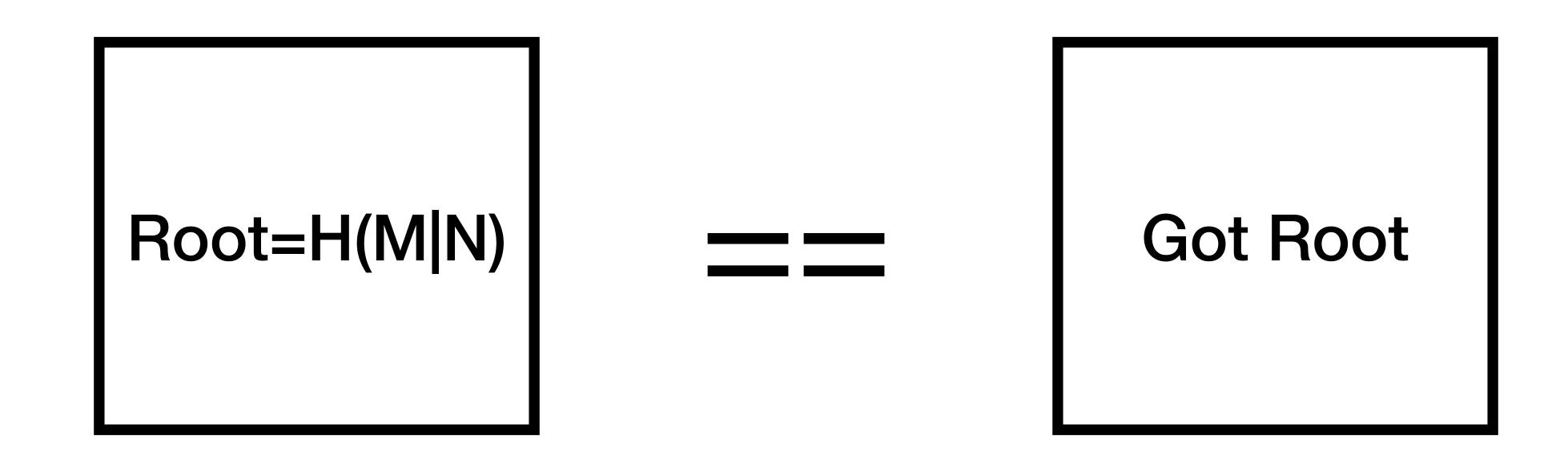


Calculate Root

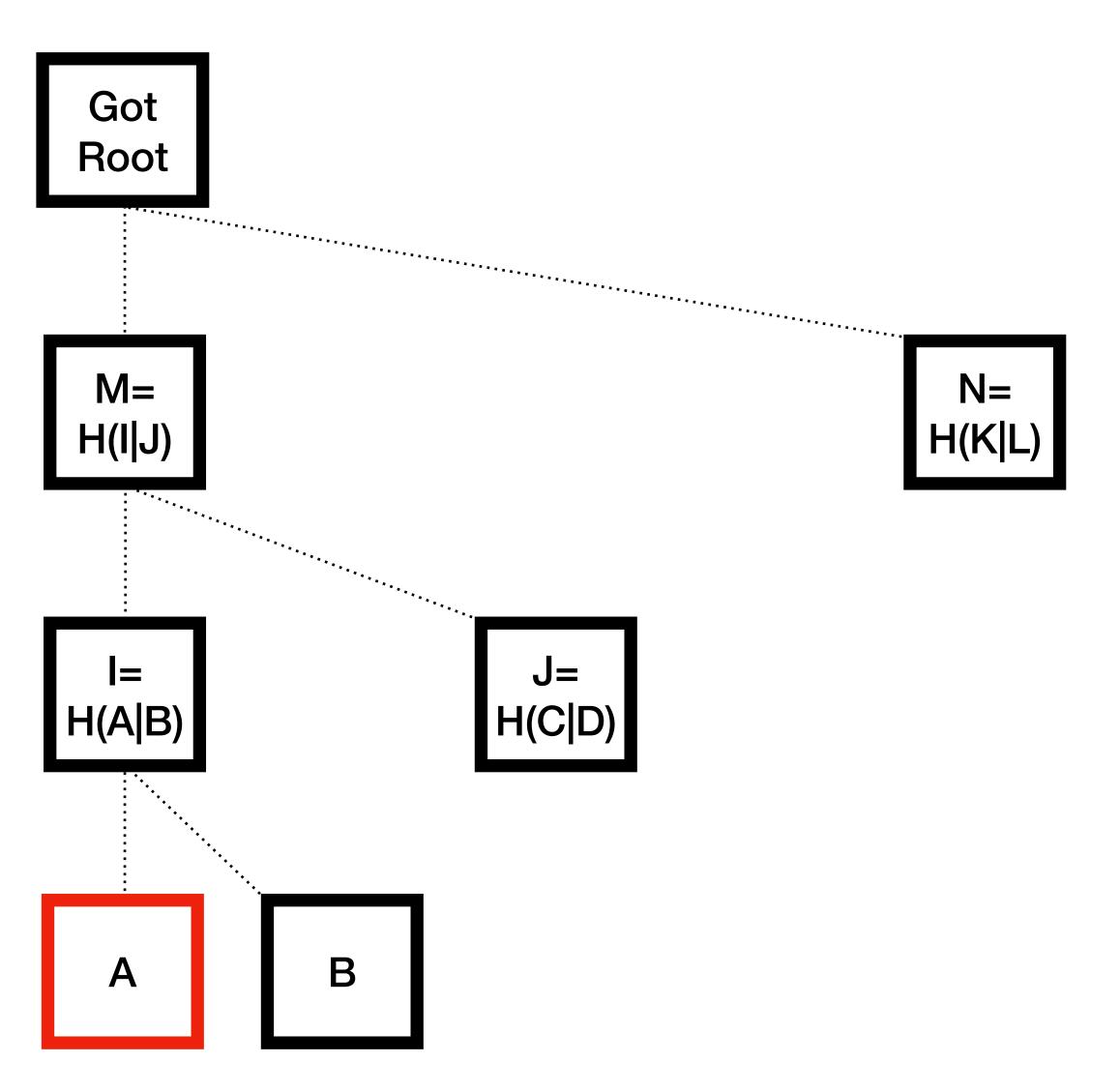


Compare roots

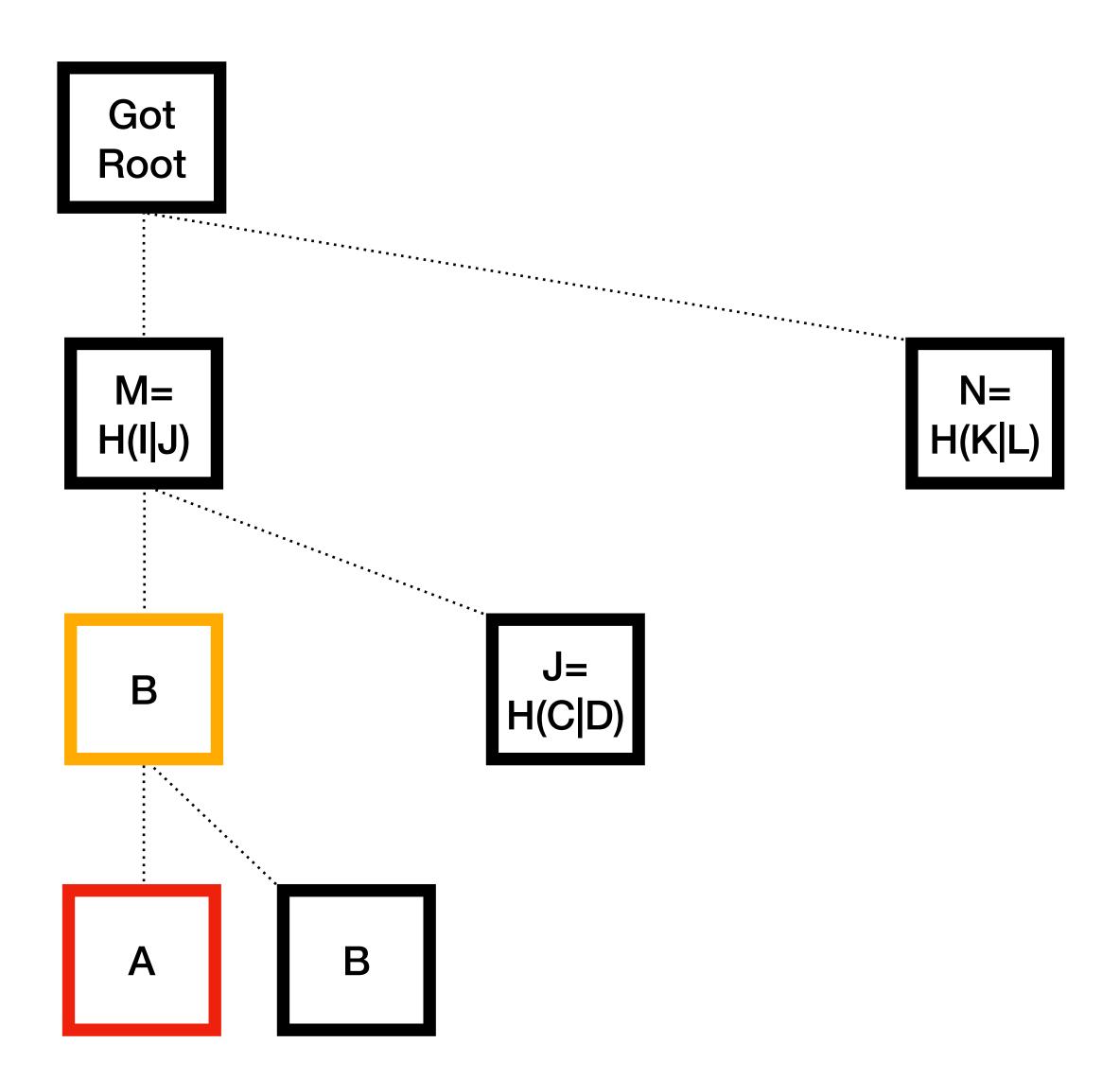
Continue if the roots are equal. Ban peer if not equal



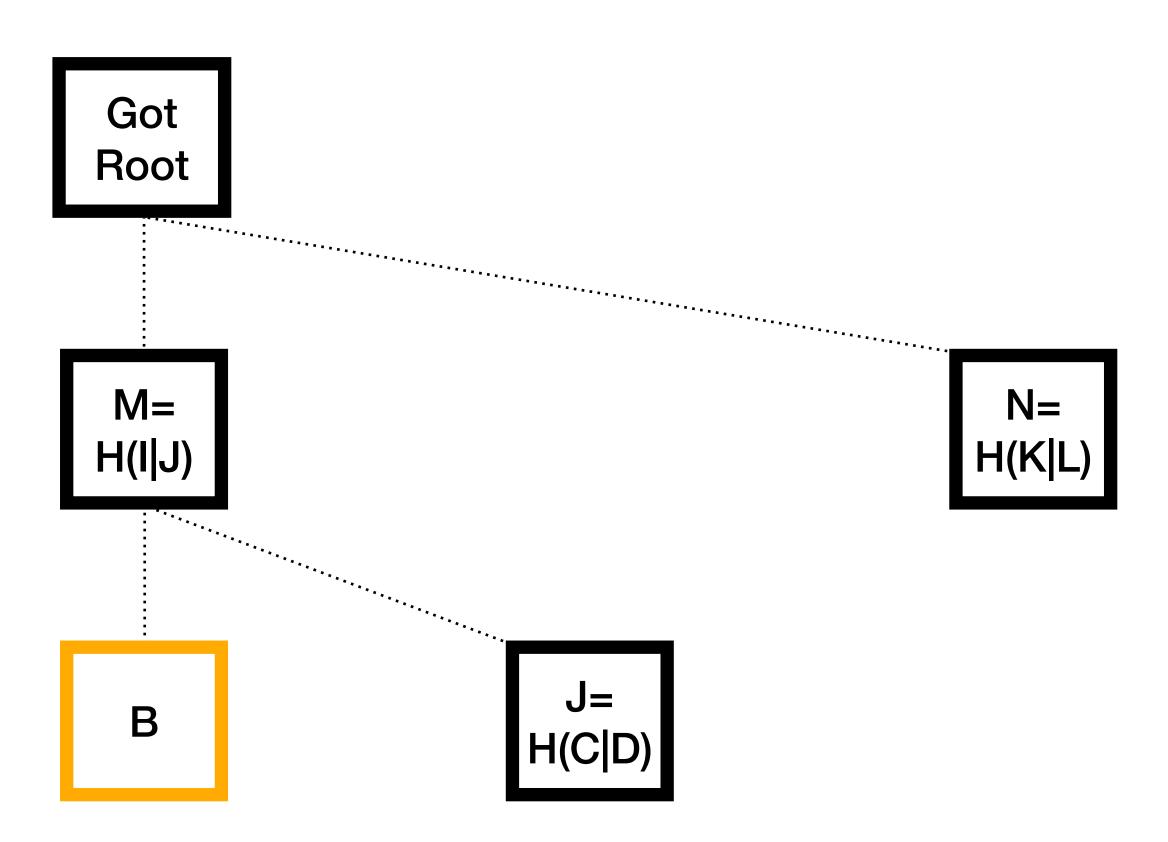
Calculate root after deletion



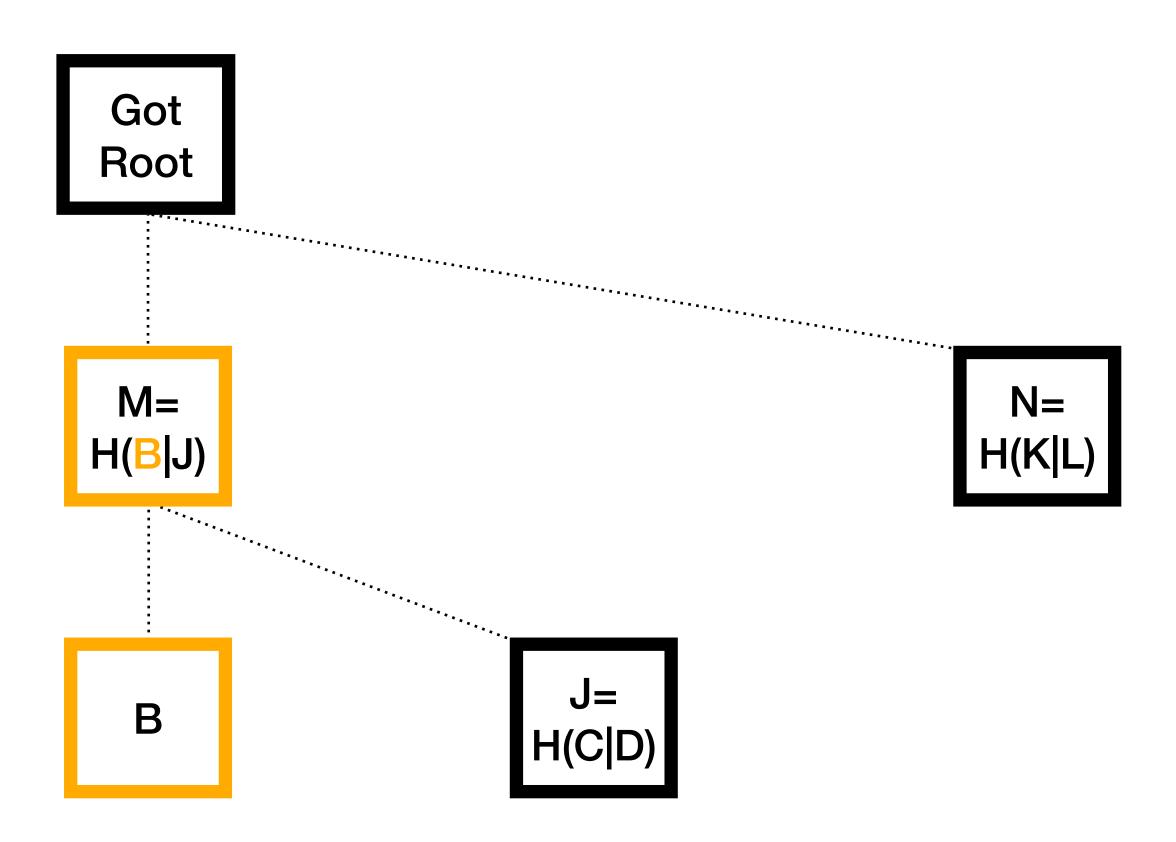
Move up B to I



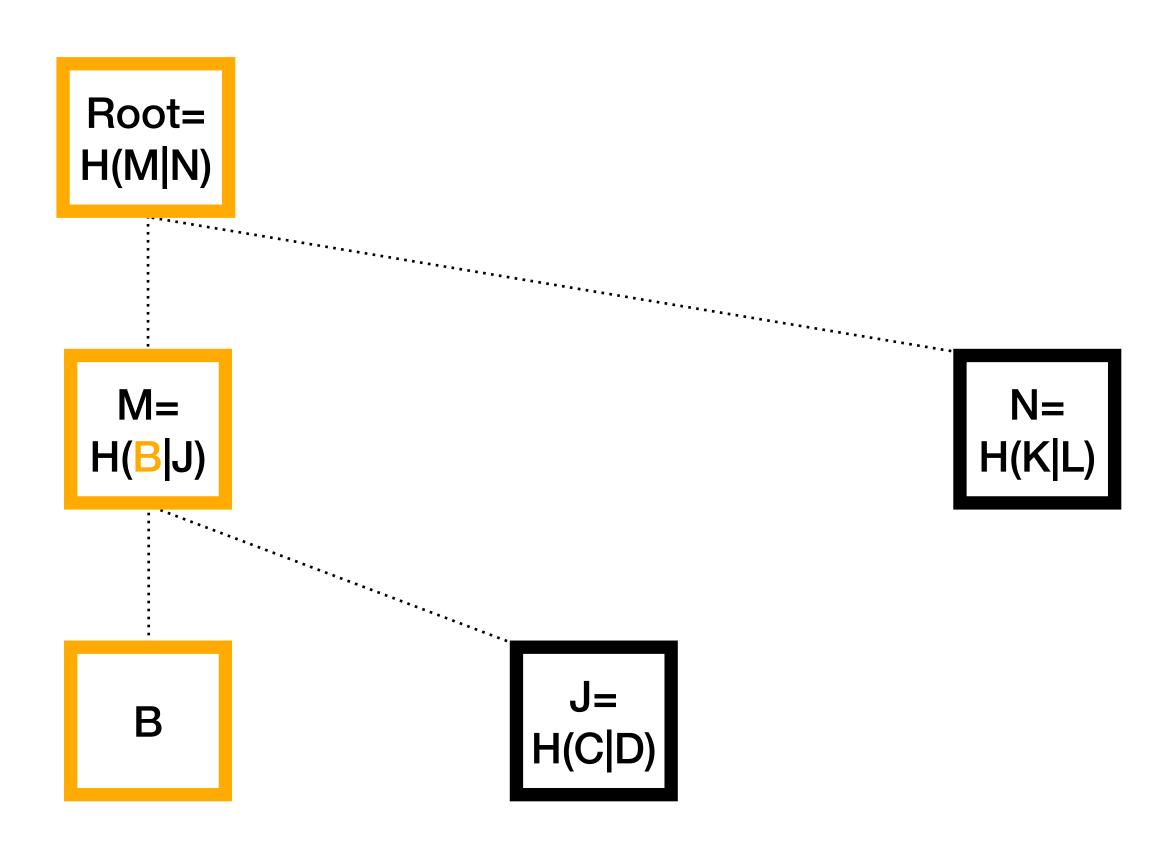
Remove old nodes for A&B



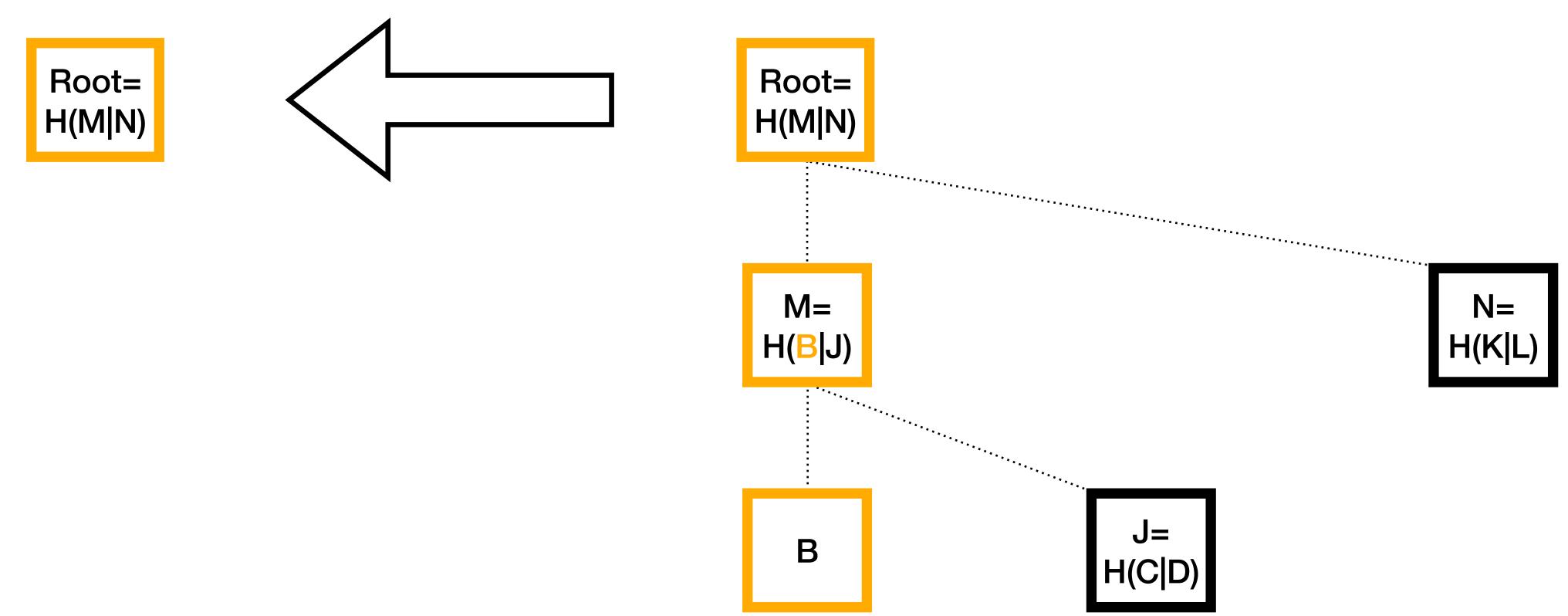
Calculate H(BJ)



Calculate new root

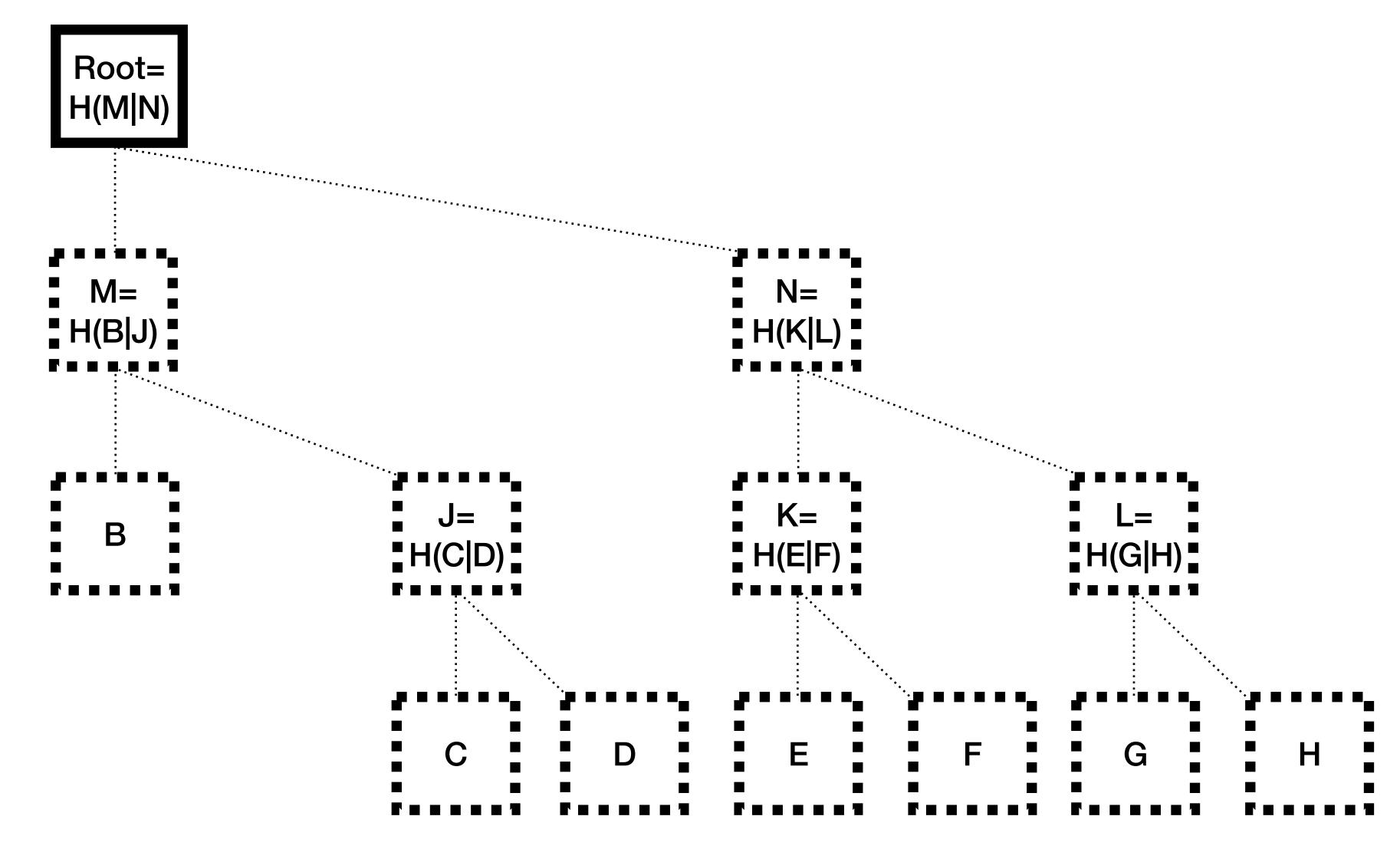


Copy over the new root to be saved

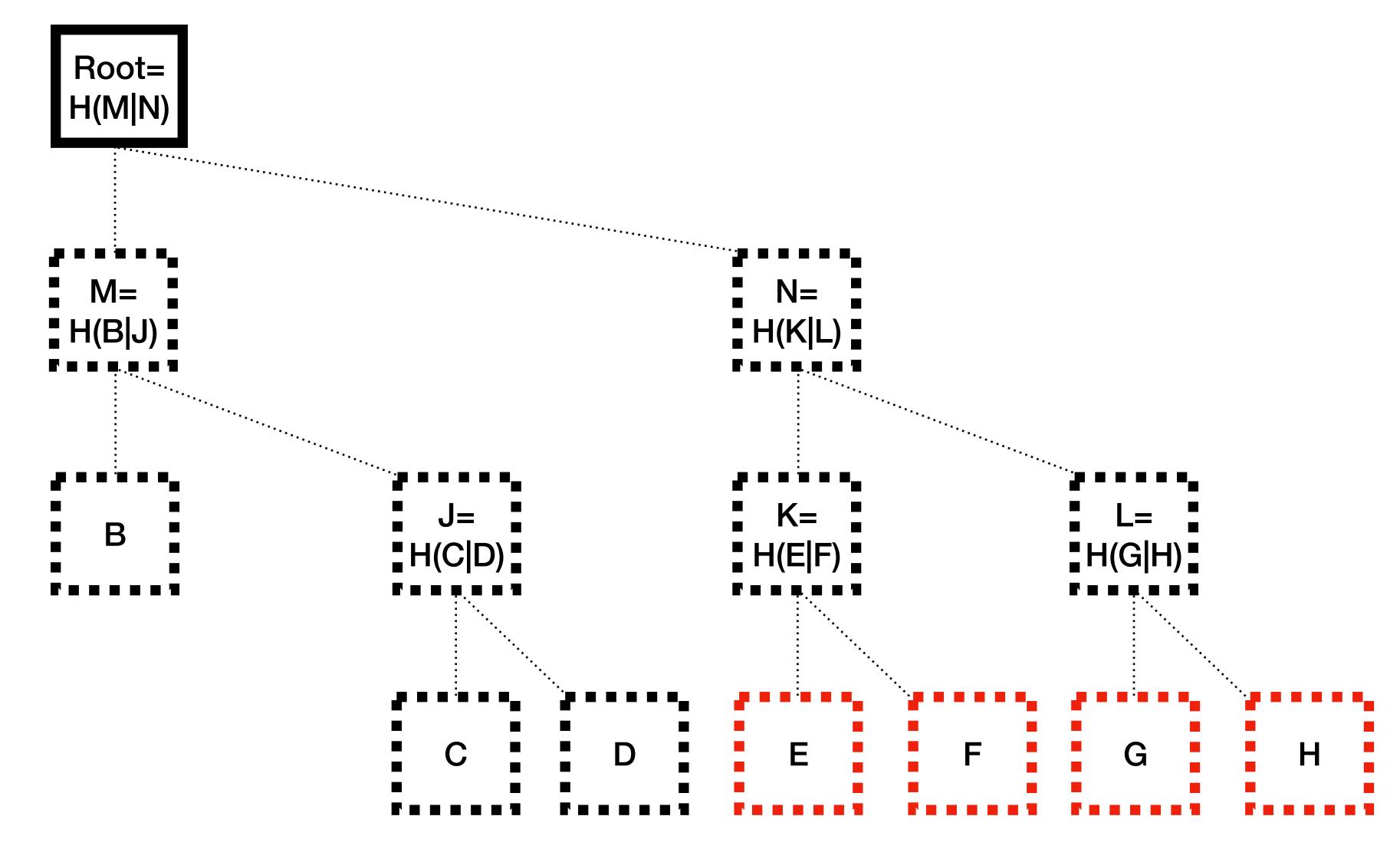


Done

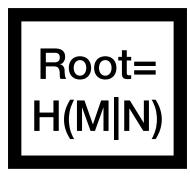
After deleting A



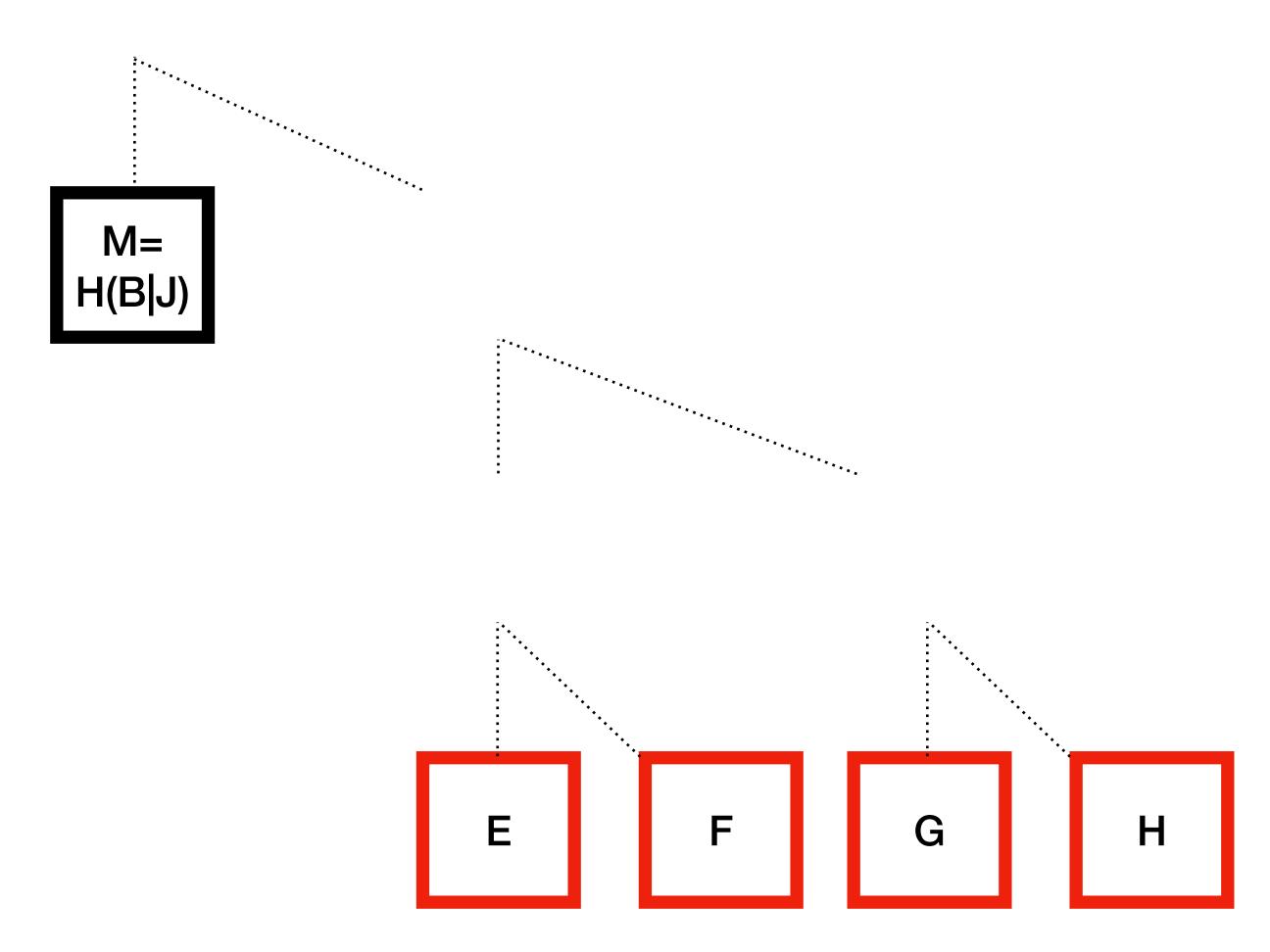
Delete E, F, G, H. (Batched deletions)



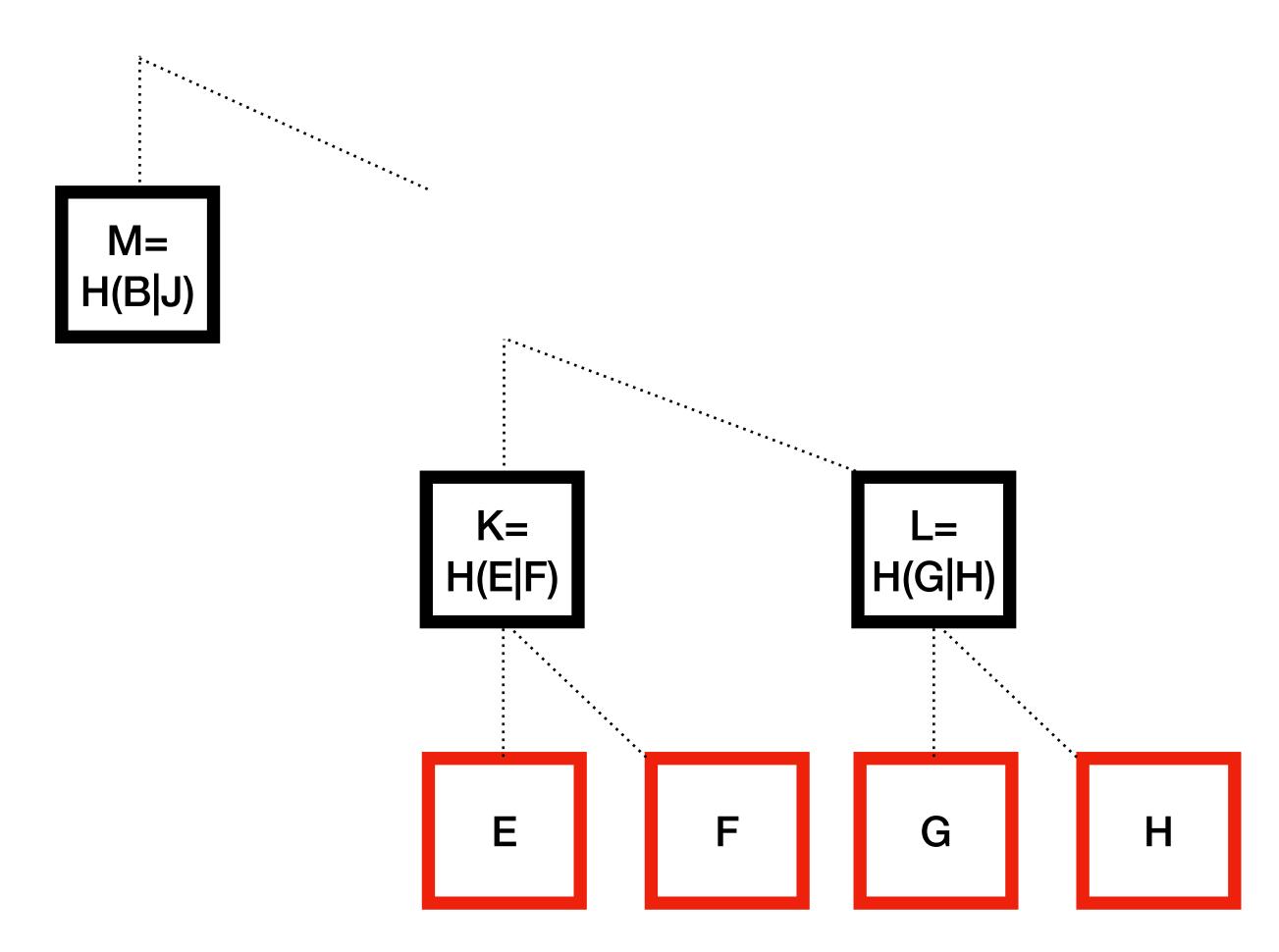
Start off with only the root



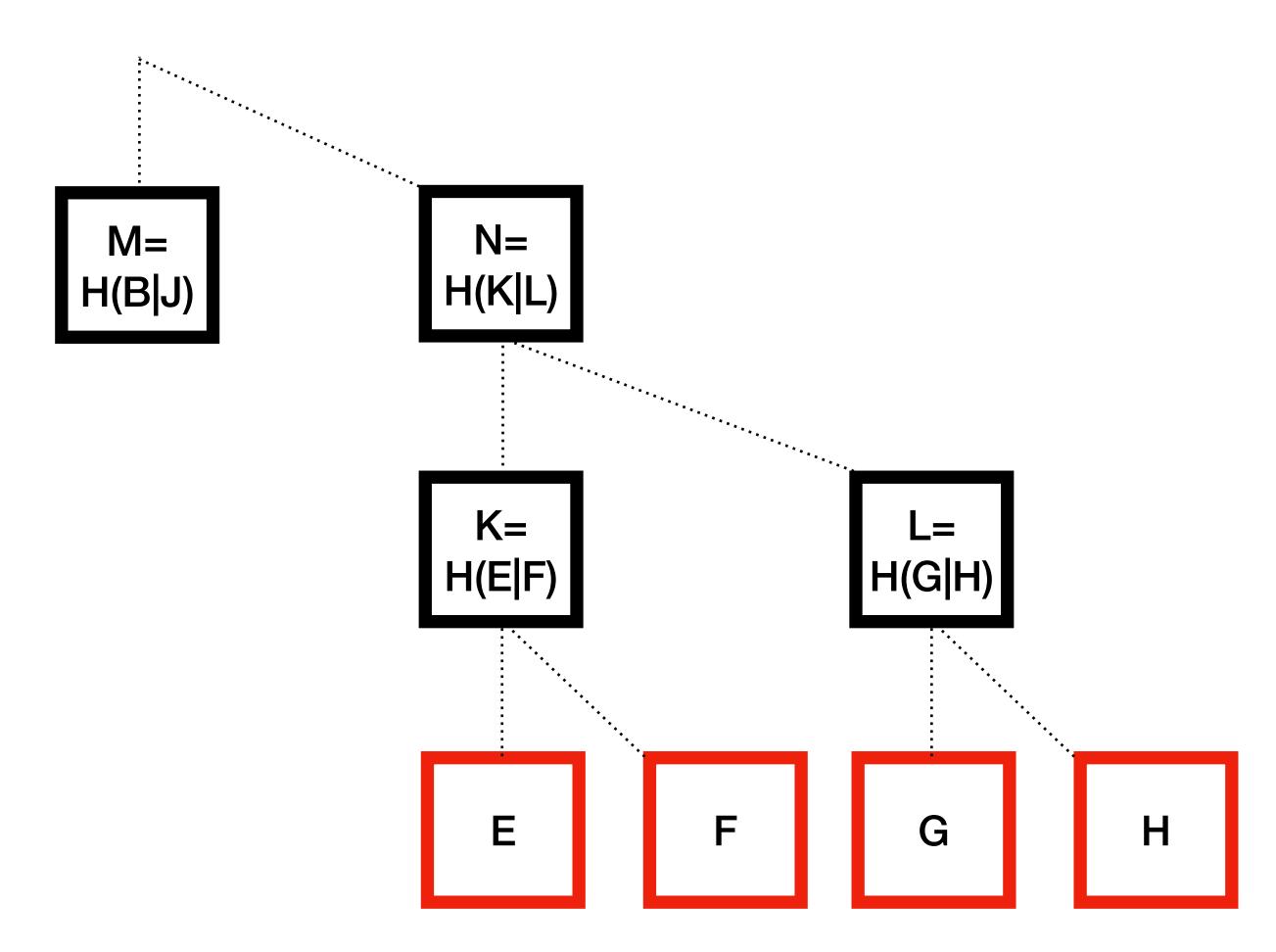
Receive the proof



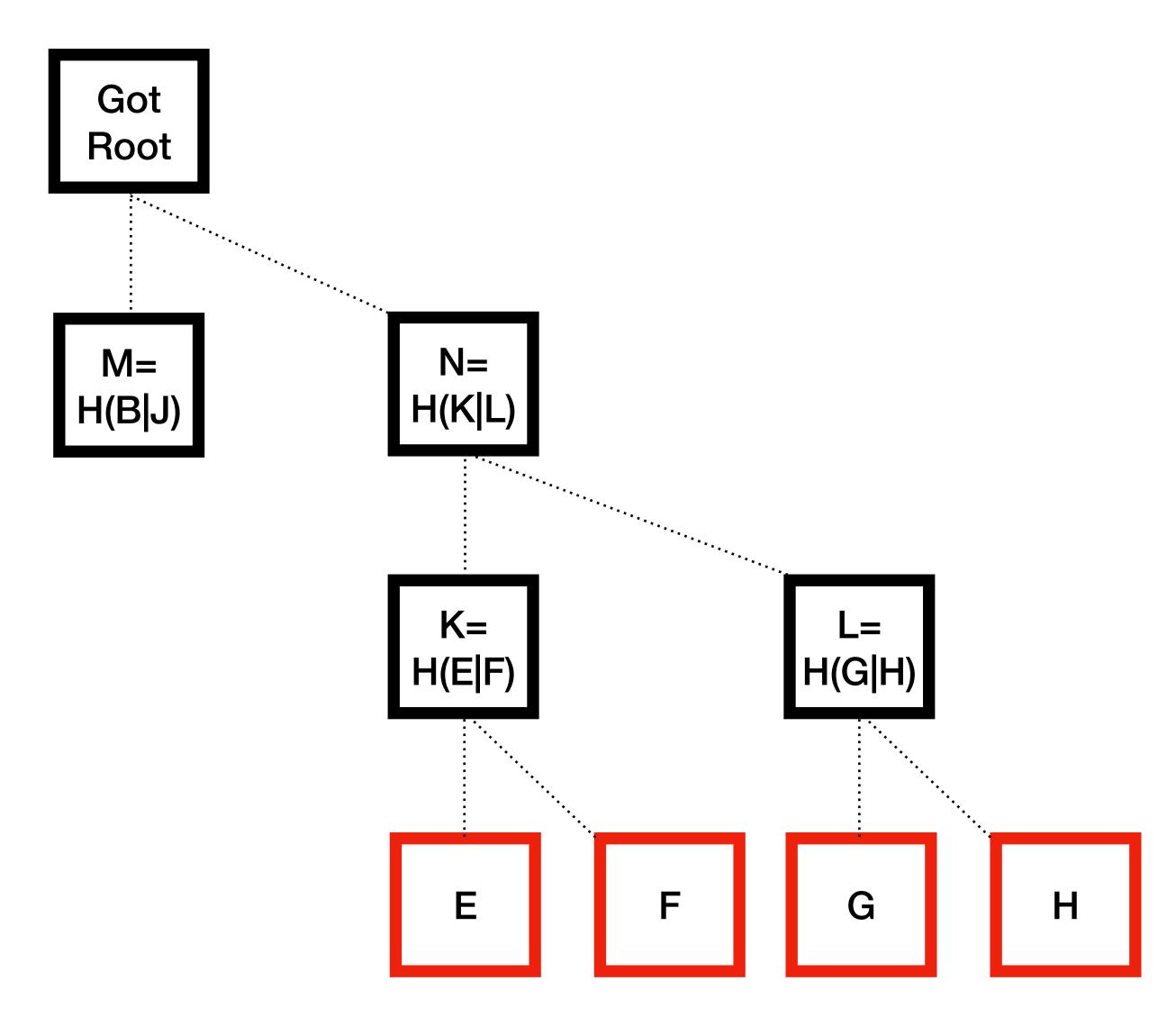
Calculate K and L



Calculate N

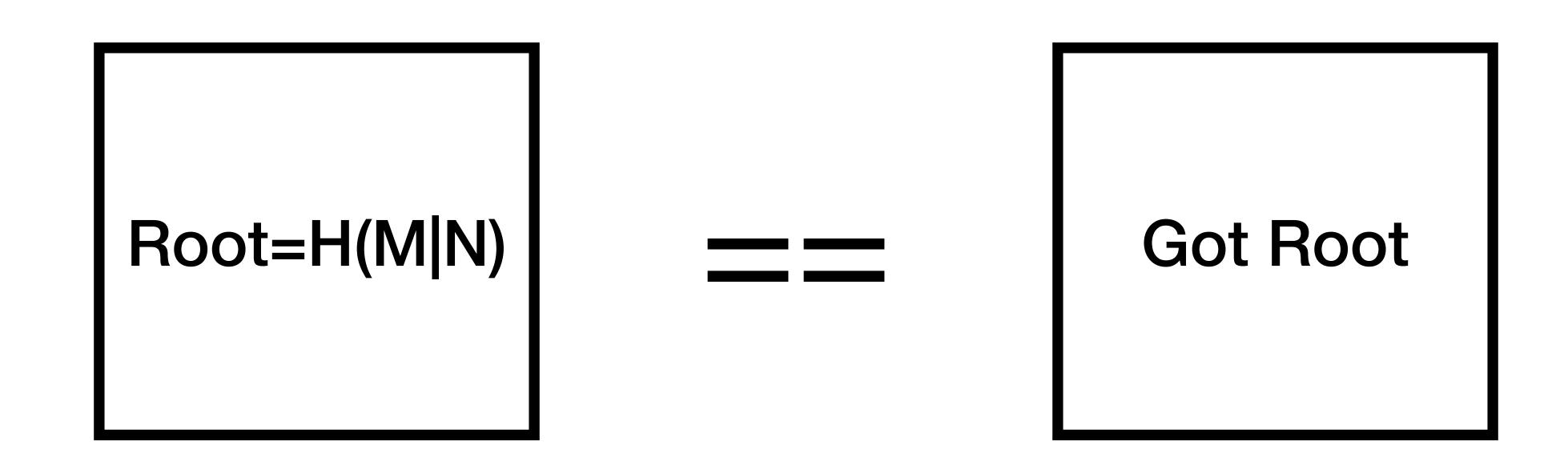


Calculate the Root

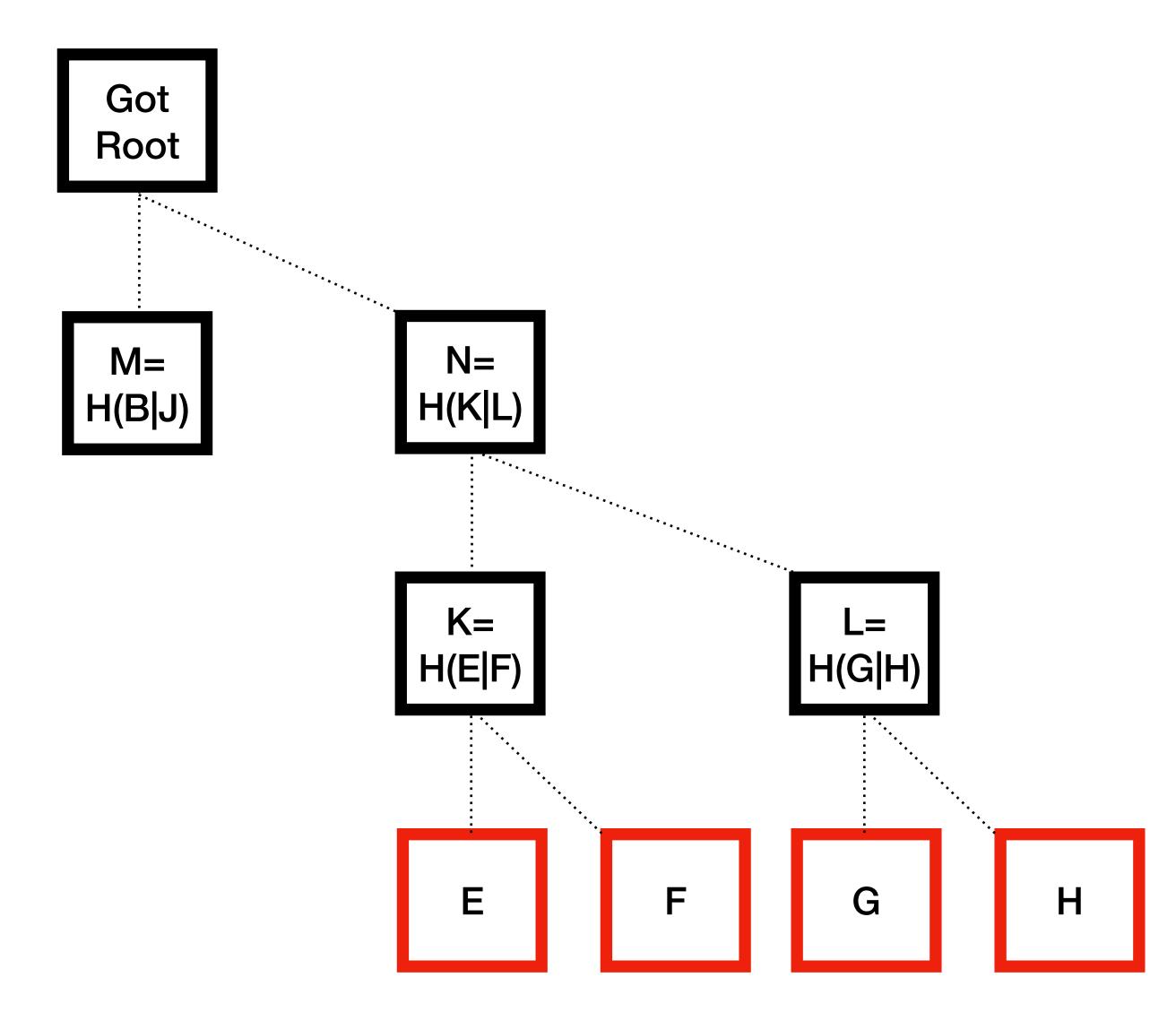


Compare roots

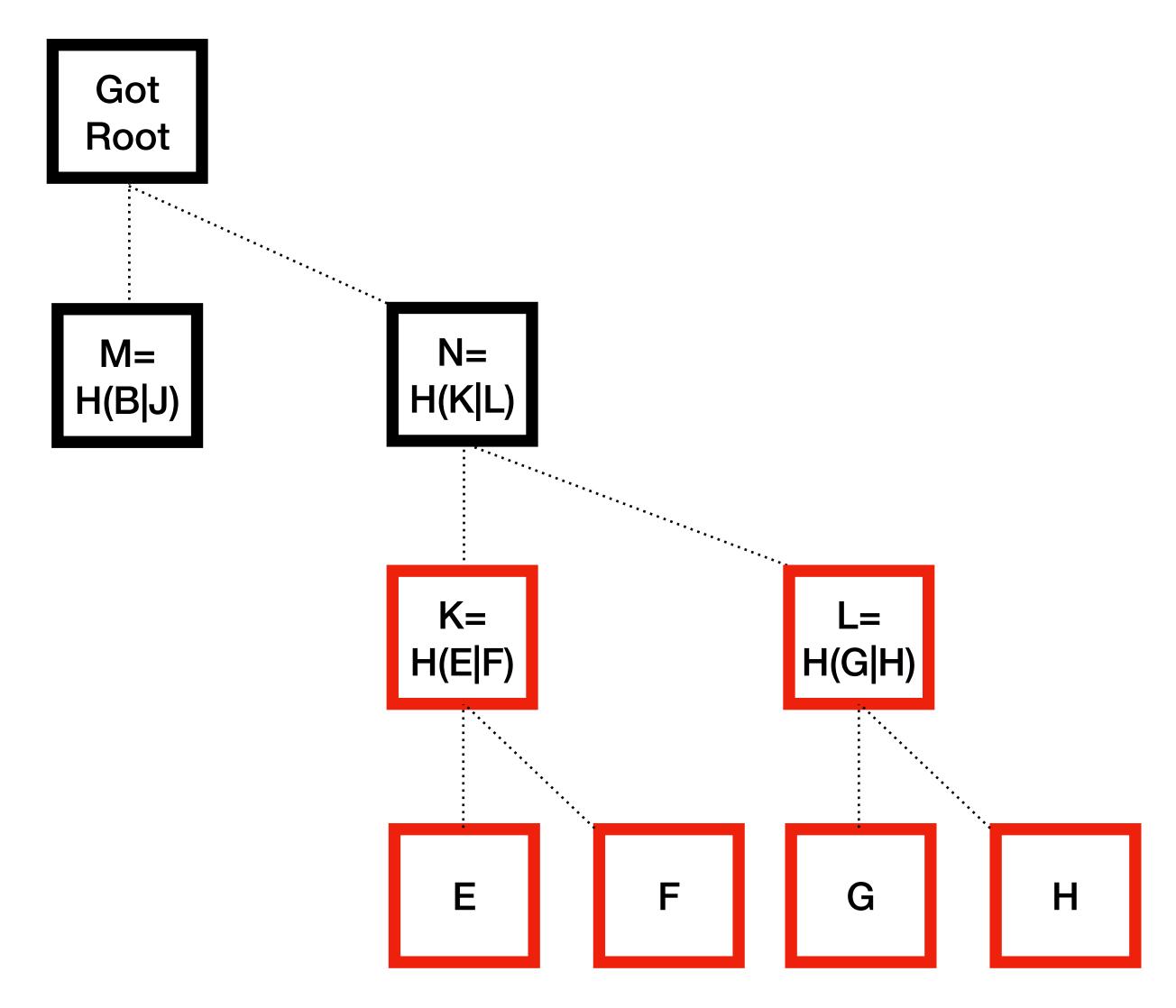
Continue if the roots are equal. Ban peer if not equal



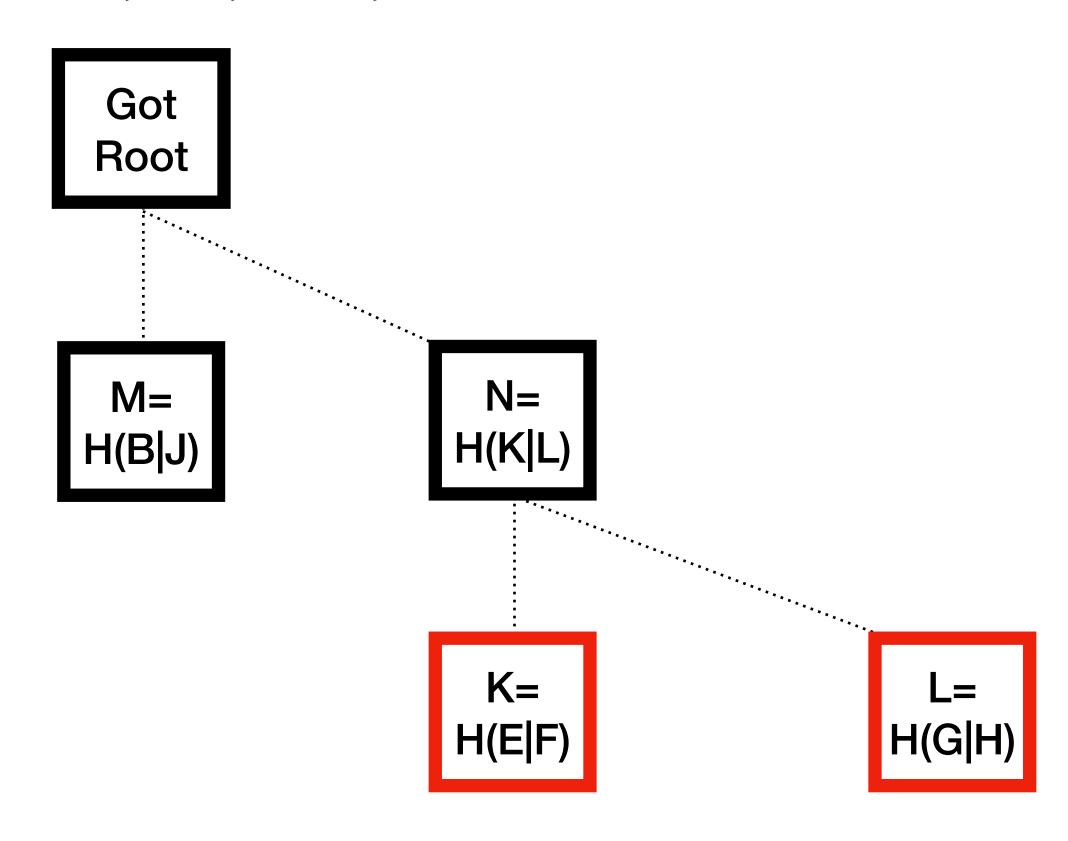
Calculate new root



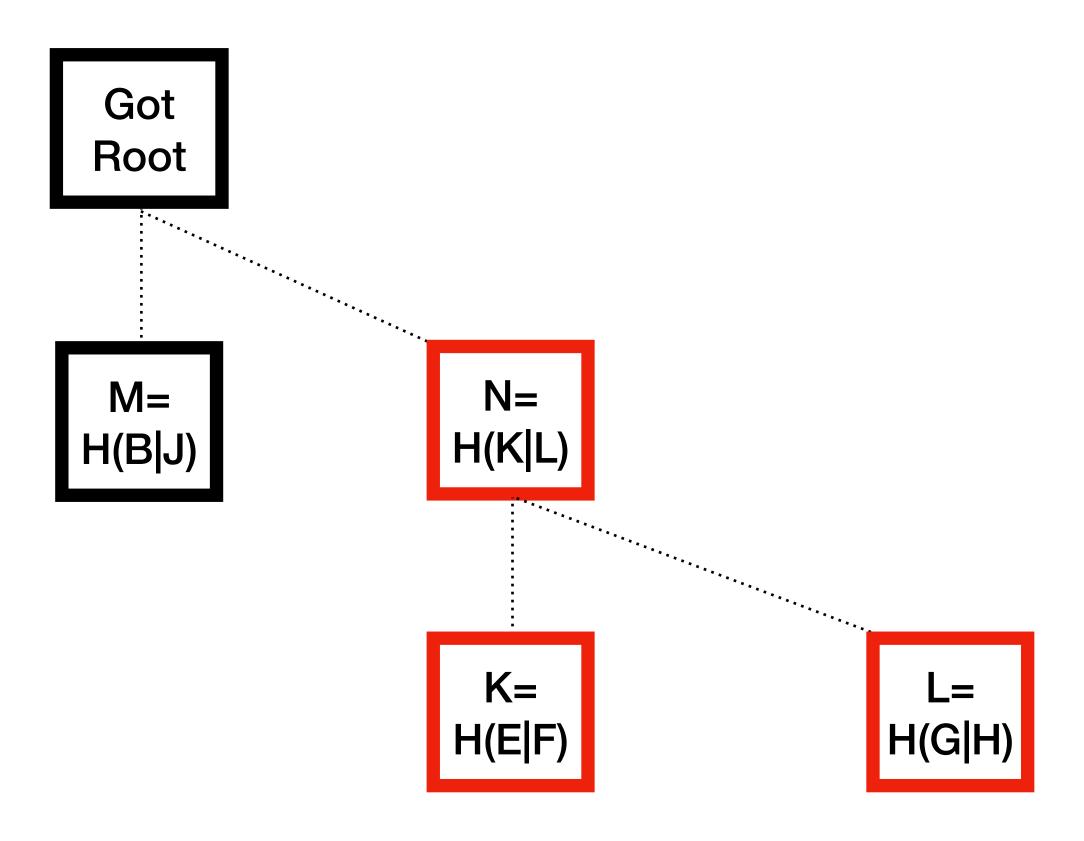
Mark K & L as node to delete



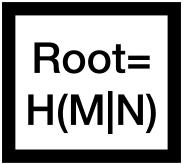
Remove old nodes for E, F, G, H

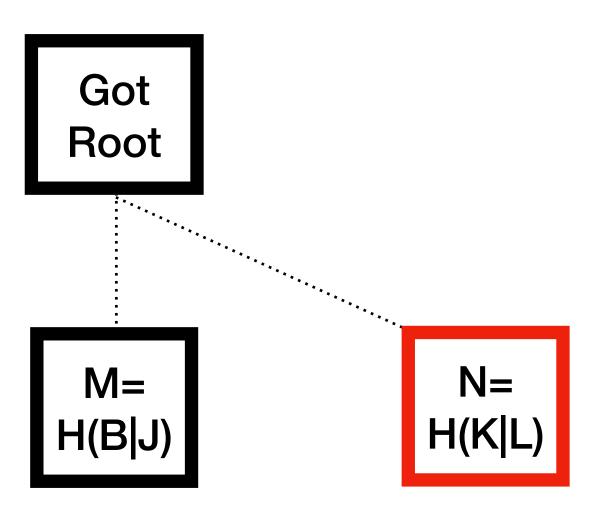


Mark N as node to delete

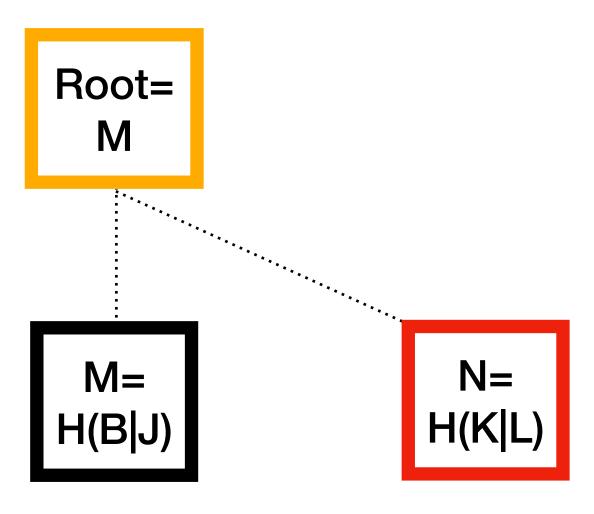


Remove old nodes for K & L

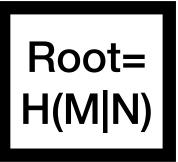


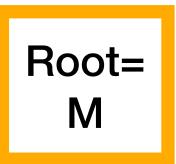


Move up M

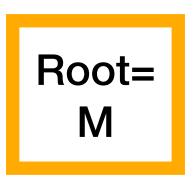


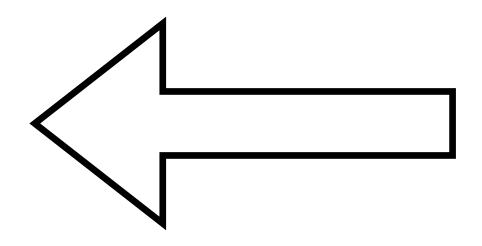
Remove old nodes for M & N

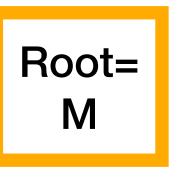




Copy over the new root to be saved



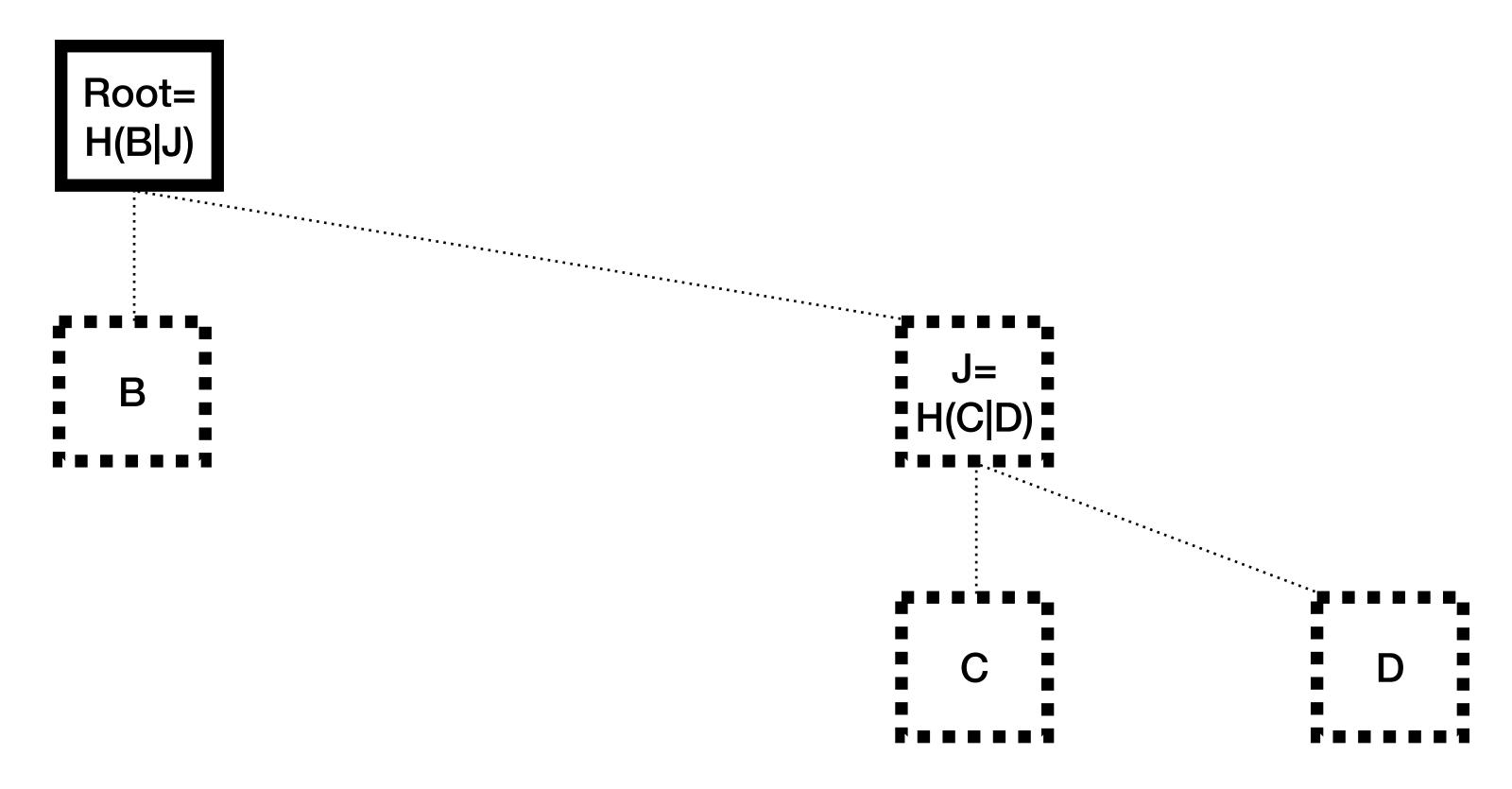




Done

Root= M

After deleting E, F, G, H



Consensus

Add, Verify, Delete

- That's all the relevant algorithm for consensus
- Implemented in 174 lines of Python: github.com/utreexo/pytreexo
- Thanks theStack! (Sebastian Falbesoner)

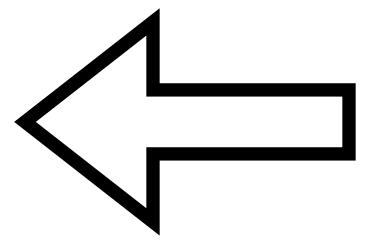
Role of levelDB

It let's you

- 1. Add a UTXO
- 2. Delete a UTXO
- 3. Tell you the existence of a UTXO
- 4. Provide the data for verification

LevelDB Fetching a UTXO

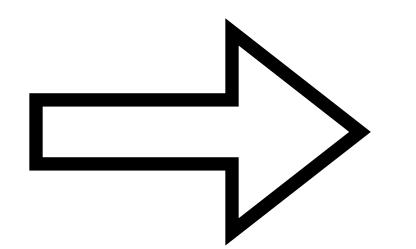




091dd74c2abc5dc5a656450288 0657037d09f56545b5f08365b6 5d21dcc19d2f:1

LevelDB Fetching a UTXO





Amount: 291312

BlockHeight: 791794

IsCoinbase: False

PkScript: 0014703a38f47197ede65d38c3f79f60093f956d6e43

No LevelDB with Utreexo



How the data is provided

What Uteexo nodes download

- Data is provided by the peer
- Sent along with every block or transaction

TX serialization for Utreexo nodes

Not finalized!

Version	
Flag	
TxIn Count	
TxIns	
TxOut Count	
TxOuts	
Witness	
Locktime	
Utreexo Proof Data	

TX serialization for Utreexo nodes

Not finalized!

Version	
Flag	
TxIn Count	
TxIns	
TxOut Count	
TxOuts	
Witness	
Locktime	
Utreexo Proof Data	

Utreexo Proof Data

Not finalized!

Merkle Proof

UTXO
Data

Block serialization for Utreexo nodes

Not finalized!

Version	
Previous Block Hash	
Merkle Root	
Timestamp	
Difficulty Bits	
Nonce	
Transaction Count	
Transactions	
Batched Utreexo Proof Data	

Block serialization for Utreexo nodes

Not finalized!

Version	
Previous Block Hash	
Merkle Root	
Timestamp	
Difficulty Bits	
Nonce	
Transaction Count	
Transactions	
Batched Utreexo Proof Data	

Batched Utreexo Proof Data

Not finalized!

Batched Merkle Proof

UTXO data Count

UTXO Datas

2 serialization methods

 For calculating the hash to be committed into the accumulator

Block Hash	The block hash where the tx was included
TxHash	The transaction hash for the UTXO
Index (Vout)	The index within the TX for this UTXO
Block Height + IsCoinBase (blockHeight << 1)&IsCoinBase	Block height and Coinbase indicator
Amount	Amount in satoshis
PkScript length	Length of the pkscript in varint
PkScript	PkScript itself

Block Hash	The block hash where the tx was included
TxHash	The transaction hash for the UTXO
Index (Vout)	The index within the TX for this UTXO
Block Height + IsCoinBase (blockHeight << 1)&IsCoinBase	Block height and Coinbase indicator
Amount	Amount in satoshis
PkScript length	Length of the pkscript in varint
PkScript	PkScript itself

Block Hash	The block hash where the tx was included
TxHash	The transaction hash for the UTXO
Index (Vout)	The index within the TX for this UTXO
Block Height + IsCoinBase (blockHeight << 1)&IsCoinBase	Block height and Coinbase indicator
Amount	Amount in satoshis
PkScript length	Length of the pkscript in varint
PkScript	PkScript itself

Block Hash	The block hash where the tx was included
TxHash	The transaction hash for the UTXO
Index (Vout)	The index within the TX for this UTXO
Block Height + IsCoinBase (blockHeight << 1)&IsCoinBase	Block height and Coinbase indicator
Amount	Amount in satoshis
PkScript length	Length of the pkscript in varint
PkScript	PkScript itself

Block Hash	The block hash where the tx was included
TxHash	The transaction hash for the UTXO
Index (Vout)	The index within the TX for this UTXO
Block Height + IsCoinBase (blockHeight << 1)&IsCoinBase	Block height and Coinbase indicator
Amount	Amount in satoshis
PkScript length	Length of the pkscript in varint
PkScript	PkScript itself

Block Hash	The block hash where the tx was included
TxHash	The transaction hash for the UTXO
Index (Vout)	The index within the TX for this UTXO
Block Height + IsCoinBase (blockHeight << 1)&IsCoinBase	Block height and Coinbase indicator
Amount	Amount in satoshis
PkScript length	Length of the pkscript in varint
PkScript	PkScript itself

Block Hash	The block hash where the tx was included
TxHash	The transaction hash for the UTXO
Index (Vout)	The index within the TX for this UTXO
Block Height + IsCoinBase (blockHeight << 1)&IsCoinBase	Block height and Coinbase indicator
Amount	Amount in satoshis
PkScript length	Length of the pkscript in varint
PkScript	PkScript itself

Block Hash	The block hash where the tx was included
TxHash	The transaction hash for the UTXO
Index (Vout)	The index within the TX for this UTXO
Block Height + IsCoinBase (blockHeight << 1)&IsCoinBase	Block height and Coinbase indicator
Amount	Amount in satoshis
PkScript length	Length of the pkscript in varint
PkScript	PkScript itself

2 serialization methods

- For calculating the hash to be committed into the accumulator
- For sending data to peers/storage on disk

For disk storage/p2p transfer

Block Height + IsCoinBase (blockHeight << 1)&IsCoinBase	Block height and Coinbase indicator
Amount	Amount in satoshis
IsReconstructablePkScript	True if pkscript can be fetched from the spending TxIn. Nil pkscript for p2pkh&p2sh
PkScript length	Length of the pkscript in varint
PkScript	PkScript itself

Block Height + IsCoinBase (blockHeight << 1)&IsCoinBase	Block height and Coinbase indicator
Amount	Amount in satoshis
IsReconstructablePkScript	True if pkscript can be fetched from the spending TxIn. Nil pkscript for p2pkh&p2sh
PkScript length	Length of the pkscript in varint
PkScript	PkScript itself

Block Height + IsCoinBase (blockHeight << 1)&IsCoinBase	Block height and Coinbase indicator
Amount	Amount in satoshis
IsReconstructablePkScript	True if pkscript can be fetched from the spending Txln. Nil pkscript for p2pkh&p2sh
PkScript length	Length of the pkscript in varint
PkScript	PkScript itself

Block Height + IsCoinBase (blockHeight << 1)&IsCoinBase	Block height and Coinbase indicator
Amount	Amount in satoshis
IsReconstructablePkScript	True if pkscript can be fetched from the spending TxIn. Nil pkscript for p2pkh&p2sh
PkScript length	Length of the pkscript in varint
PkScript	PkScript itself

Block Height + IsCoinBase (blockHeight << 1)&IsCoinBase	Block height and Coinbase indicator
Amount	Amount in satoshis
IsReconstructablePkScript	True if pkscript can be fetched from the spending TxIn. Nil pkscript for p2pkh&p2sh
PkScript length	Length of the pkscript in varint
PkScript	PkScript itself

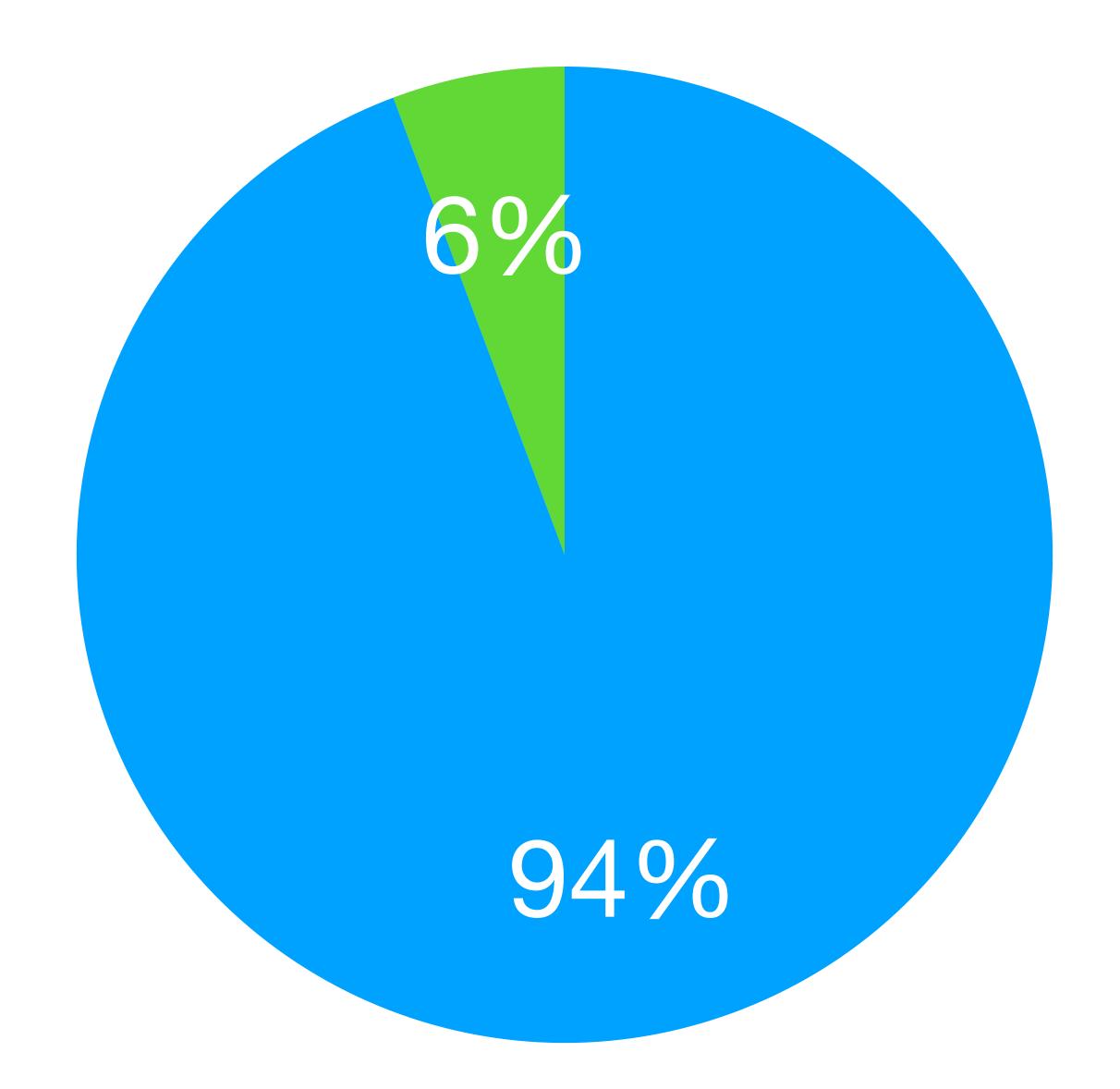
Block Height + IsCoinBase (blockHeight << 1)&IsCoinBase	Block height and Coinbase indicator
Amount	Amount in satoshis
IsReconstructablePkScript	True if pkscript can be fetched from the spending TxIn. Nil pkscript for p2pkh&p2sh
PkScript length	Length of the pkscript in varint
PkScript	PkScript itself

Total overhead

Extra data that a utreexo node will download

- 364GB of extra data (as of block 710,000)
- Caching and batch proving multiple blocks is being explored





Current Progress

github.com/utreexo

Main github organization

- Accumulator implementation in Go
- Full node with Utreexo
- Full node with electrum-personal-server capability. Full support for protocol v1.4.1.
- Pytreexo

github.com/mit-dci/rustreexo

Accumulator in Rust

- Accumulator implementation in Rust
- Currently missing bridge node capability (2023-05-30)

github.com/Davidson-Souza/Floresta

Full node in Rust

- Full node with Utreexo in Rust
- Supports electrum personal server capabilities

- Accumulator design
- Working full node
- Wallet support
- P2P protocol that supports caching
- Efficient mempool with Utreexo



- Accumulator design
 - Working full node
 - Wallet support
 - P2P protocol that supports caching
 - Efficient mempool with Utreexo



- Accumulator design
 - Working full node
 - Wallet support
 - P2P protocol that supports caching
 - Efficient mempool with Utreexo



- Accumulator design
- Working full node
 - Wallet support
 - P2P protocol that supports caching
 - Efficient mempool with Utreexo

Rust implementation

Things going on the Rust side by Davidson



- Working compact utreexo node
 - C bindings
 - Python bindings
 - Javascript bindings

Twitter: @kcalvinalvinn



Slides:

github.com/kcalvinalvin/ slides-for-btcprague

