



Blockchain Technologies for Data Scientists

Bruno Gonçalves

www.data4sci.com

<https://github.com/DataForScience/blockchain-data>

Disclaimer

The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of my employers. The examples provided with this tutorial were chosen for their didactic value and are not meant to be representative of my day to day work.

References



How Money Works

<https://www.continentalcurrency.ca/wp-content/uploads/2015/06/continental-currency-exchange.jpg>
https://upload.wikimedia.org/wikipedia/commons/2/2b/Olaus_Magnus_-_On_Trade_Without_Using_Money.jpg

- Many different types of currency have been used throughout the years

- Gold
- Cattle
- Salt
- Rocks
- Sea Shells
- etc...



- Fundamental functions (and properties) of money:
 - Medium of Exchange (**Fungibility**, **Portability**)
 - Unit of Account (**Cognizability**)
 - Store of Value (**Durability**, **Stability of Value / Scarcity**)



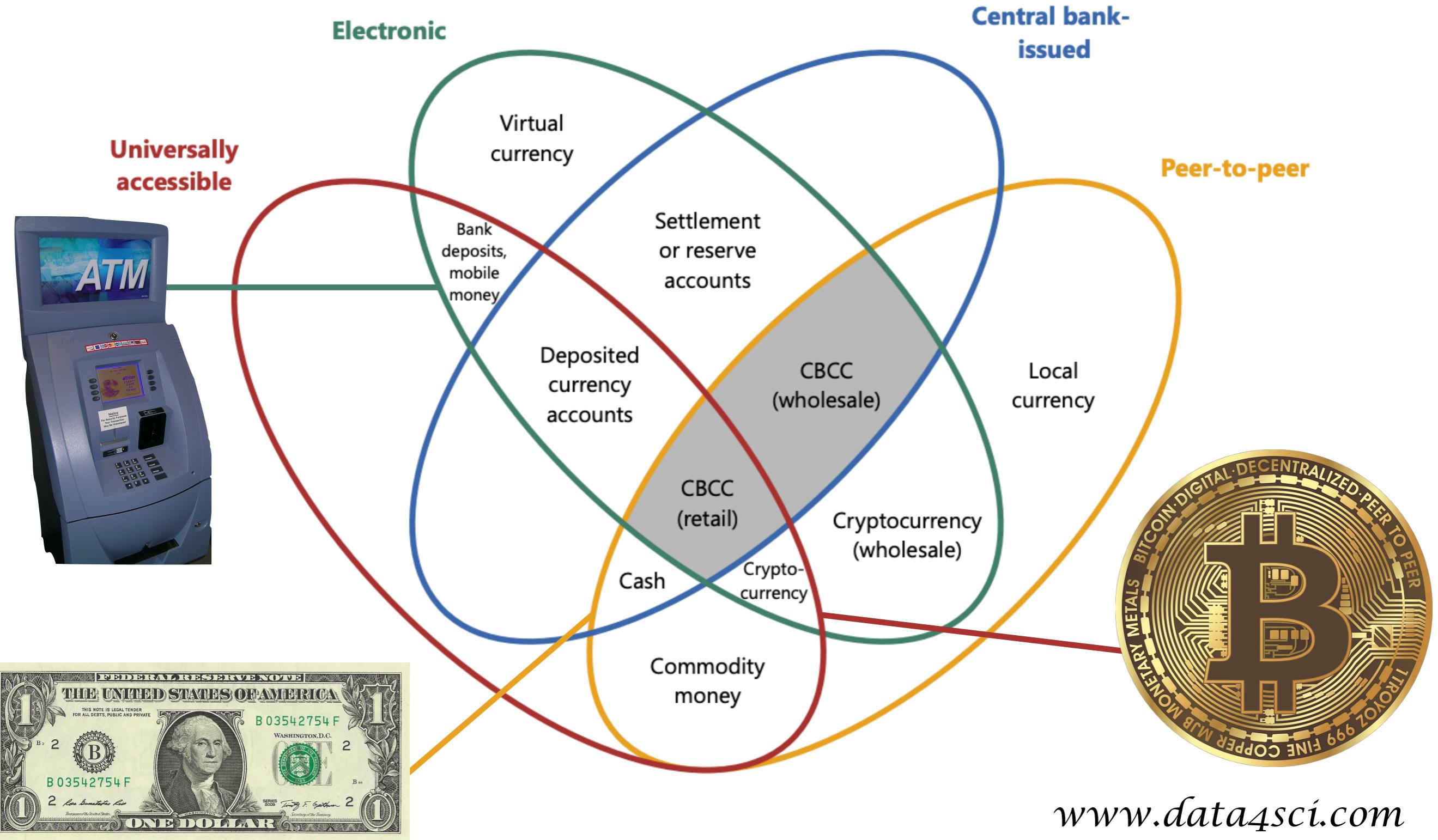
Money Flower

https://www.bis.org/publ/qtrpdf/r_qt1709.pdf

https://upload.wikimedia.org/wikipedia/commons/2/23/US_one_dollar_bill%2C_obverse%2C_series_2009.jpg

<https://upload.wikimedia.org/wikipedia/commons/thumb/d/d3/49024-SOS-ATM.JPG/1600px-49024-SOS-ATM.JPG>

https://pixabay.com/p-3125488/?no_redirect



The Ingredients of a Crypto-Currency

<https://pixabay.com/photos/bank-money-finance-shares-save-2907728/>

- Fundamental functions (and properties) of money:
 - Medium of Exchange (**Fungibility**, **Portability**)
 - Unit of Account (**Cognizability**)
 - Store of Value (**Durability**, **Stability of Value / Scarcity**)
- Digital currencies have obvious advantages:
 - **Fungibility** - A bit is a bit
 - **Portability** - We can easily carry terabytes in our pockets or digitally transfer them across the world
 - **Cognizability** - It's just a number
 - **Durability** - Information can easily be stored for decades or even centuries
- And obvious problems:
 - **Stability of Value / Scarcity**
 - There is no limit to how many you can create
 - They can be easily duplicated (no trust)
 - etc...



The Ingredients of a Crypto-Currency

<https://pixabay.com/photos/bank-money-finance-shares-save-2907728/>

<https://news.bitcoin.com/10-years-ago-bitcoins-genesis-block-changed-the-course-of-history/>

- How to avoid the need for a centralized authority?

- Distributed Ledger

- How to avoid double spending?

- Consensus algorithm

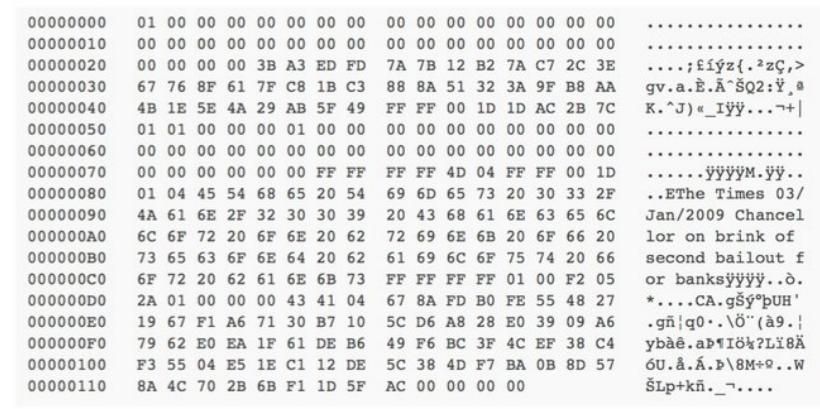
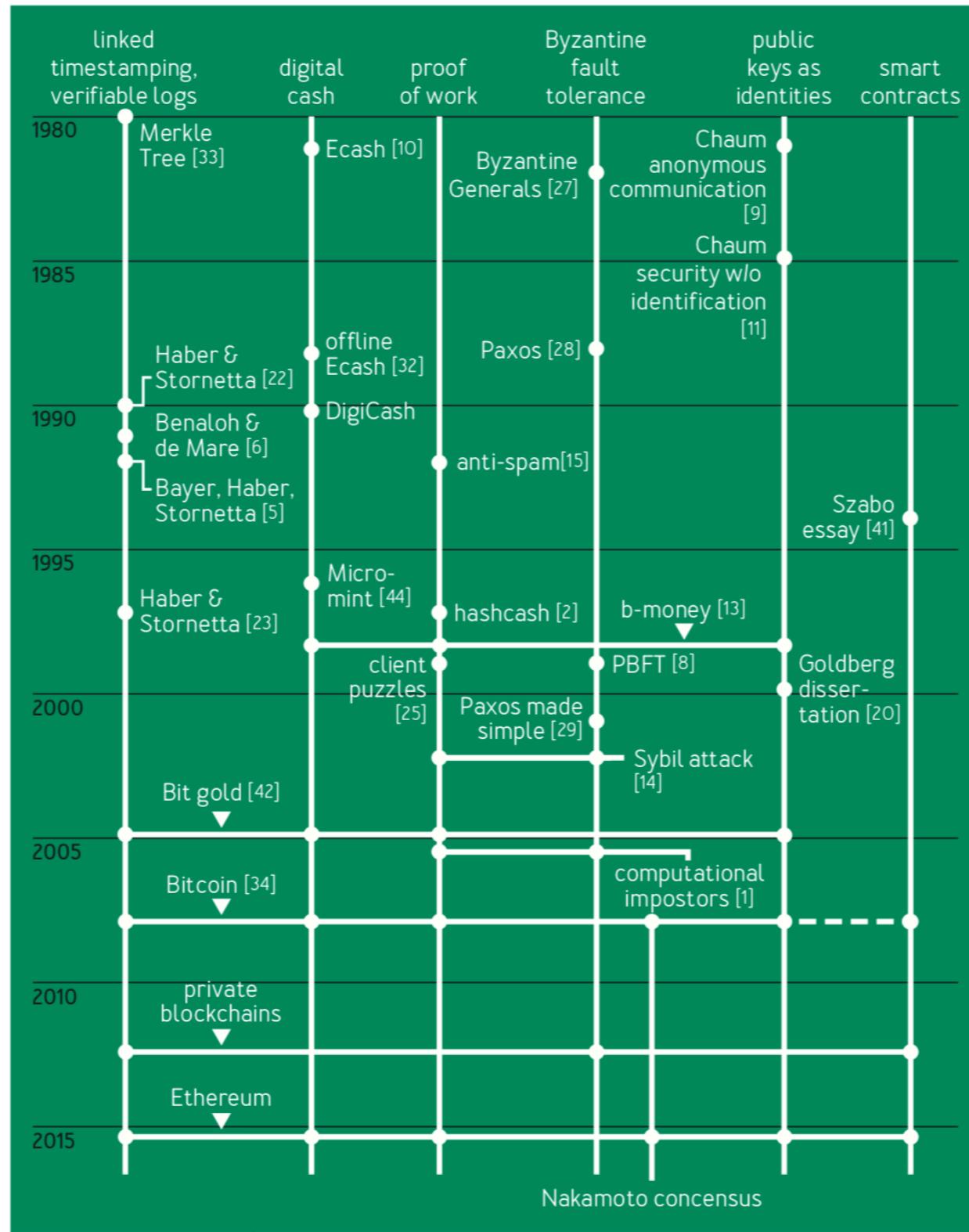
- How to prevent cheaters?

- Reward miners for being honest



The road to Bitcoin

<https://queue.acm.org/detail.cfm?id=3136559>



How does it work?

<https://bitcoin.org/bitcoin.pdf>

- According to "Satoshi Nakamoto":

The steps to run the network are as follows:

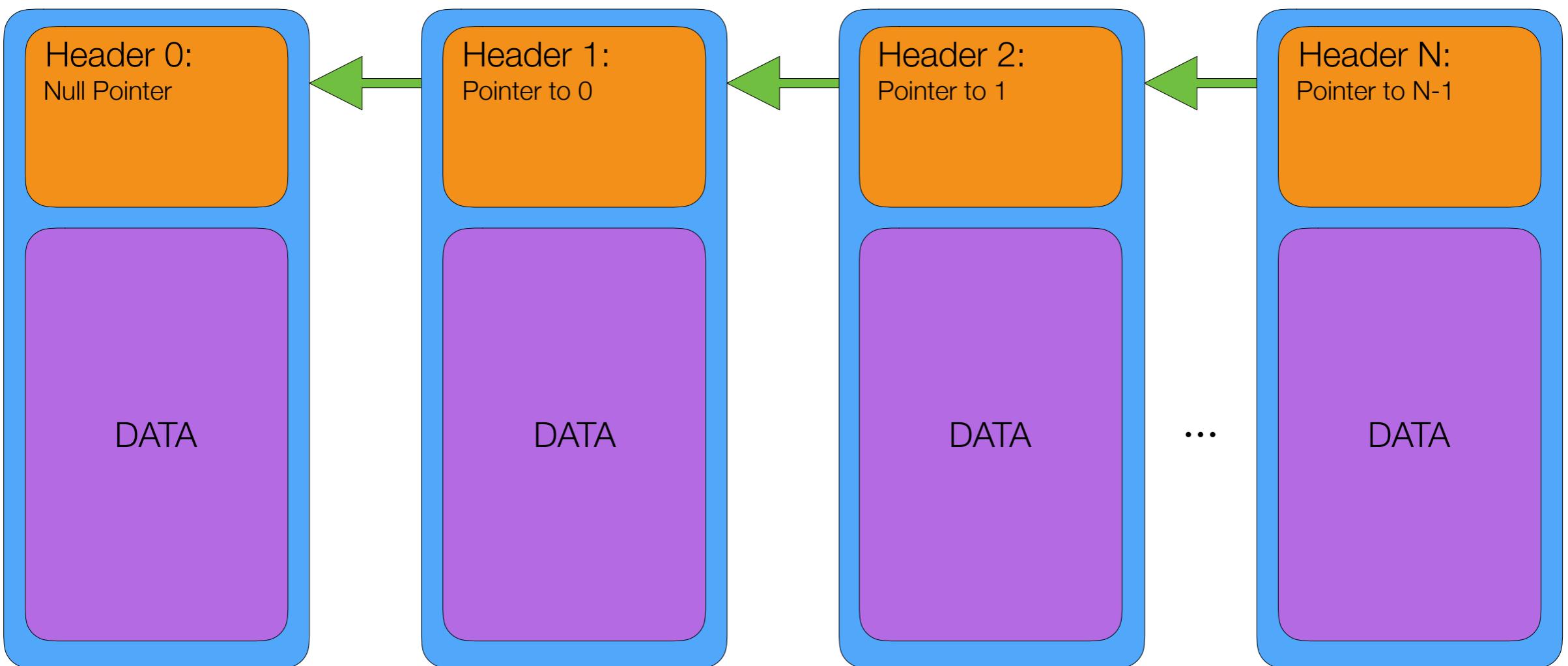
- 1) New transactions are broadcast to all nodes.
- 2) Each node collects new transactions into a block.
- 3) Each node works on finding a difficult proof-of-work for its block.
- 4) When a node finds a proof-of-work, it broadcasts the block to all nodes.
- 5) Nodes accept the block only if all transactions in it are valid and not already spent.
- 6) Nodes express their acceptance of the block by working on creating the next block in the chain, using the hash of the accepted block as the previous hash.



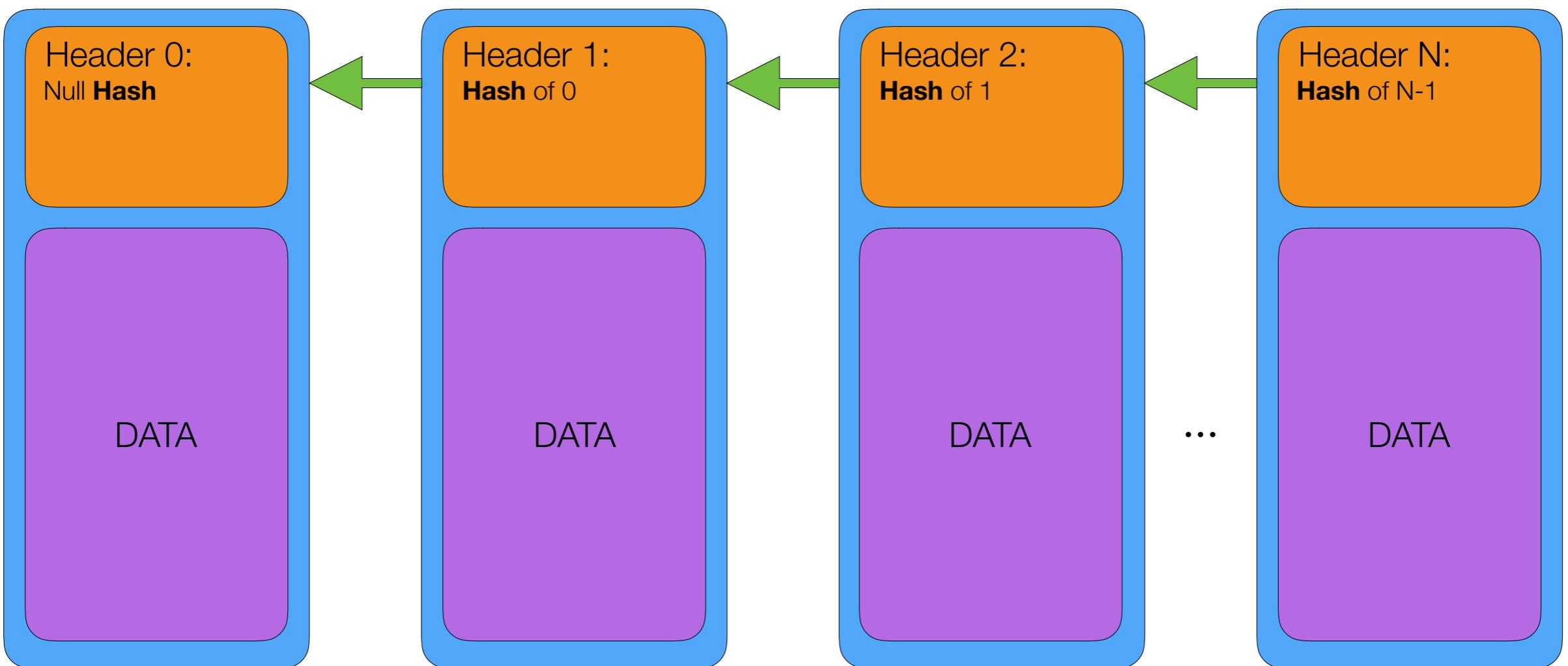
- Essentially:

- Transactions are processed in blocks
- Miners verify the transactions in each block [Mining Reward]
- Blocks are cryptographically signed to prevent tampering [Proof of Work]
- Each block points to the previous block to form a chain
- The longest chain is defined as the correct one [Consensus]

Linked List

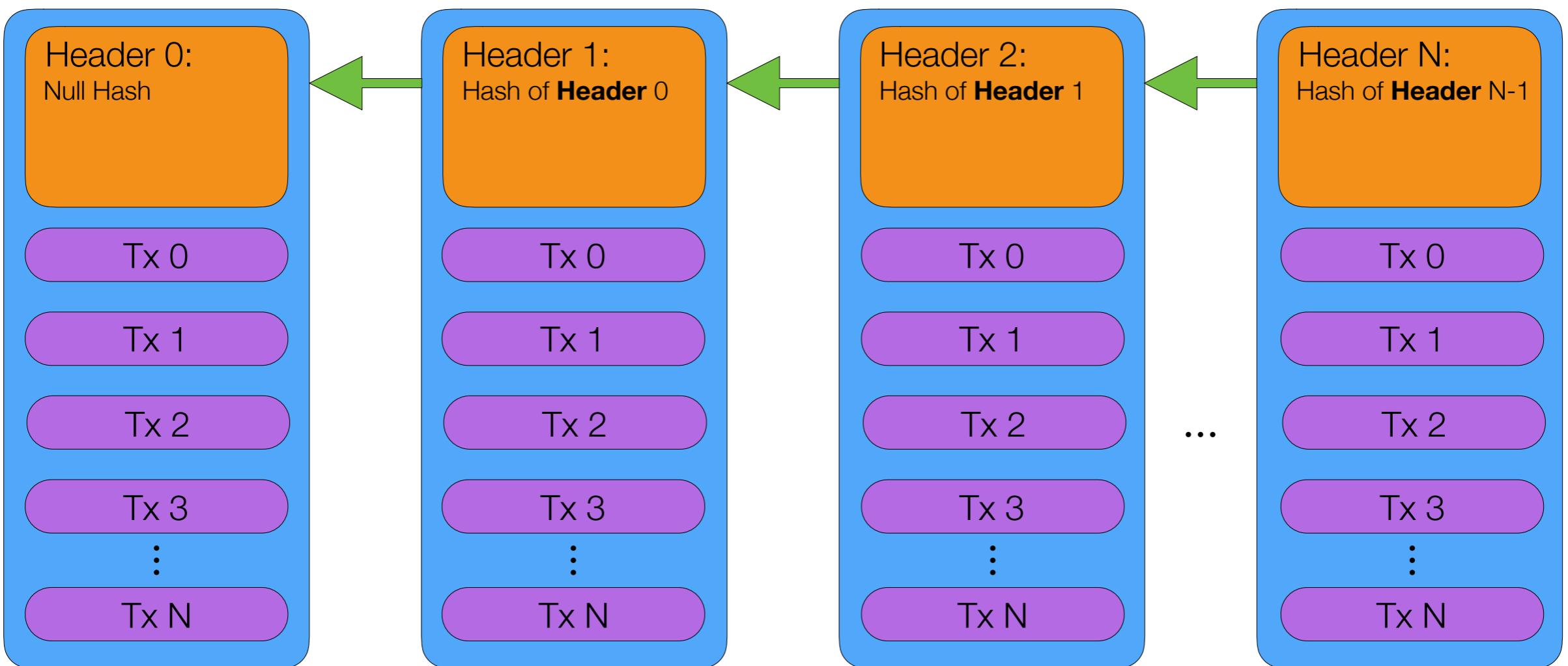


Blockchain



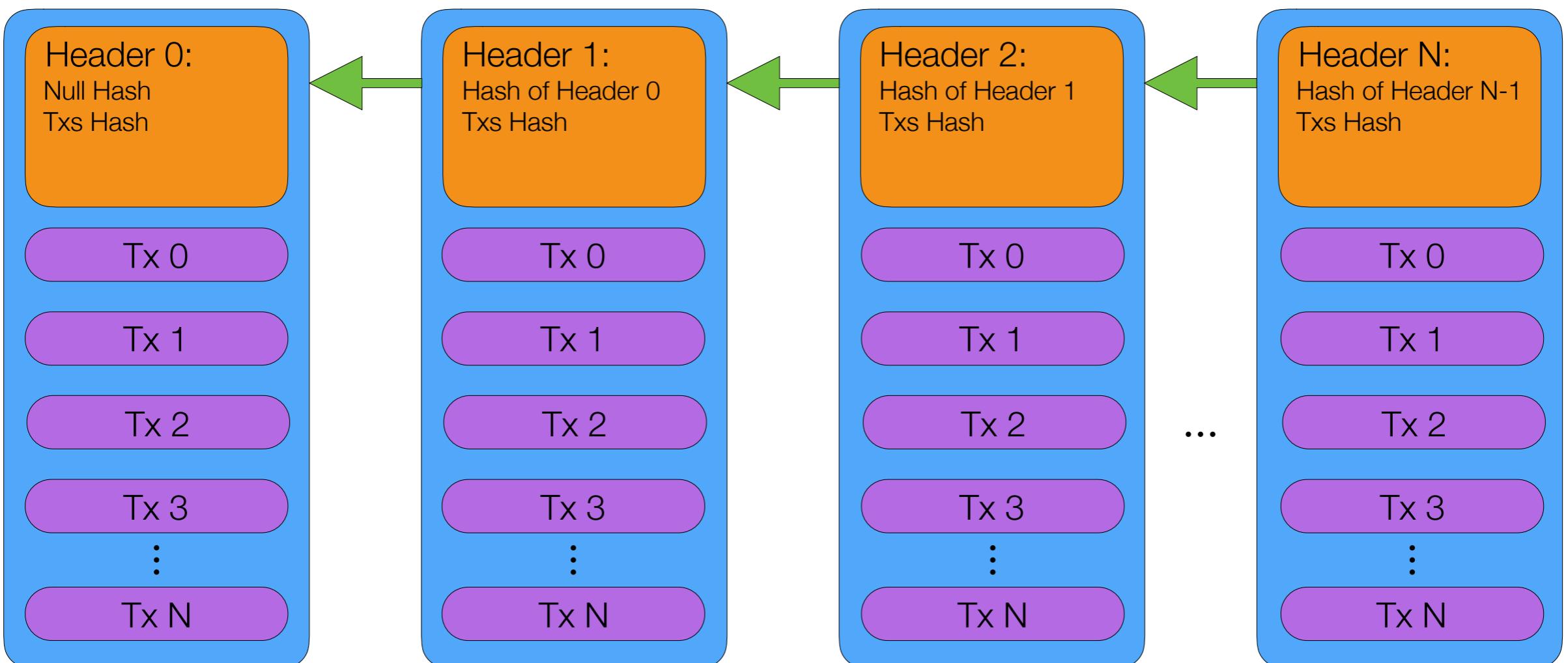
- The hashing function ensures that any changes to a validated previous block invalidates it.

Blockchain



- The hashing function ensures that any changes to a validated previous block invalidates it.
- Keep hash only the previous header for efficiency

Blockchain



- The hashing function ensures that any changes to a validated previous block invalidates it.
- Keep hash only the previous header for efficiency
- Header must include hash of data

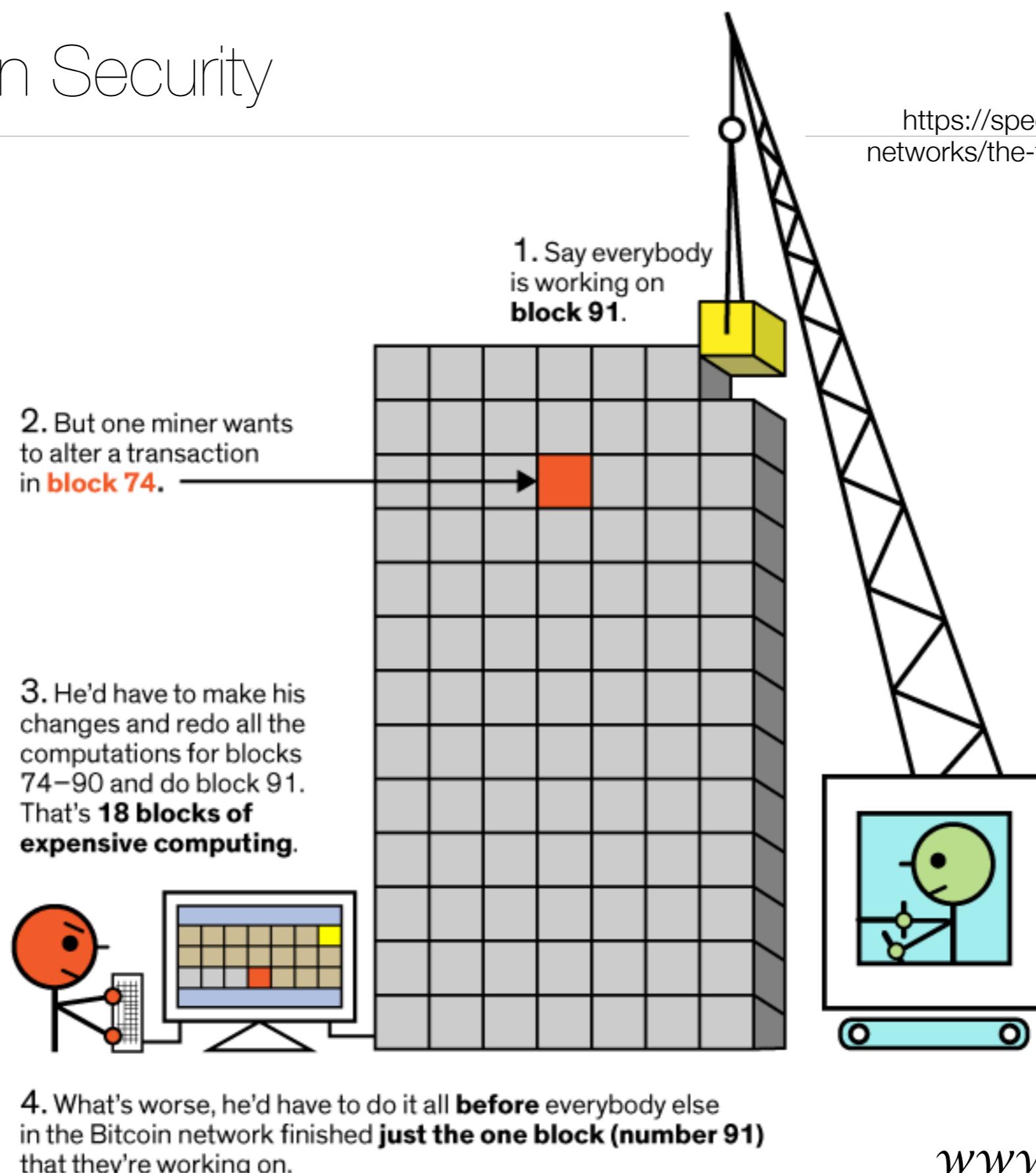
Proof Of Work

- Each block is identified by a hash value
- To control the rate at which blocks can be verified, and avoid malicious attacks the hash value of the block must obey certain conditions
- Miners add values to the data in the header (a **nonce**) until the hash fulfills these conditions
- The cryptographic properties of the hashing function guarantee that each value of the nonce is a roll of the dice (**fairness**)
- The “**difficulty**” level of the chain establishes how hard it is to obtain a valid hash
- Any change to the block data **invalidates** the hash
- The difficulty is adjusted every two weeks to account for the **hashing capability** of the network

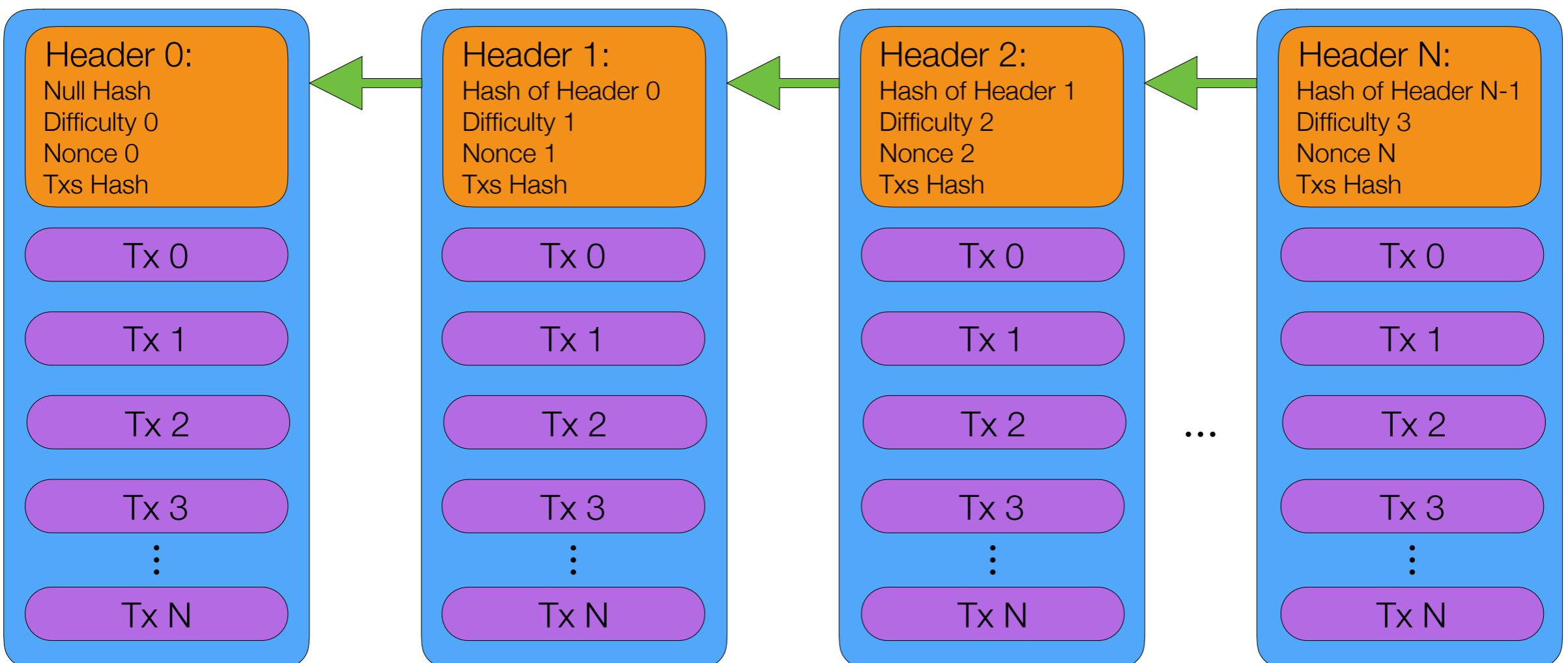


Blockchain Security

<https://spectrum.ieee.org/computing/networks/the-future-of-the-web-looks-a-lot-like-bitcoin>

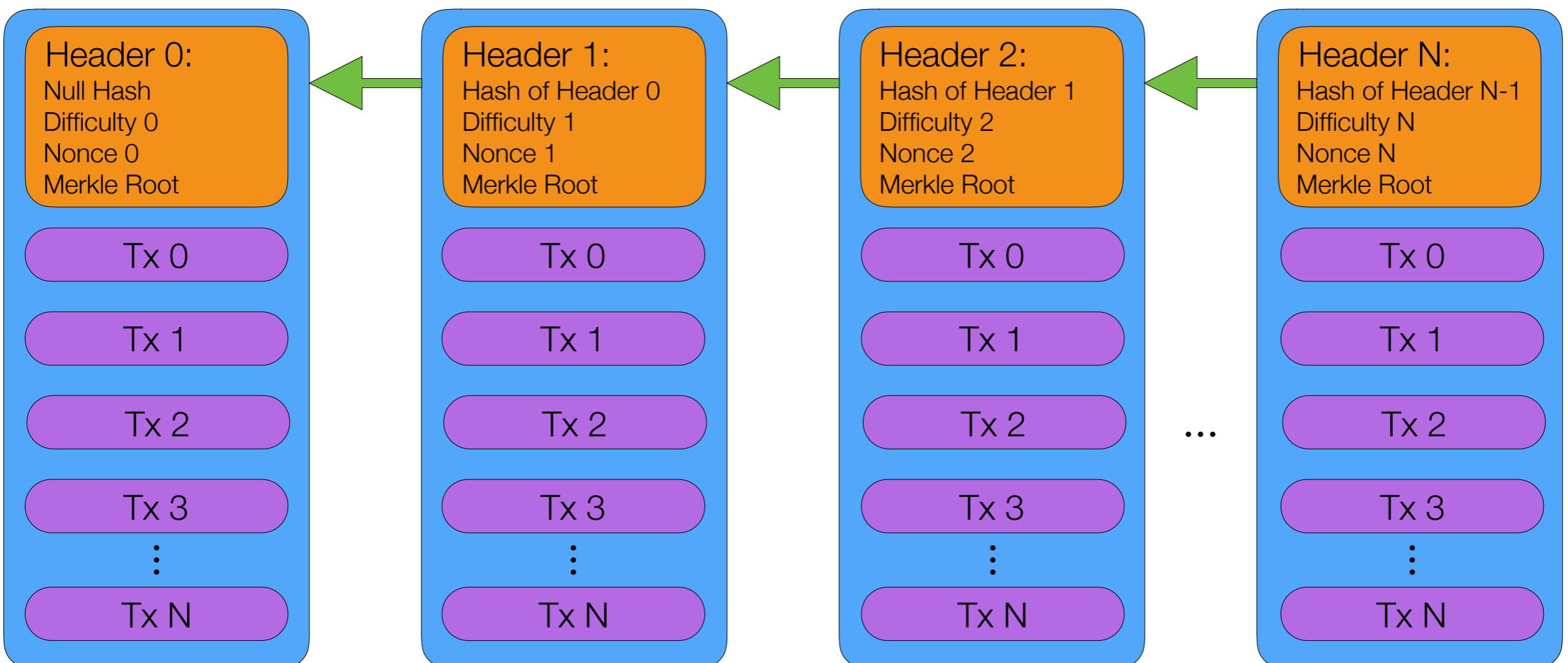


Blockchain



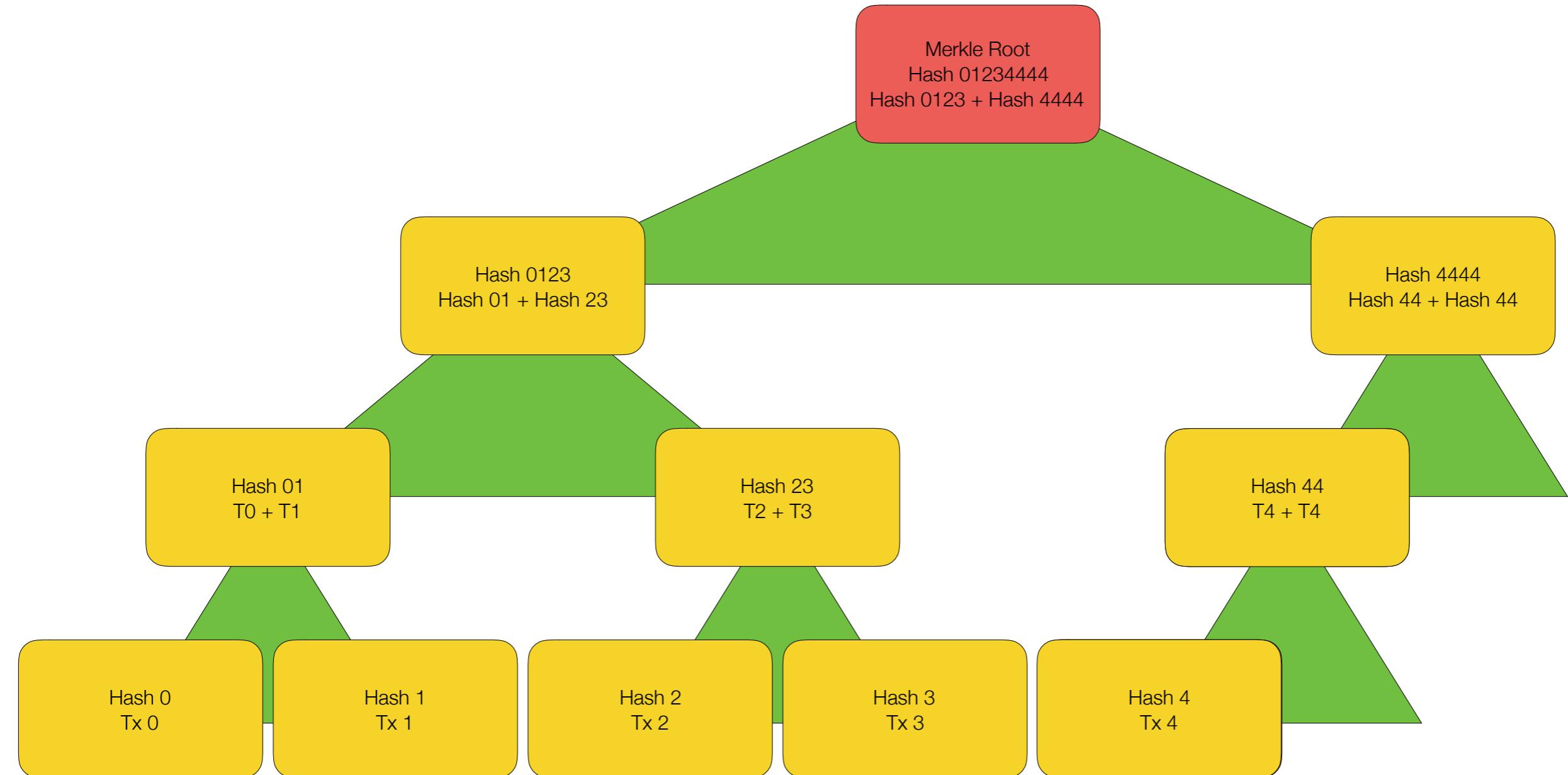
- The hashing function ensures that any changes to a validated previous block invalidates it.
- Keep hash only the previous header for efficiency
- Header must include hash of data

Blockchain



- The hashing function ensures that any changes to a validated previous block invalidates it.
- Keep hash only the previous header for efficiency
- Header must include hash of data

Merkle Tree



Merkle Tree

```
def get_root(transactions):
    missing = len(transactions)

    next_transactions = []

    while missing > 1:
        print(missing)
        next_transactions = []

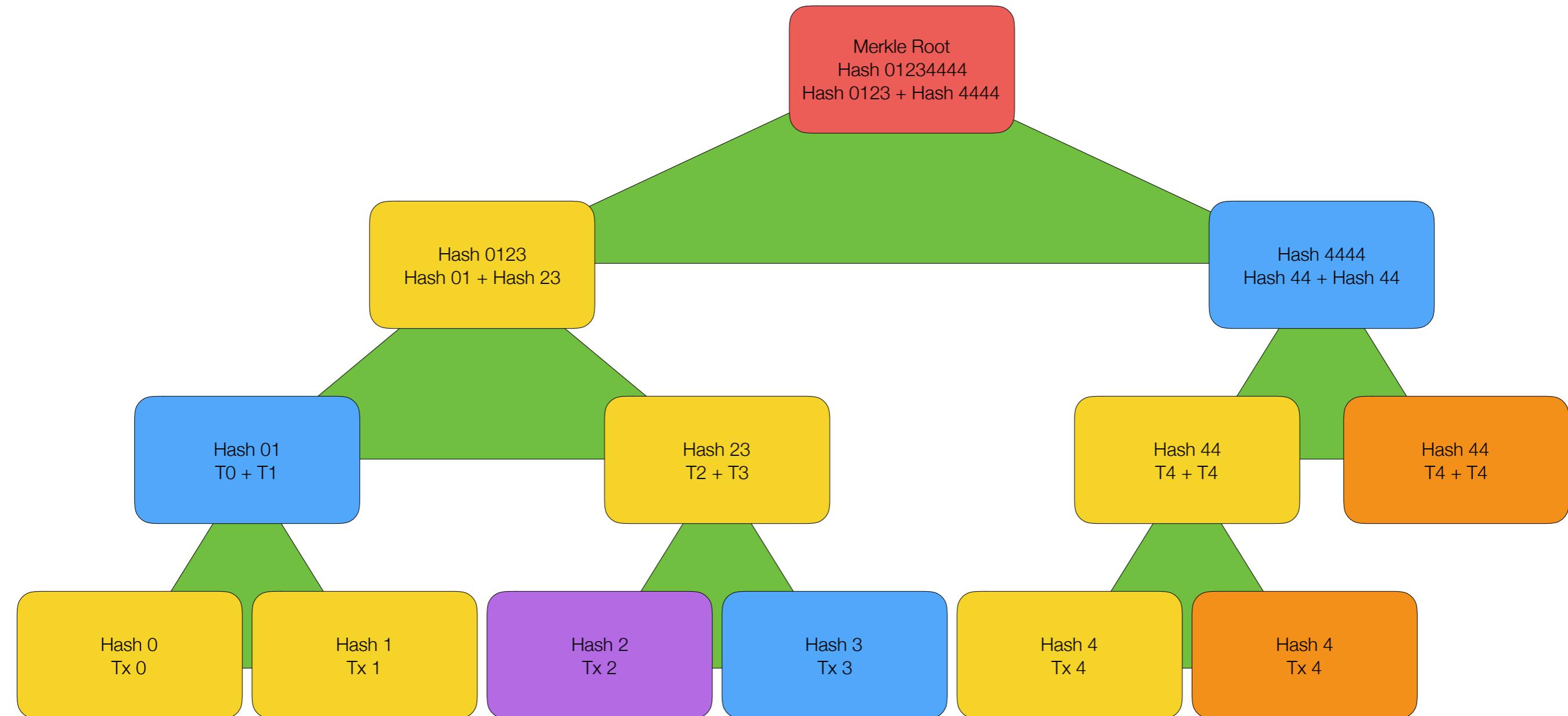
        # If it's odd, repeat the last transaction
        if missing % 2 == 1:
            transactions.append(transactions[-1])
            missing += 1

        for i in range(0, missing, 2):
            input = transactions[i] + transactions[i+1]
            output = doubleSha256_new(input)
            next_transactions.append(output)
            print(i, i+1, next_transactions[-1])

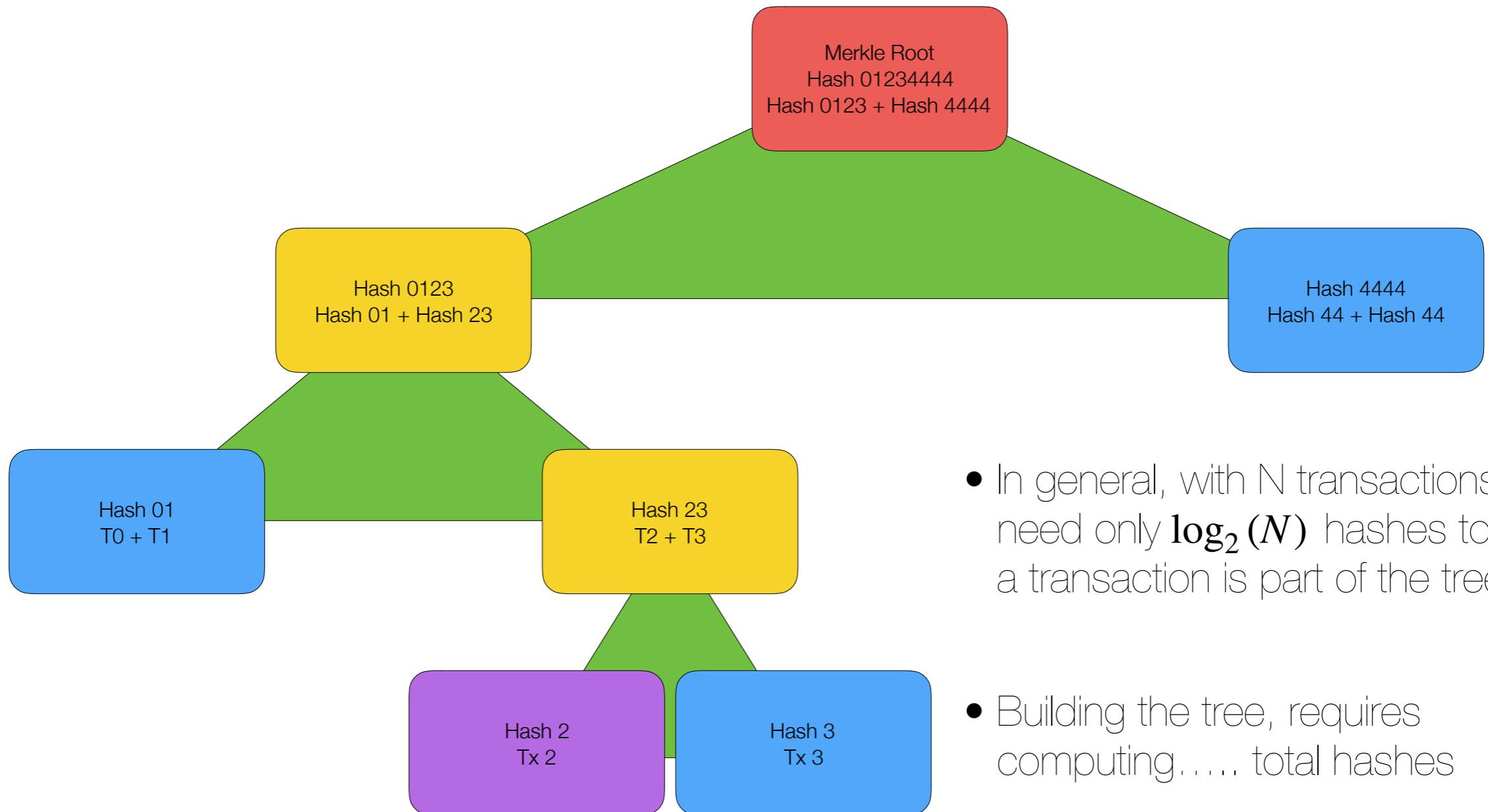
    missing = len(next_transactions)
    transactions = next_transactions[:]

    return invert_byteorder(next_transactions[0])
```

Merkle Proof - Tx 2



Merkle Proof - Tx 2



- In general, with N transactions you need only $\log_2(N)$ hashes to verify if a transaction is part of the tree.
- Building the tree, requires computing total hashes

Merkle Proof

```
def get_root(transactions):
    missing = len(transactions)

    next_transactions = []

    while missing > 1:
        print(missing)
        next_transactions = []

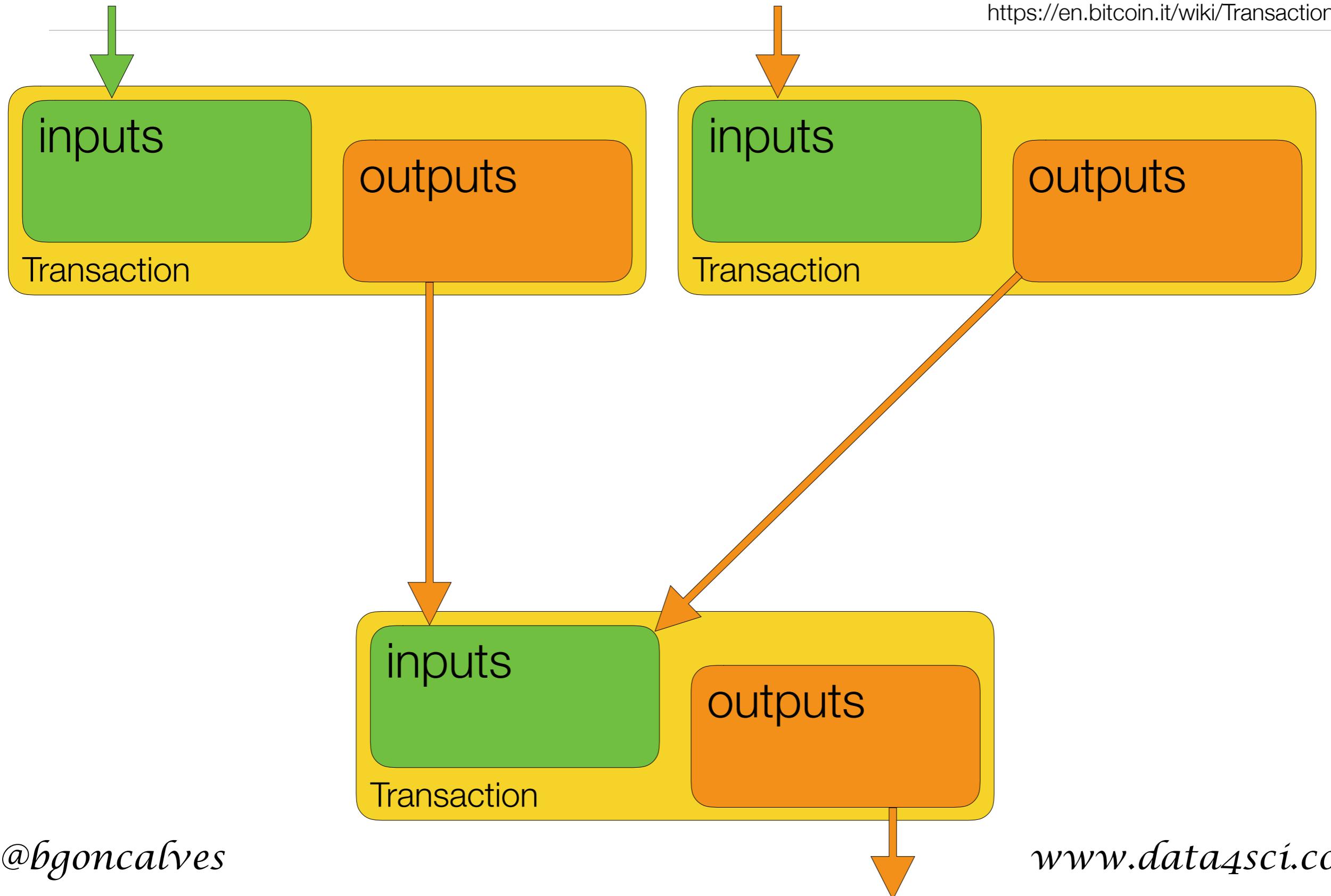
        # If it's odd, repeat the last transaction
        if missing % 2 == 1:
            transactions.append(transactions[-1])
            missing += 1

        for i in range(0, missing, 2):
            input = transactions[i] + transactions[i+1]
            output = doubleSha256_new(input)
            next_transactions.append(output)
            print(i, i+1, next_transactions[-1])

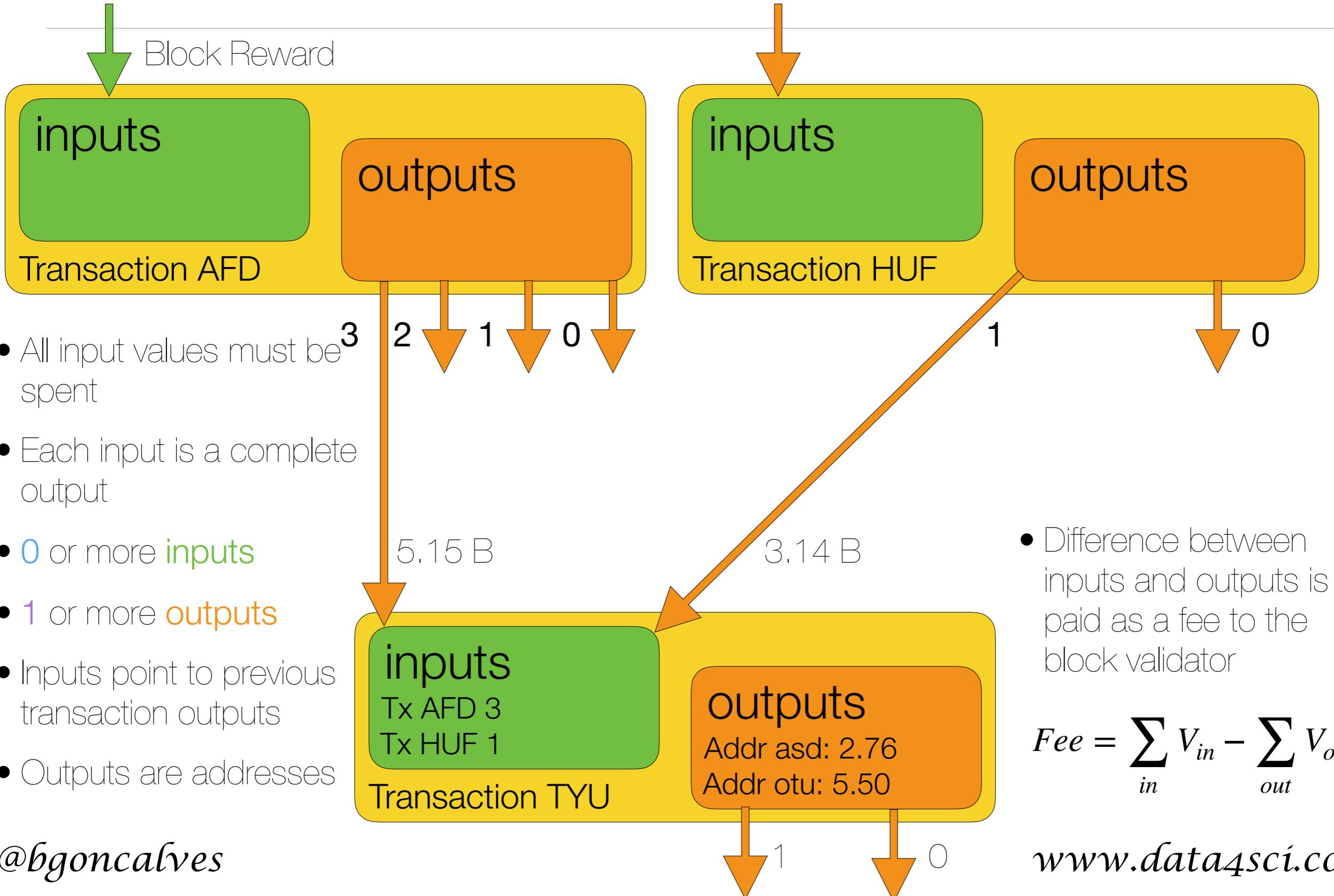
    missing = len(next_transactions)
    transactions = next_transactions[:]

    return invert_byteorder(next_transactions[0])
```

Transactions



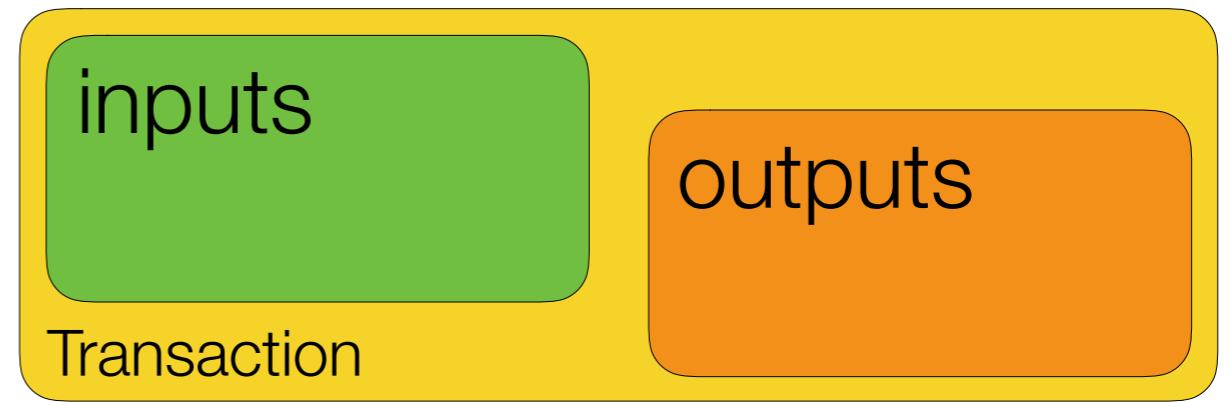
Transactions



Transactions

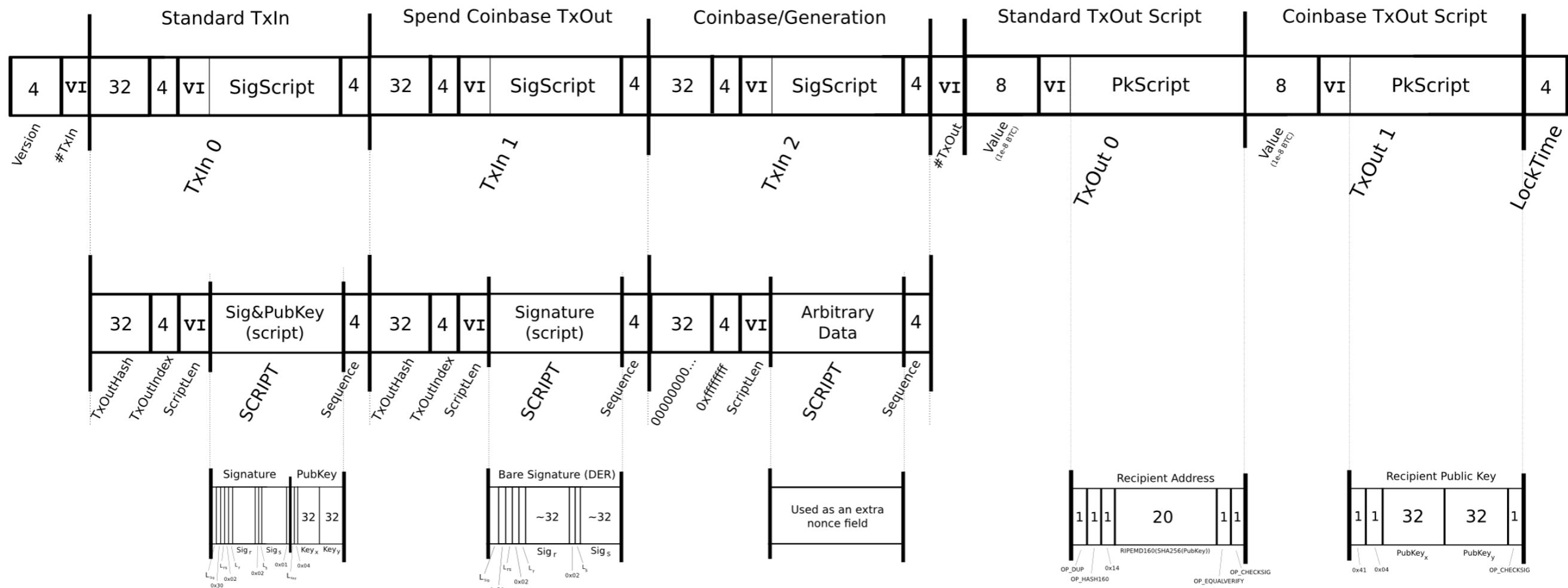
<https://en.bitcoin.it/wiki/Transaction>

- All input values must be spent
- Each input is a complete output
- 0 or more **inputs**
- 1 or more **outputs**
- Inputs point to previous transaction outputs
- Outputs are addresses
- First transaction in a block is the block reward and goes to the miners (**coinbase** transaction)
- **coinbase** transactions include all fees for the transactions included in the block
- Base block reward is how new **Bitcoins** are produced



Transactions

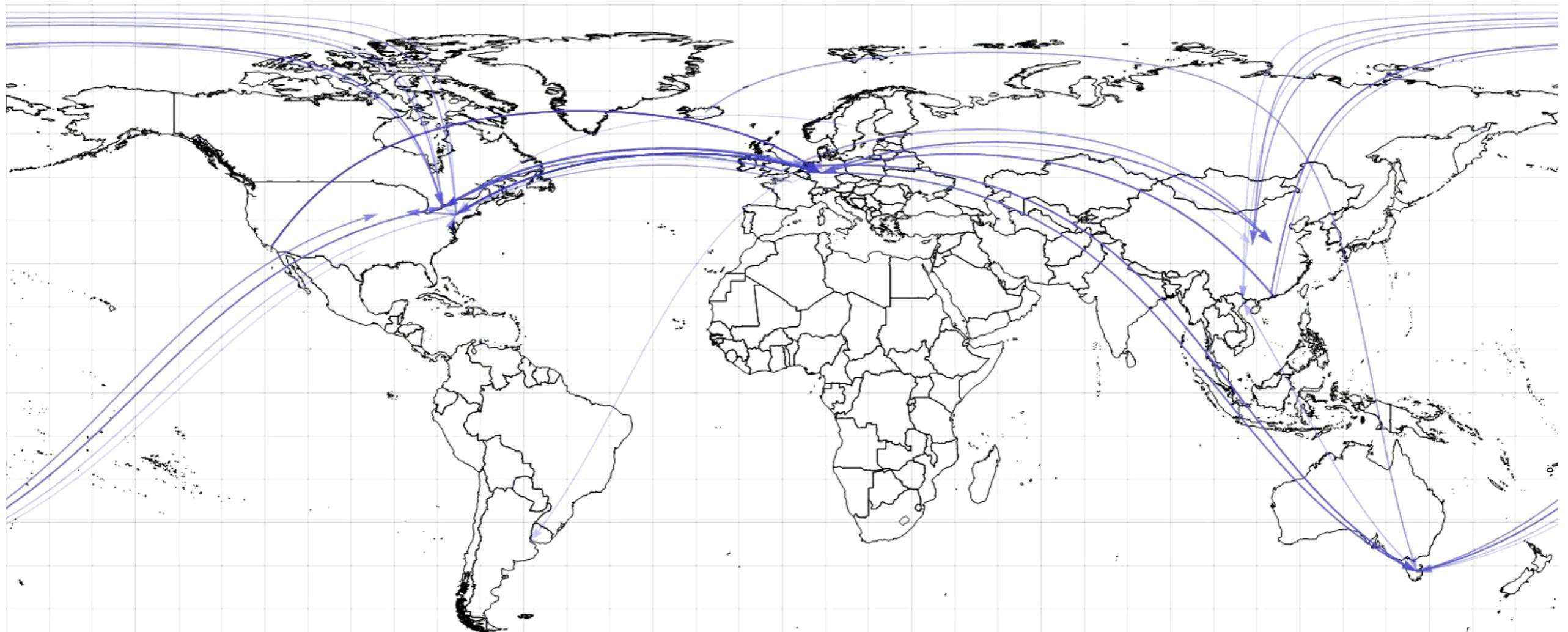
<https://en.bitcoin.it/wiki/Transaction>



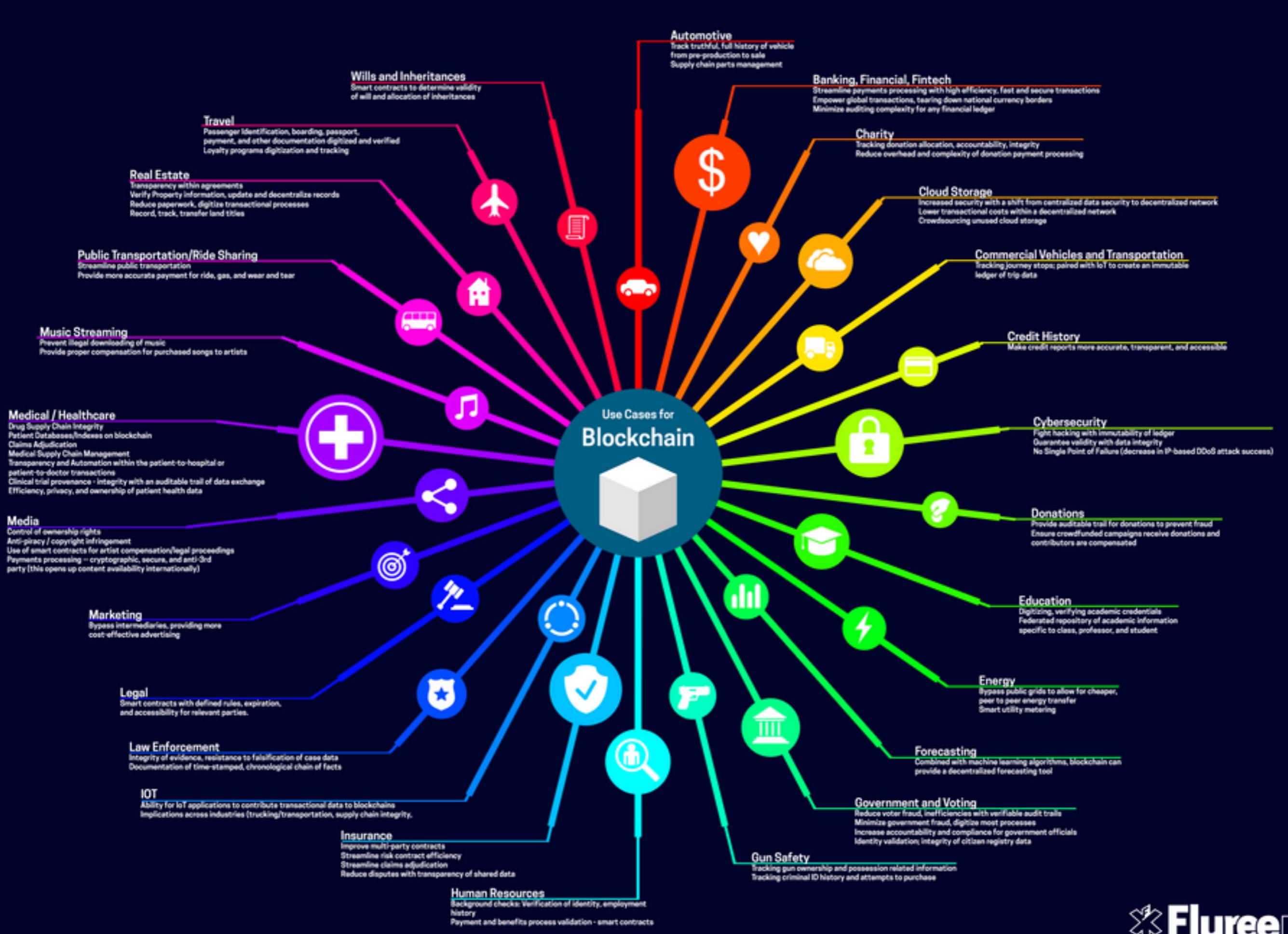
etotheipi@gmail.com / 1Gffm7LKXcNPrtxy6yF4JBoe5rVka4sn1

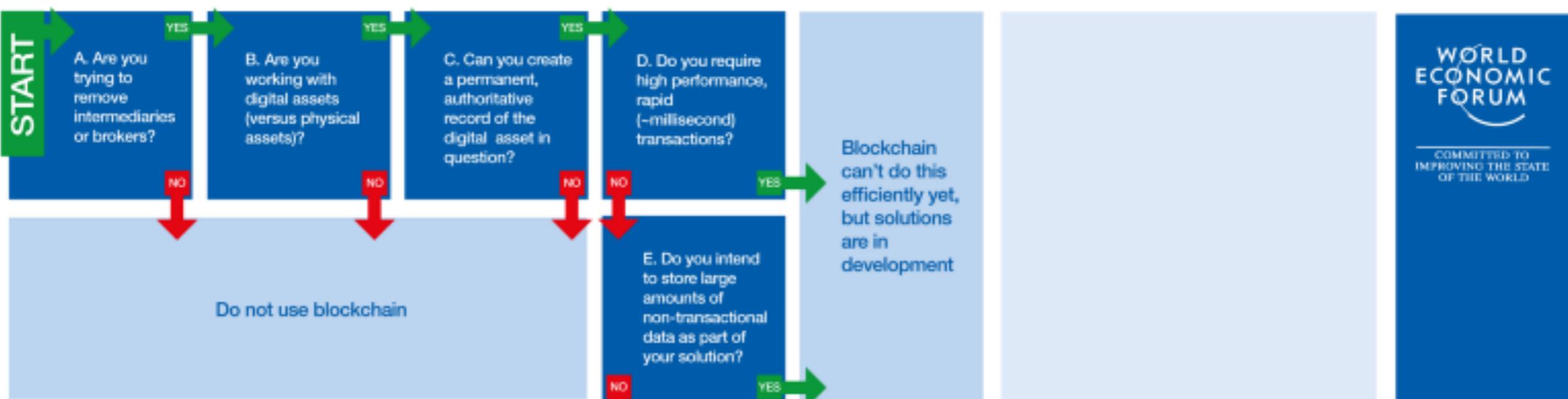
Bitcoin transactions around the world

<http://www.vo.elte.hu/papers/2017/bitcoin/>



11/25/2013

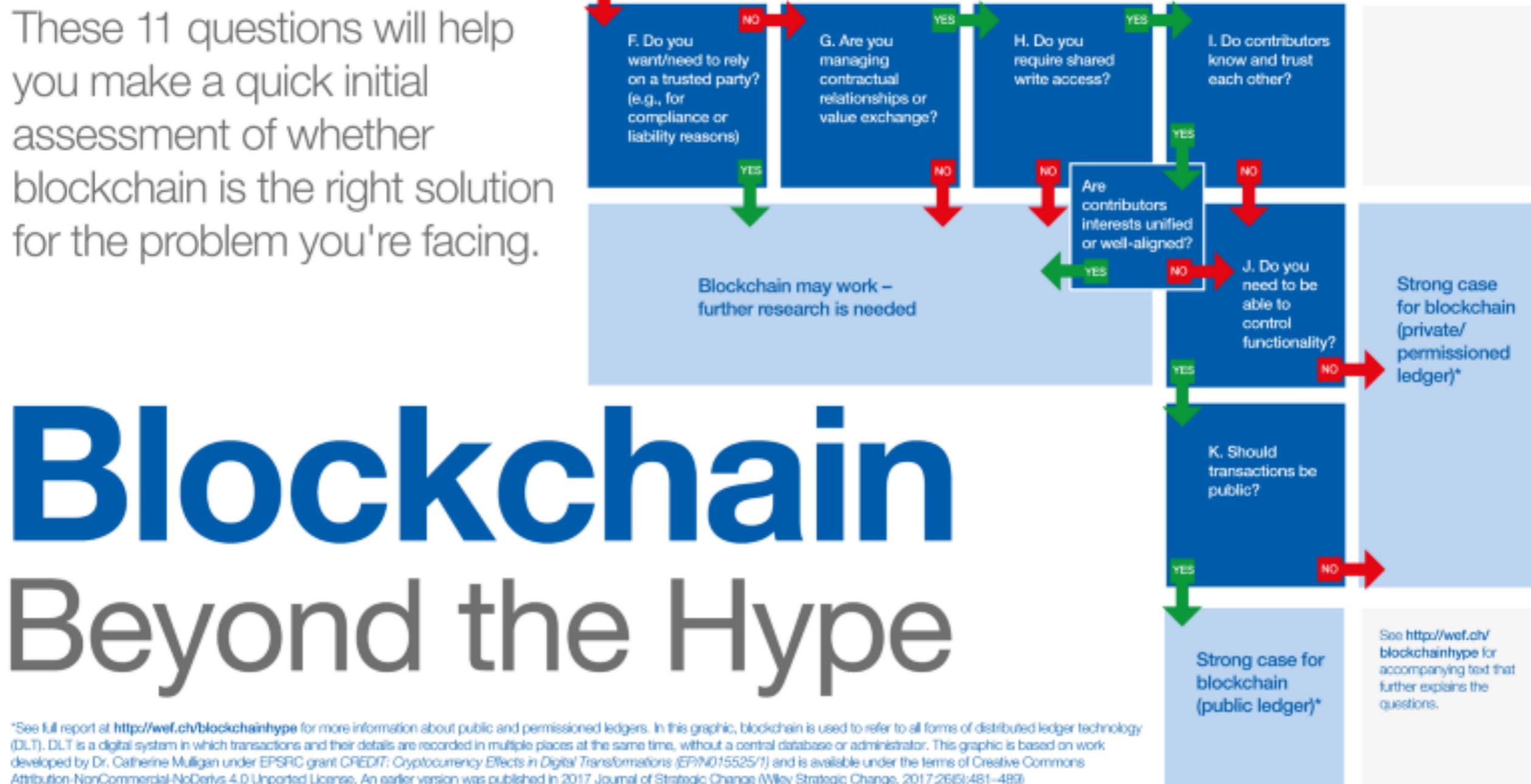




These 11 questions will help you make a quick initial assessment of whether blockchain is the right solution for the problem you're facing.

Blockchain Beyond the Hype

*See full report at <http://wef.ch/blockchainhype> for more information about public and permissioned ledgers. In this graphic, blockchain is used to refer to all forms of distributed ledger technology (DLT). DLT is a digital system in which transactions and their details are recorded in multiple places at the same time, without a central database or administrator. This graphic is based on work developed by Dr. Catherine Mulligan under EPSRC grant CREDIT: Cryptocurrency Effects in Digital Transformations (EP/N015525/1) and is available under the terms of Creative Commons Attribution-NonCommercial-NoDerivs 4.0 Unported License. An earlier version was published in 2017 Journal of Strategic Change (Wiley Strategic Change, 2017;26(5):481–489).



BTC

BTC

BTC is the original cryptocurrency, and was released as open-source software in 2009. Using a new distributed ledger known as the blockchain, the Bitcoin protocol allows for users to make peer-to-peer transactions using digital currency while avoiding the "double spending" problem.

No central authority or server verifies transactions, and instead the legitimacy of a payment is determined by the decentralized network itself.

KEY FEATURES

- Blockchain - Foundational Technology
- Fast P2P Payments Worldwide
- No Double Spend Problem
- Low Processing Fees
- Decentralized
- Available To Anyone
- Anonymity (Partial)
- Transparent

INTERESTING FACT

In 2010, a programmer bought two pizzas for 10,000 BTC in one of the first real-world bitcoin transactions. Today, 10,000 BTC is equal to roughly \$381 million - a big price to pay for satisfying hunger pangs.

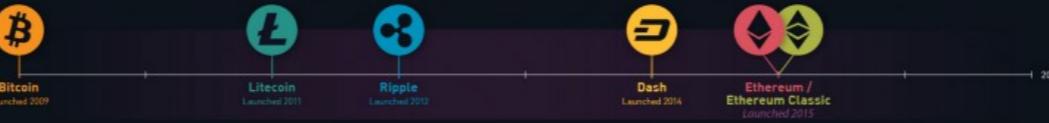
While regrettable, it is also forgivable. No one knew Bitcoin would be accepted by hundreds of thousands of merchants just years later. Now cryptocurrency enthusiasts around the world celebrate "Bitcoin Pizza Day" every year on May 22nd, the date the transaction took place.

BOTTOM LINE FOR BITCOIN

Bitcoin is the original cryptocurrency with the most liquidity and significant network effects - it also has brand name recognition around the world, with an eight year track record.

THE CRYPTO UNIVERSE

COMPARING SIX MAJOR CURRENCIES



DASH

DASH

DASH is an attempt to improve on Bitcoin in two main areas: speed of transactions, and anonymity.

To do this, it has a two-tier architecture with miners and also "masternodes" that help the network perform advanced functions such as near-instant transactions and coin mixing to provide additional privacy.

KEY DIFFERENTIATIONS FROM BITCOIN

- Two-Tier Architecture
- Advanced Transactions
- Decentralized Autonomous Organization (DAO)
- Improved Anonymity

INTERESTING FACT

For each block mined, a 10% reward goes to the treasury to help fund improvements and projects around Dash.

Dash is the first ever Decentralized Autonomous Organization (DAO), a type of organization run through rules encoded in computer programs and smart contracts.

BOTTOM LINE FOR DASH

The innovations behind Dash are interesting, and could help to make the coin more consumer-friendly than other alternatives.

ETHEREUM

ETHEREUM

ETHEREUM is an open software platform based on blockchain technology that enables developers to build and deploy decentralized applications.

In the Ethereum blockchain, instead of mining for bitcoin, miners work to earn ether, a type of crypto token that fuels the network. Beyond a tradable cryptocurrency, ether is also used by application developers to pay for transaction fees and services on the Ethereum network.

KEY DIFFERENTIATIONS FROM BITCOIN

- Platform For Making Blockchain Applications
- Multiple Industry Uses
- Uses Smart Contracts
- Ether Powers The Network

INTERESTING FACT

Ethereum has quickly skyrocketed in value since its introduction in 2015, and it is now the 2nd most valuable cryptocurrency by market cap.

It's increased in value by 2,230% in just the last year - a huge boom for early investors.

BOTTOM LINE FOR ETHERUM

Ethereum serves a different purpose than other cryptocurrencies, but it's quickly grown to displace all but Bitcoin in value. Some experts are so bullish on Ethereum that they even see it becoming the world's top cryptocurrency in just a short span of time - but only time will tell.

ETHEREUM CLASSIC

ETHEREUM CLASSIC

ETHEREUM CLASSIC In 2016, the Ethereum community faced a difficult decision: The DAO, a venture capital firm built on top of the Ethereum platform, had \$50 million in ether stolen from it through a security vulnerability.

The majority of the Ethereum community decided to help the DAO by "hard forking" the currency, and then changing the blockchain to return the stolen proceeds back to The DAO.

The minority thought this idea violated the key foundation of immutability that the blockchain was designed around, and kept the original Ethereum blockchain the way it was. Hence, the "Classic" label.

KEY DIFFERENTIATIONS FROM BITCOIN

- Platform For Making Blockchain Applications
- Multiple Industry Uses
- Uses Smart Contracts
- Ether Powers The Network

INTERESTING FACT

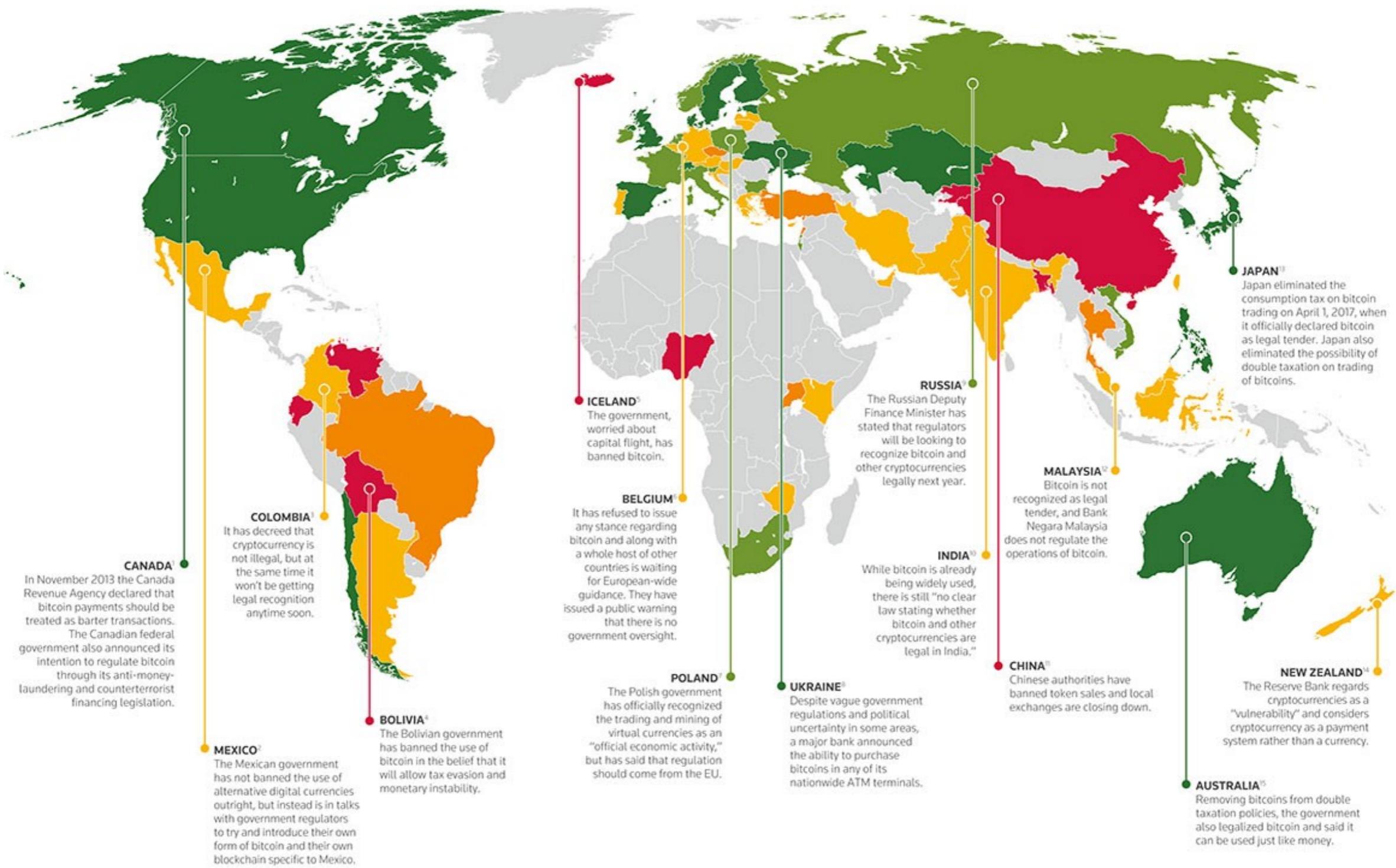
Because Ethereum and Ethereum Classic stem from the same code, they run in parallel, doing virtually the same thing.

The DAO hacker still holds over 3 million of Ethereum Classic, which could be "dumped" somewhere along the line - a baked-in risk that could drag on price.

BOTTOM LINE FOR ETHERUM CLASSIC

As time goes on, Ethereum Classic has been carving out a separate identity from its bigger sibling. With similar capabilities and a different set of principles, Ethereum Classic could still have upside.

A WORLD OF CRYPTOCURRENCIES



1. <https://www.fxempire.com/education/article/what-does-the-future-of-bitcoin-look-like-404636>

2. <https://www.cryptocompare.comcoins/guides/how-legal-is-bitcoin-and-crypto-currencies>

3. <http://esidebitcoins.com/news/colombia-clarifies-stance-bitcoin-is-not-illegal-60667>

4. <https://www.cryptocompare.comcoins/guides/how-legal-is-bitcoin-and-crypto-currencies>

5. <https://themerkle.com/bitcoins-legal-status-worldwide>

6. <https://www.cryptocompare.comcoins/guides/how-legal-is-bitcoin-and-crypto-currencies>

7. <https://cointelegraph.com/news/poland-officially-recognizes-trading-in-bitcoin-and-other-cryptocurrencies>

8. <https://www.forbes.com/sites/realspin/2017/03/20/ukraine-is-silently-leading-a-digital-currency-revolution/#f7e902d53465c>

9. <https://www.fxempire.com/education/article/what-does-the-future-of-bitcoin-look-like-404636>

10. <https://news.bitcoin.com/india-supreme-court-bitcoin-legalization-taxation>

11. <https://qz.com/1079900/huobi-and-bitcoin-chinas-two-biggest-bitcoin-exchanges-will-halt-all-trading-services-for-local-customers/>

12. <https://www.mta.com.my/opinion/columnists/2017/06/24/0010/are-we-ready-bitcoin>

13. <https://news.bitcoin.com/india-supreme-court-bitcoin-legalization-taxation>

14. <https://www.stuff.co.nz/business/93087302/cryptocurrency-exchange-offers-more-than-bitcoin-for-new-zealand-dollars>

15. <https://news.bitcoin.com/india-supreme-court-bitcoin-legalization-taxation>

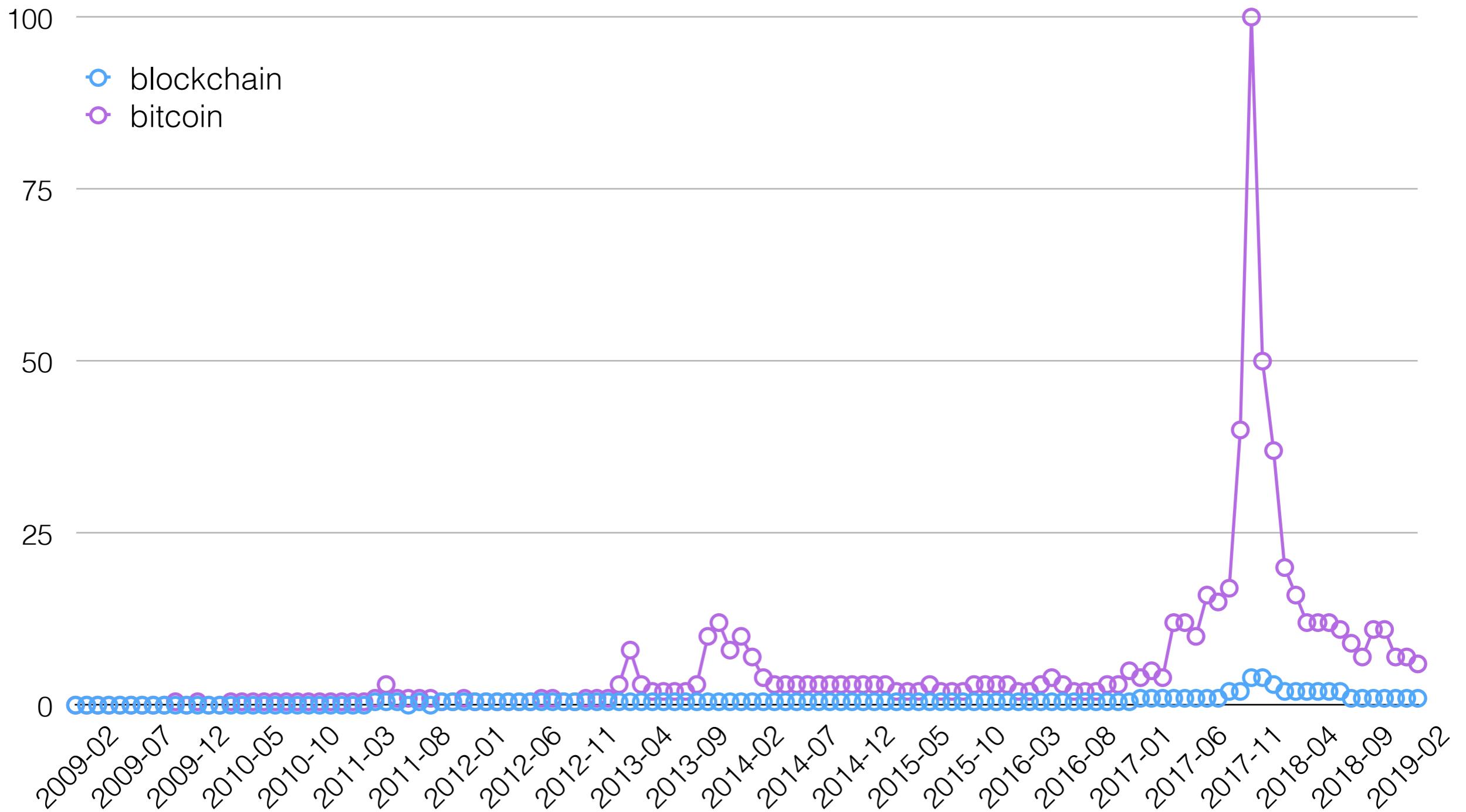


Read about all of the countries at
thomsonreuters.com/know360app

How much is it worth?



Bitcoin interest over time

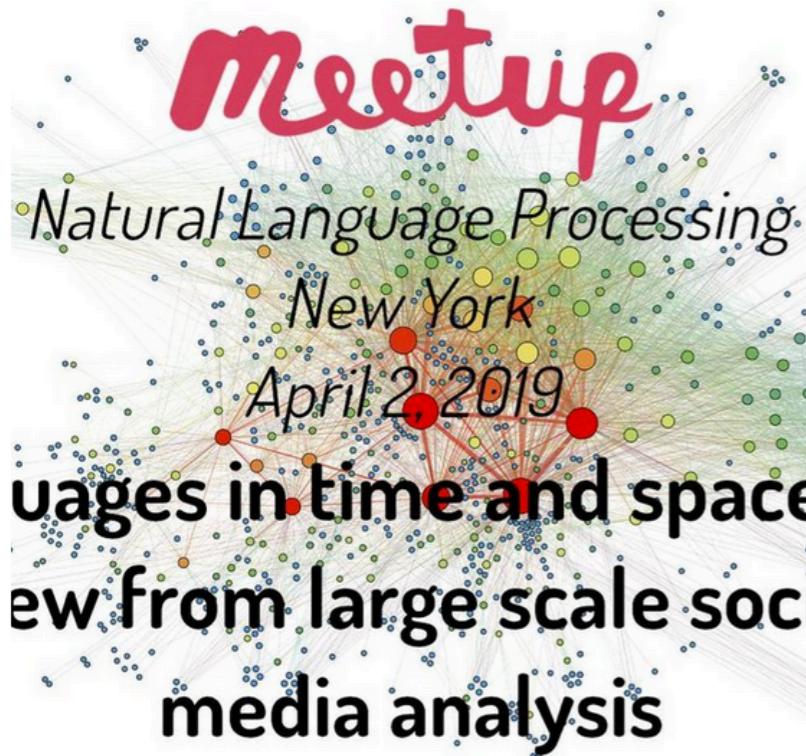


Bitcoin price





Questions?

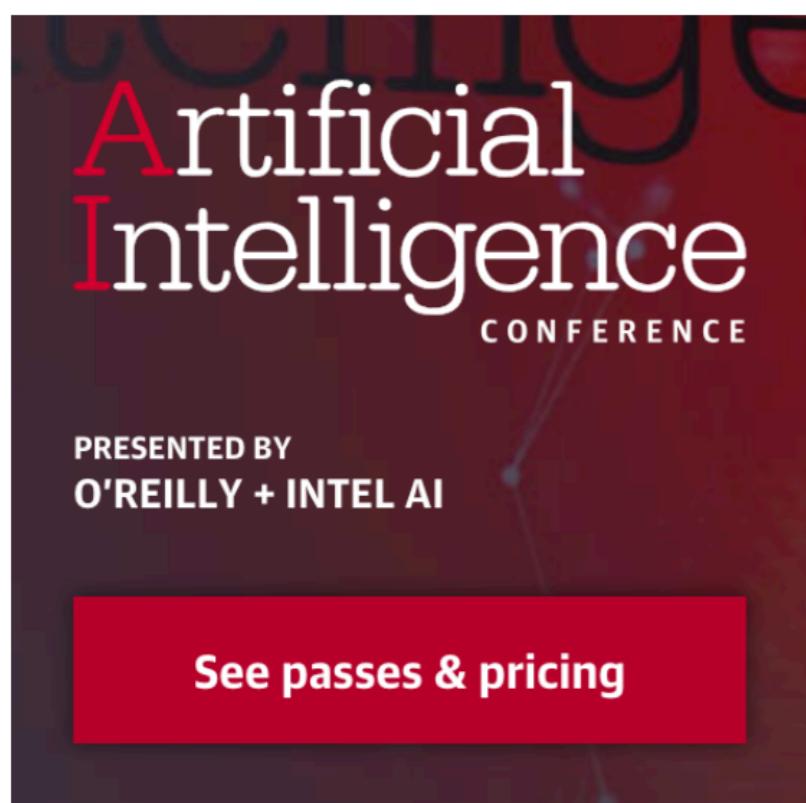


Languages in time and space: the view from large scale social media analysis
Apr 2, 2019

The advent of online social services coupled with GPS-enabled smartphones resulted in the accumulation of massive amounts of data documenting our individual and social behavior. Using sources such as Twitter, Wikipedia, Google Books and others, this talk will present how languages are used across both time and space. In particular, we will analyze how language dialects can be defined based on how they are used in the real world. We will also analyze how English usage changes from place to place and over time and how languages can be used to identify communities within the urban environment.

[REGISTER](#)

[LEARN MORE](#)



Recurrent neural networks for time series analysis
Apr 16, 2019

Time series are everywhere around us. Understanding them requires taking into account the sequence of values seen in previous steps and even long-term temporal correlations. Join Bruno Gonçalves to learn how to use recurrent neural networks to model and forecast time series and discover the advantages and disadvantages of recurrent neural networks with respect to more traditional approaches.

[REGISTER](#)

[LEARN MORE](#)