



# Stochastic Models for Blockchain Analysis

Pierre-O. Goffard

Université de Strasbourg  
[goffard@unistra.fr](mailto:goffard@unistra.fr)

16 mars 2025

# Agenda

- 1 Introduction
- 2 Consensus protocol
- 3 Stochastic models for blockchain analysis

# Agenda

## Introduction

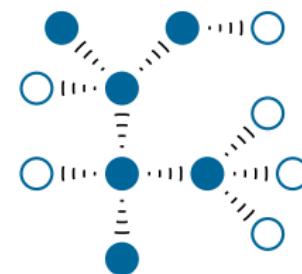
- 1 Introduction
- 2 Consensus protocol
- 3 Stochastic models for blockchain analysis

# Blockchain

## Introduction

A data ledger made of a sequence of blocks maintained by a achieving consensus in a Peer-To-Peer network.

- Decentralized
- Public/private
- Permissionned/permissionless
- Immutable
- Incentive compatible



# Consensus protocols

## Introduction

The mechanism to make all the nodes agree on a common data history.

The four dimensions of blockchain systems analysis

**1** Efficiency (Queueing theory)

- Throughputs
- Transaction confirmation time

**2** Decentralization (Entropy)

- Fair distribution of the accounting right

**3** Security (Fluctuation Theory)

- Resistance to attacks

**4** Incentive mechanism (Risk theory)

- Incentive for the nodes to contribute

# Applications of blockchain : Cryptocurrency

## Introduction



S. Nakamoto, "Bitcoin : A peer-to-peer electronic cash system." Available at <https://bitcoin.org/bitcoin.pdf>, 2008.



- Transaction anonymity
- No need for a trusted third party

# Agenda

## Consensus protocol

1 Introduction

2 Consensus protocol

3 Stochastic models for blockchain analysis

# Consensus protocol

## Consensus protocol

### Definition

Algorithm to allows the full nodes to agree on a common data history

It must rely on the scarce resources of the network

- bandwidth
- computational power
- storage (disk space)

# Types of consensus protocols

## Consensus protocol

### 1 Voting based

 L. Lamport, R. Shostak, and M. Pease, "The byzantine generals problem," *ACM Transactions on Programming Languages and Systems*, pp. 382–401, July 1982.

-  Communication overhead
-  Denial of service

### 2 Leader based

- Proof-of-Work (computational power)
- Proof-of-Capacity and Proof-of-Spacetime (storage)
- Proof-of-Interaction (bandwidth, locally grown)
- Proof-of-Stake (tokens)

# Proof-of-Work

Consensus protocol

## Objective

Elect a leader based on computational effort to append the next block.

# What's inside a block ?

Consensus protocol

A block consists of

- a header
- a list of "transactions" that represents the information recorded through the blockchain.

The header usually includes

- the date and time of creation of the block,
- the block height which is the index inside the blockchain,
- the hash of the block
- the hash of the previous block.

## Question

What is the hash of a block ?

<https://blockchain.info/rawblock/0>

# Cryptographic Hash function

## Consensus protocol

A function that maps data of arbitrary size (message) to a bit array of fixed size (hash value)

$$h : \{0,1\}^* \mapsto \{0,1\}^d.$$

A good hash function is

- deterministic
- quick to compute
- One way

→ For a given hash value  $\bar{h}$  it is hard to find a message  $m$  such that

$$h(m) = \bar{h}$$

- Collision resistant
  - Impossible to find  $m_1$  and  $m_2$  such that

$$h(m_1) = h(m_2)$$

- Chaotic

$$m_1 \approx m_2 \Rightarrow h(m_1) \neq h(m_2)$$

# SHA-256

Consensus protocol

The SHA-256 function which converts any message into a hash value of 256 bits.

## Example

The hexadecimal digest of the message

Is DeFi the future?

is

60a147c28568dc925c347bce20c910ef90f3774e2501ac63344f3411b6a6bf79

# Hidden prediction

Consensus protocol



**Matt Levine**   
@matt\_levine

Here is a SHA-256 hash of a prediction I am making:

64b70b0494580b278d7f1f551d482a3fb952a4b018b43090ffeb87b662d34847.



M. Levine, "The crypto story." Bloomberg business week, Oct. 2022.

# Mining a block

## Consensus protocol

```
Block Hash: 1fc23a429aa5aaf04d17e9057e03371f59ac8823b1441798940837fa2e318aaa
Block Height: 0
Time:2022-02-25 12:42:04.560217
Nonce:0
Block data: [{"sender": "Coinbase", "recipient": "Satoshi", "amount": 100, "fee": 0}, {"sender": "Satoshi", "recipient": "Pierre-O", "amount": 5, "fee": 2}]
Previous block hash: 0
Mined: False
-----
```

Figure – A block that has not been mined yet.

# Mining a block

Consensus protocol

The maximum value for a 256 bits number is

$$T_{\max} = 2^{256} - 1 \approx 1.16e^{77}.$$

Mining consists in drawing at random a nonce

$$\text{Nonce} \sim \text{Unif}(\{0, \dots, 2^{32} - 1\}),$$

until

$$h(\text{Nonce} | \text{Block info}) < T,$$

where  $T$  is referred to as the target.

Difficulty of the cryptopuzzle

$$D = \frac{T_{\max}}{T}.$$

<https://pierre-olivier.goffard.me/bitcoin-hashing/>

# Mining a block

## Consensus protocol

If we set the difficulty to  $D = 2^4$  then the hexadecimal digest must start with at least 1 leading 0

```
Block Hash: 0869032ad6b3e5b86a53f9dded5f7b09ab93b24cd5a79c1d8c81b0b3e748d226
Block Height: 0
Time:2022-02-25 13:41:48.039980
Nonce:2931734429
Block data: [{"sender": "Coinbase", "recipient": "Satoshi", "amount": 100, "fee": 0}, {"sender": "Satoshi", "recipient": "Pierre-O", "amount": 5, "fee": 2}]
Previous block hash: 0
Mined: True
-----
```

Figure – A mined block with a hash value having one leading zero.

The number of trials is geometrically distributed

- Exponential inter-block times
- Length of the blockchain = Poisson process

# Conflict resolution in blockchain

## Consensus protocol

### Fork

A fork arises when there is a disagreement between the nodes resulting in several branches in the blockchain.

### LCR

The *Longest Chain Rule* states that if there exist several branches of the blockchain then the longest should be trusted.

## In practice

- A branch can be considered legitimate if it is  $k \in \mathbb{N}$  blocks ahead of its pursuers.
- Fork can be avoided when

$$\text{block appending time} > \text{propagation delay}$$

# Bitcoin protocol

## Consensus protocol

- One block every 10 minutes on average
- Depends on the hashrate of the network
- Difficulty adjustment every 2,016 blocks ( $\approx$  two weeks)
- Reward halving every 210,000 blocks

---

. <https://www.bitcoinblockhalf.com/>

# Mining equipments

## Consensus protocol

How it started

- CPU, GPU

How it is going

- Application Specific Integrated Chip (ASIC)
  - Network electricity consumption
  - E-Waste
  - Centralization issue

# Proof of Stake

Consensus protocol

PoW is slow and ressource consuming. Let  $\{1, \dots, N\}$  be a set of miners and  $\{\pi_1, \dots, \pi_N\}$  be their share of cryptocoins.

## PoS

- 1 Node  $i \in \{1, \dots, N\}$  is selected with probability  $\pi_i$  to append the next block

Nodes are chosen according to what they own.

- Nothing at stake problem
- Rich gets richer ?
- <https://www.peercoin.net/>



F. Saleh, "Blockchain without waste : Proof-of-stake," *The Review of Financial Studies*, vol. 34, pp. 1156–1190, jul 2020.

# Ethereum Blockchain

Consensus protocol

The world-computer

- Smart contracts, decentralized applications
- Proof-of-stake (since 2022)
- a block is created every 12 seconds<sup>1</sup>
- Decentralized Finance (DeFi)

---

1. <https://txcity.io/v/eth-btc>

# Agenda

## Stochastic models

- 1 Introduction
- 2 Consensus protocol
- 3 Stochastic models for blockchain analysis

# Efficiency

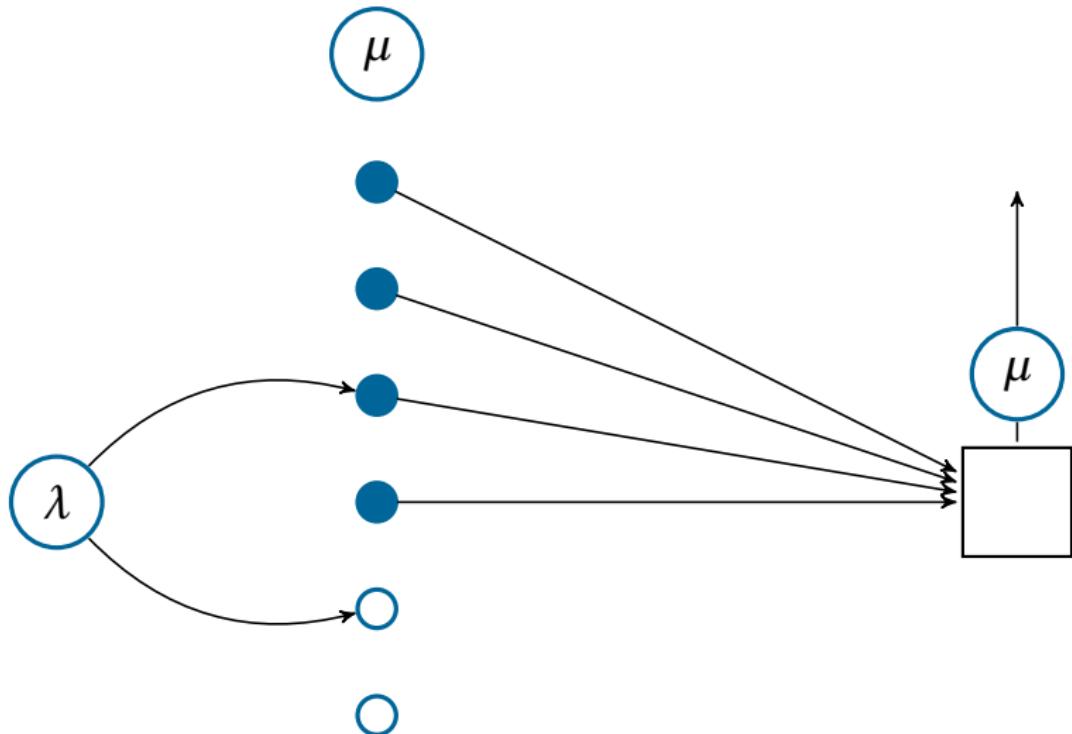
## Stochastic models

Efficiency is characterized by

- Throughputs : Number of transaction being processed per time unit
- Latency : Average transaction confirmation time

# Efficiency

Stochastic models



# Queueing setting

## Stochastic models

- Poisson arrival with rate  $\lambda > 0$  for the transactions
- Poisson arrival with rate  $\mu > 0$  for the blocks
- Block size  $b \in \mathbb{N}^*$   $\Rightarrow$  Batch service

⚠ The server is always busy

This is somekind of  $M/M^b/1$  queue.



Y. Kawase and S. Kasahara, "Transaction-confirmation time for bitcoin : A queueing analytical approach to blockchain mechanism," in *Queueing Theory and Network Applications*, pp. 75–88, Springer International Publishing, 2017.



N. T. J. Bailey, "On queueing processes with bulk service," *Journal of the Royal Statistical Society : Series B (Methodological)*, vol. 16, pp. 80–87, jan 1954.



D. R. Cox, "The analysis of non-markovian stochastic processes by the inclusion of supplementary variables," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 51, pp. 433–441, jul 1955.

# Queue length distribution

## Stochastic models

The queueing process eventually reaches stationarity if

$$\mu \cdot b > \lambda. \quad (1)$$

We denote by  $N^q$  the length of the queue upon stationarity.

### The blockchain efficiency theorem

Assume that (1) holds then  $N^q$  is geometrically distributed

$$\mathbb{P}(N^q = n) = (1-p) \cdot p^n,$$

where  $p = 1/z^*$  and  $z^*$  is the only root of

$$-\frac{\lambda}{\mu} z^{b+1} + z^b \left( \frac{\lambda}{\mu} + 1 \right) - 1,$$

such that  $|z^*| > 1$ .

# Latency and throughputs

## Stochastic models

### Little's law

Consider a stationary queueing system and denote by

- $1/\lambda$  the mean of the unit inter-arrival times
- $L$  be the mean number of units in the system
- $W$  be the mean time spent by units in the system

We have

$$L = \lambda \cdot W$$



J. D. C. Little, "A proof for the queuing formula : $L = \lambda W$ ," *Operations Research*, vol. 9, pp. 383–387, jun 1961.

- Latency is the confirmation time of a transaction

$$\text{Latency} = W + \frac{1}{\mu} = \frac{\mathbb{E}(N^q)}{\lambda} + \frac{1}{\mu} = \frac{p}{(1-p)\lambda} + \frac{1}{\mu}.$$

- Throughput is the number of transaction confirmed per time unit

$$\text{Throughput} = \mu \mathbb{E}(N^q \mathbb{I}_{N^q \leq b} + b \mathbb{I}_{N^q > b}) = \mu \sum_{n=0}^b n(1-p)p^n + bp^{b+1}.$$

# Perspectives

## Stochastic models

### 1 Include some priority consideration to account for the transaction fees



Y. Kawase, , and S. Kasahara, "Priority queueing analysis of transaction-confirmation time for bitcoin," *Journal of Industrial & Management Optimization*, vol. 16, no. 3, pp. 1077–1098, 2020.

### 2 Go beyond the Poisson process framework



Q.-L. Li, J.-Y. Ma, and Y.-X. Chang, "Blockchain queue theory," in *Computational Data and Social Networks*, pp. 25–40, Springer International Publishing, 2018.



Q.-L. Li, J.-Y. Ma, Y.-X. Chang, F.-Q. Ma, and H.-B. Yu, "Markov processes in blockchain systems," *Computational Social Networks*, vol. 6, jul 2019.

# Risk of centralization ?

## Stochastic models

### PoS algorithm

- Draw a coin at random
- The owner of the coin append a block and collect the reward
- The block appender is more likely to get selected during the next round

Similar to Polya's urn



- Consider an urn of  $N$  balls of color in  $E = \{1, \dots, p\}$
- Draw a ball of color  $x \in E$
- Replace the ball together with  $r$  balls of color  $x$

$p$  is the number of peers and  $r$  is the size of the block reward.

### Theorem

The proportion of coins owned by each peer is stable on average over the long run



I. Roşu and F. Saleh, "Evolution of shares in a proof-of-stake cryptocurrency," *Management Science*, vol. 67, pp. 661–672, feb 2021.

# Extensions and perspectives

## Stochastic models

- How to include more peers along the way ?
- What if the peers are not simply buy and hold investors ?
- Find ways to monitor decentralization and take action if necessary

# Double spending attack

Stochastic models

- 1 Mary transfers 10 BTCs to John
- 2 The transaction is recorded in the public branch of the blockchain and John ships the good.
- 3 Mary transfers to herself the exact same BTCs
- 4 The malicious transaction is recorded into a private branch of the blockchain
  - Mary has friends among the miners to help her out
  - The two chains are copycat up to the one transaction

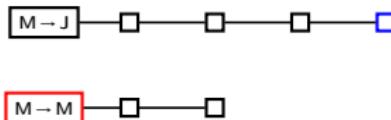
Fact (Bitcoin has only one rule)

The longest chain is to be trusted

# Double Spending attack

## Stochastic models

Vendor are advised to wait for  $\alpha \in \mathbb{N}$  of confirmations so that the honest chain is ahead of the dishonest one.



In the example, vendor awaits  $\alpha = 4$  confirmations, the honest chain is ahead of the dishonest one by  $z = 2$  blocks.

### Fact (PoW is resistant to double spending)

- Attacker does not own the majority of computing power
- Suitable  $\alpha$

Double spending is unlikely to succeed.



S. Nakamoto, "Bitcoin : A peer-to-peer electronic cash system." Available at <https://bitcoin.org/bitcoin.pdf>, 2008.

# Double Spending Attack

## Stochastic models

Assume that

- $R_0 = z \geq 1$  (the honest chain is  $z$  blocks ahead)
- at each time unit a block is created
  - in the honest chain with probability  $p$
  - in the dishonest chain with probability  $q = 1 - p$

The process  $(R_n)_{n \geq 0}$  is a random walk on  $\mathbb{Z}$  with

$$R_n = z + Y_1 + \dots + Y_n,$$

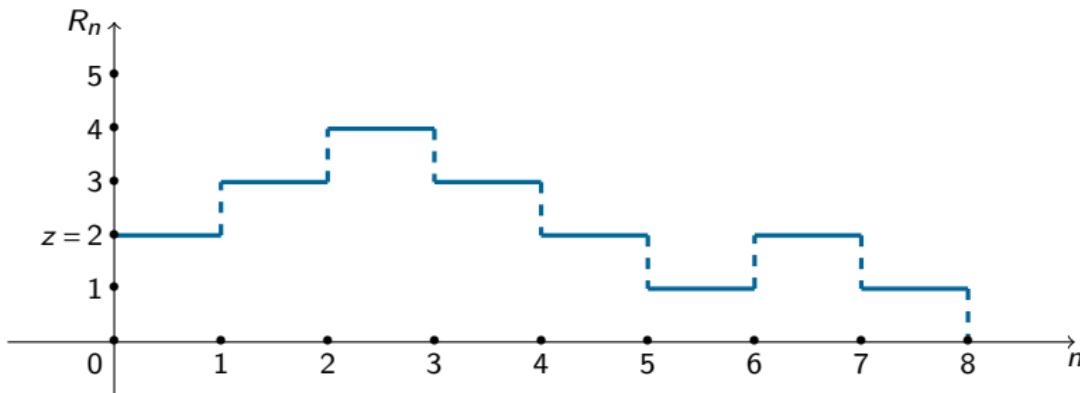
where  $Y_1, \dots, Y_n$  are the i.i.d. steps of the random walk.

# Double Spending Attack

Stochastic models

Double spending occurs at time

$$\tau_z = \inf\{n \in \mathbb{N} ; R_n = 0\}.$$



Double spending theorem

If  $p > q$  then the double-spending probability is given by

$$\psi(z) = \mathbb{P}(\tau_z < \infty) = \left(\frac{q}{p}\right)^z.$$

# Double spending with Poisson processes

## Stochastic models

Let the length of honest and dishonest chain be driven by counting processes

- Honest chain  $\Rightarrow z + N_t$ ,  $t \geq 0$ , where  $z \geq 1$ .
- Malicious chain  $\Rightarrow M_t$ ,  $t \geq 0$
- Study the distribution of the first-*rendez-vous* time

$$\tau_z = \inf\{t \geq 0, M_t = z + N_t\}.$$



P.-O. Goffard, "Fraud risk assessment within blockchain transactions," *Advances in Applied Probability*, vol. 51, pp. 443–467, jun 2019.  
<https://hal.archives-ouvertes.fr/hal-01716687v2>.



R. Bowden, H. P. Keeler, A. E. Krzesinski, and P. G. Taylor, "Modeling and analysis of block arrival times in the bitcoin blockchain," *Stochastic Models*, vol. 36, pp. 602–637, jul 2020.

# Blockchain as a research topic

## Stochastic models

- Computer science
  - Peer-to-peer networks and consensus algorithm
  - Cryptography and security
- Economics
  - Game theory to study the incentive mechanism at play
  - Nature of the cryptoassets
- Operations research
  - Optimization of complex system
- Financial math
  - Valuation models for cryptoassets
- Machine learning and statistics
  - Open data
  - Interaction between blockchain users
  - (Social) network analysis
  - Clustering of public keys and addresses in the bitcoin blockchain.

# TradFi Pain Points I

## 1 Access barrier

- TradFi Criteria ⇒ bank account
- DeFi No Barrier

## 2 Centralization

- TradFi Bank are record keepers ⇒ Cyber risk
- DeFi Decentralized Ledger

## 3 High costs and intermediation

- TradFi Transaction fees, account maintenance fees, wire transfer fees,...
- DeFi Intermediaries ⇒ Smart contracts (Gas fees)

## 4 Slow transaction settlement

- TradFi Cross border transactions ⇒ Takes day to be settled
- DeFi Near instant settlement (30 min on the bitcoin blockchain)

## 5 Transparency and auditability

- TradFi Difficulty to verify the accuracy of transactions and asset holding
- DeFi Publicly available ledger and open source code

# TradFi Pain Points II

## 6 Censorship and restrictions

- TradFi Asset in custody, censorship by governments
- DeFi No interference of central authority

## 7 Global accessibility

- TradFi Respect the operating hours
- DeFi Operates 24/7

## 8 Fractional ownership

- TradFi Real estate and art work are not divisible
- DeFi Tokenized real world assets

## 9 Innovation and interoperability

- TradFi Use outdated IT solutions
- DeFi Interoperability between platforms and project



A. Lipton and A. Treccani, *Blockchain and Distributed Ledgers*.  
WORLD SCIENTIFIC, apr 2021.

# Cryptocurrencies

- 1 No central authority (Decentralized network)
- 2 Ledger to record all the transactions and coin ownership (blockchain)
- 3 A coin generation process (block finding reward)
  - Incentive to the full nodes
- 4 Ownership can be proved cryptographically (wallet associated to a public/private key)
- 5 Transactions can be issued by an entity proving ownership of the cryptographic unit (through the private key)
- 6 The system cannot process more than one transaction associated to the same cryptographic unit (double spending)



J. Lansky, "Possible state approaches to cryptocurrencies," *Journal of Systems Integration*, vol. 9, pp. 19–31, jan 2018.

# Cryptocurrency implementation

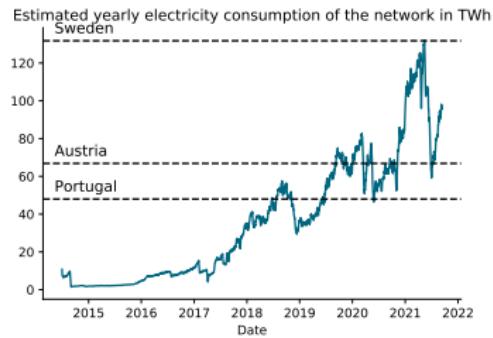
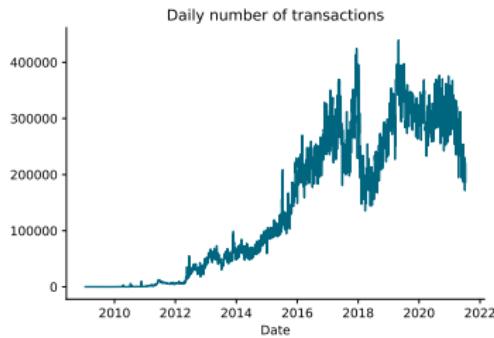
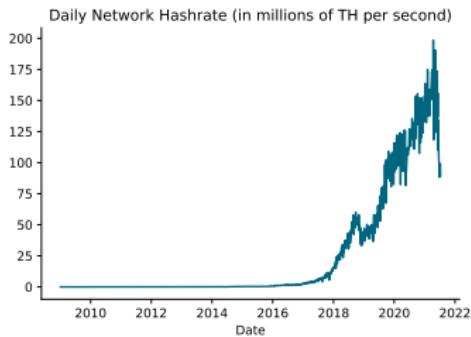
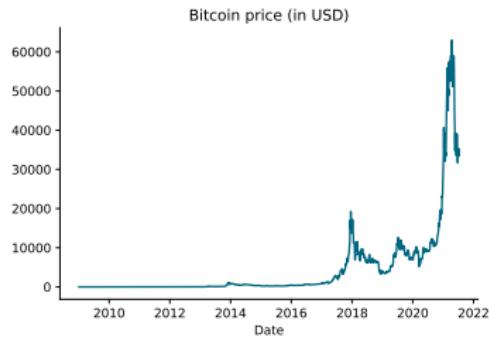
## Blockchain parameters

- Consensus protocol (PoW or PoS)
  - ↳ Hash function (SHA-256 for Bitcoin and scrypt for LiteCoin)
  - ↳ Hybrid PoW/PoS (PeerCoin)
- Block generation time (<https://txcity.io/v/eth-btc>)
  - ↳ every 10 minutes for Bitcoin
  - ↳ every 12 sec for Ethereum
- Block finding reward
  - ↳ Halved every 210,000 blocks in Bitcoin. It started at 50 BTC, is now 6.25 BTC  
<https://www.bitcoinblockhalf.com/>
- Total coin supply
  - ↳ 21,000,000 in total for Bitcoin
- Transaction fees
  - ↳ GAS in Ethereum

These choices lead to the creation of multiple cryptocurrencies

### Examples

Bitcoin and AltCoins (Ethereum, LiteCoin, DogeCoin, Ripple... ), see [https://en.wikipedia.org/wiki/List\\_of\\_cryptocurrencies](https://en.wikipedia.org/wiki/List_of_cryptocurrencies)



# Decentralized finance

DeFi extends the Bitcoin promises to more complex financial operations

- Collateralized lending
- Decentralized Exchange Platform
- Tokenized assets

Thanks to Smart Contract on the Ethereum blockchain.



S. M. Werner, D. Perez, L. Gudgeon, A. Klages-Mundt, D. Harz, and W. J. Knottenbelt,  
"Sok : Decentralized finance (defi)," 2021.

---

. <https://uniswap.org/>

# Decentralized Exchange platforms

## Centralized Exchange

Binance, Coinbase, Kraken,...

⇒ Order book

## Decentralized Exchange

Exchange where you trade one token for another through a smart contract on the Ethereum blockchain (ETH).

⇒ Automated Market Makers (AMM) and Liquidity Providers (LPs)

# Order book

Table – Buy Orders

Price	Quantity	Total
9,950	10	99,500
9,900	5	49,500

Table – Sell Orders

Price	Quantity	Total
10,100	2	20,200
10,200	15	153,000

and a market maker in the middle to offer liquidity...

⚠ Lots of transaction must be issued

- Slow (10-15 txs/s)
- Expensive (gas fees)

# Order book

Table – Buy Orders

Price	Quantity	Total
9,950	10	99,500
9,900	5	49,500

Table – Sell Orders

Price	Quantity	Total
10,100	2	20,200
10,200	15	153,000

and a market maker in the middle to offer liquidity...

⚠ Lots of transaction must be issued

- Slow (10-15 txs/s)
- Expensive (gas fees)

# Automated Market Makers (AMM)

A blockchain requires an algorithm ⇒ AMM

- Exchange one token against another, usually Crypto/Stable Coins
- Constant Function Market Makers
- Liquidity Providers



## Stable Coin

A bridge from fiat to crypto currency

- Fiat-Collateralized Stablecoins (e.g. USDC backed by one USD)
- Crypto-Collateralized Stablecoins (e.g. DAI backed by crypto locked in a smart contract)

# Constant Product Market Maker

Let  $k$  be a constant such that

$$x \cdot y = k,$$

- $x$  is the amount of token  $X$
- $y$  is the amount of token  $Y$

## Example

### ETH/DAI even Constant Product Market Maker

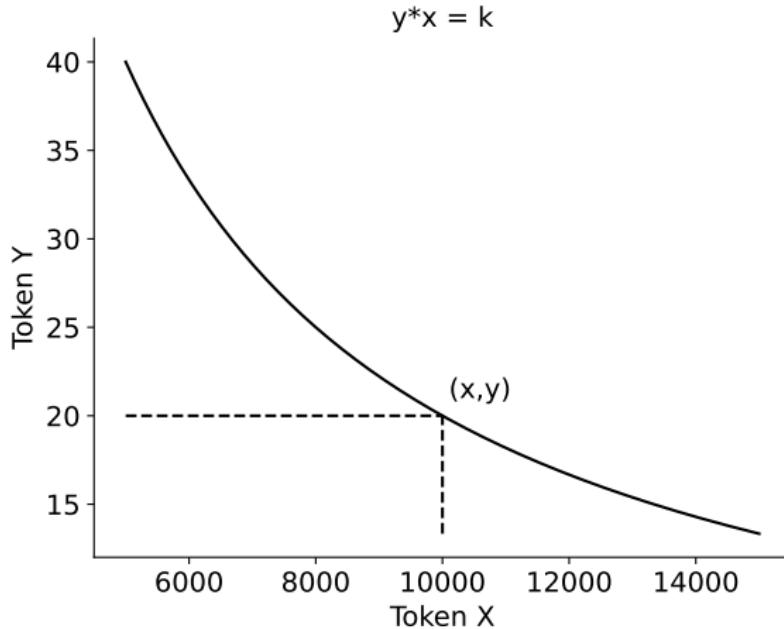
- Say that the price for one ETH is  $P = \$500$
- A first LP provides 20 ETH plus 10,000 DAI which sets

$$k = 20,000$$

- The liquidity provided is measured by

$$L = \sqrt{x \cdot y} \approx 447 \text{ (geometric mean)}$$

## Trade on a curve



- . The pool never runs out of X nor Y

## Swap $X$ for $Y$

Acquire  $dy$  of token  $Y$ , then deposit  $dx$  that solves

$$(x + dx)(y - dy) = k \Leftrightarrow dx = \frac{x \cdot dy}{y - dy}$$

and pay a fee  $\alpha \cdot dx$  to the LPs.

⇒ Price of  $Y$  rise in the pool

### Example

Arbitrageur takes 2 ETH then deposits

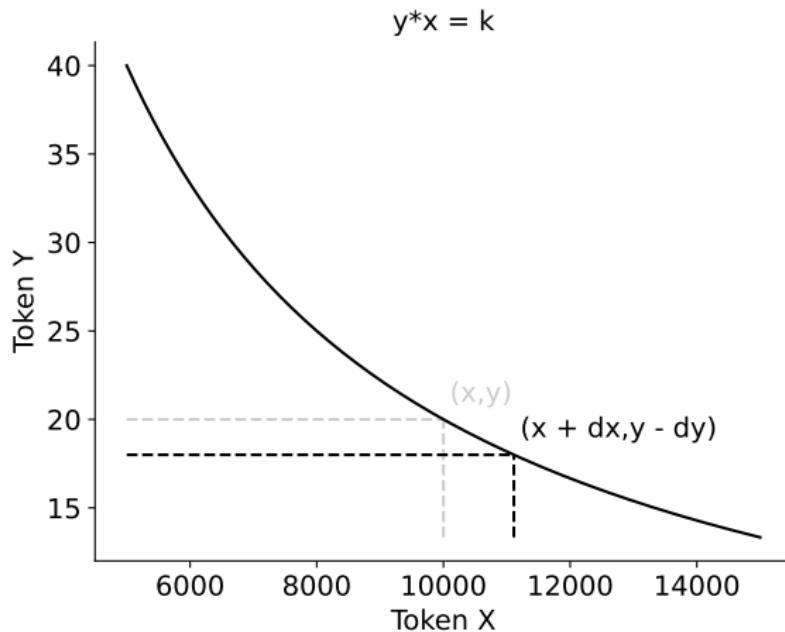
$$dx = 1,111$$

and give out  $0.3 \cdot 1,111 = 333$  worth of fees

- The price of ETH in the pool rises to

$$\frac{x + dx}{y - dy} = \$617.$$

## Trade on a curve



# Add Liquidity

Another LP provides  $dx$  of token  $X$  and thus  $dy = \frac{y}{x} dx$  (the price must not change)

- New level  $k' = (x + dx)(y + dy)$
- Liquidity rises  $L' = \sqrt{x \cdot y} + \sqrt{dx \cdot dy}$
- LPs are weighted according to the liquidity they provide
- Fees are distributed according to these weights

## Example

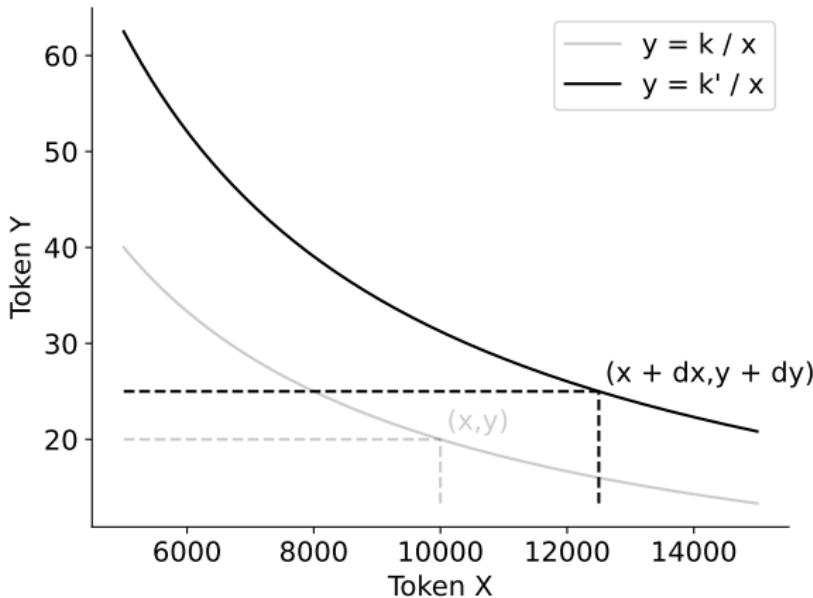
New LP deposits \$5,000 worth of tokens to the pool then

- $dx = 2,500$
- $dy = 5$
- $k' = 312,500$
- $L' \approx 559$

The weights are

$$\frac{L}{L'} = 0.8 \text{ and } \frac{L' - L}{L'} = 0.2.$$

## Trade on a new curve



# Impermanent Loss

Holding tokens VS Locking Tokens in a Smart Contract

# Impermanent Loss

The price of  $Y$  in the pool is given by

$$P = \frac{x}{y},$$

in terms of token  $X$ .

- Price of  $Y$  becomes  $P' > P$  on another trading venue then arbitrageurs wish to find

$$\underset{dy > 0}{\operatorname{argmax}} dy \cdot P' - dx \cdot (1 + \alpha) = \underset{dy}{\operatorname{argmax}} dy \cdot P' - \frac{x \cdot dy}{y - dy} (1 + \alpha)$$

- We have

$$dy^* = y - \sqrt{\frac{k(1 + \alpha)}{P'}}$$

- The arbitrageurs profit is

$$dy^* \cdot P' - dx^* (1 + \alpha)$$

- It coincides with the impermanent loss of the LPs

$$y \cdot P' + x - [(y - dy^*) \cdot P' + x + dx^* (1 + \alpha)]$$

# Impermanent Loss

The loss becomes permanent if

- The LPs withdraw their funds
- The price of the asset is not mean reverting

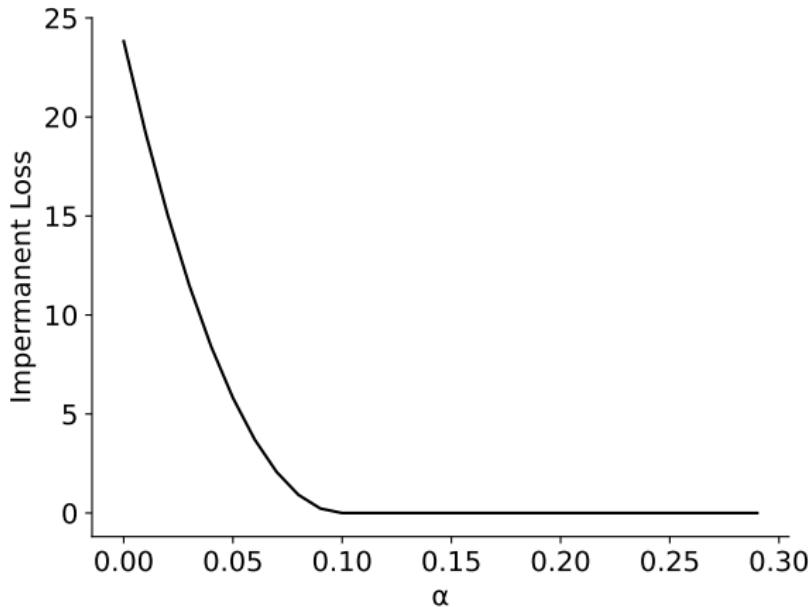
## Example

Suppose that  $P' = \$550$  and  $\alpha = 0$  then

- $dy^* = 0.93$
- The impermanent loss is then  $dy^* \cdot P' - \frac{xdy^*}{y-dy^*} = \$23$

## Impermanent loss

The trading fee allows a pool to mitigate the impermanent loss



# Decentralized insurance

## Parametric insurance

Compensation if a measurable quantity reaches a threshold

- Example : Flight delay insurance
  - [https://etherscan.io/address/  
0xdc3d8fc2c41781b0259175bdc19516f7da11cba7](https://etherscan.io/address/0xdc3d8fc2c41781b0259175bdc19516f7da11cba7)
- Use smart contract and off-chain data through oracles
- Transparent and automatic