# Ethereum EVM illustrated

exploring some mental models and implementations

Takenobu T.



### NOTE

- Please refer to the official documents in detail.
- This information is current as of Mar, 2018.
- Still work in progress.

#### Contents

- 1. Introduction
  - Blockchain
  - World state
  - Account
  - Transaction
  - Message
  - Decentralised database
  - Atomicity and order
- 2. Virtual machine
  - Ethereum virtual machine (EVM)
  - Message call
  - Exception
  - Gas and fee
  - Input and output
  - Byte order
  - Instruction set
  - Miscellaneous

#### Appendix A: Implementation

- Source code in Geth
- EVM developer utility
- Solidity ABI

#### Appendix B: User interface

- Web3 API
- Geth, Mist, Solc, Remix, Truffle, ...

#### Appendix C:

- Markle tree and RLP
- Consensus

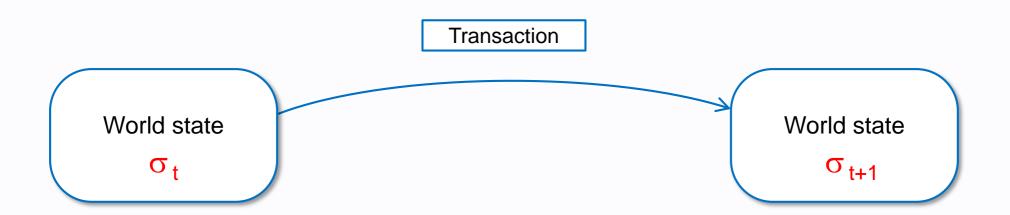
#### References

# 1. Introduction

# 1. Introduction

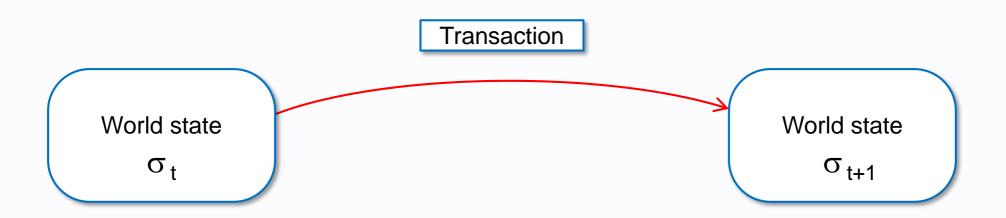
## Blockchain

#### A transaction-based state machine



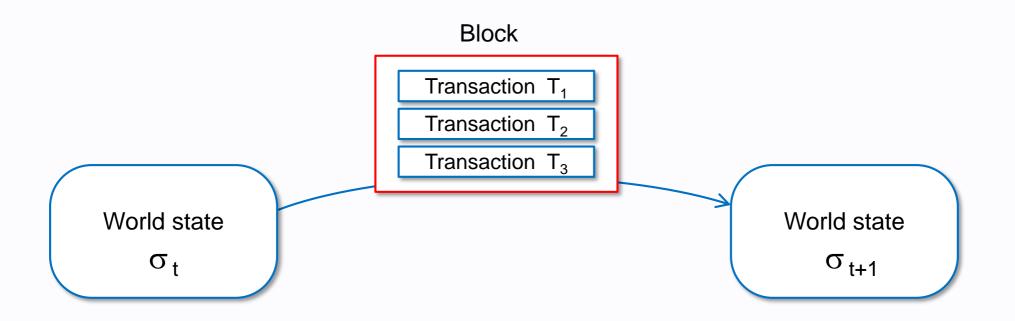
Ethereum can be viewed as a transaction-based state machine.

#### A transaction-based state machine



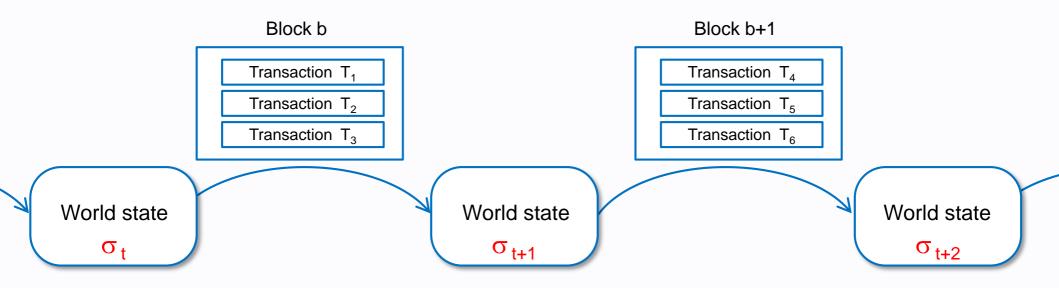
A transaction represents a valid arc between two states.

### Block and transactions



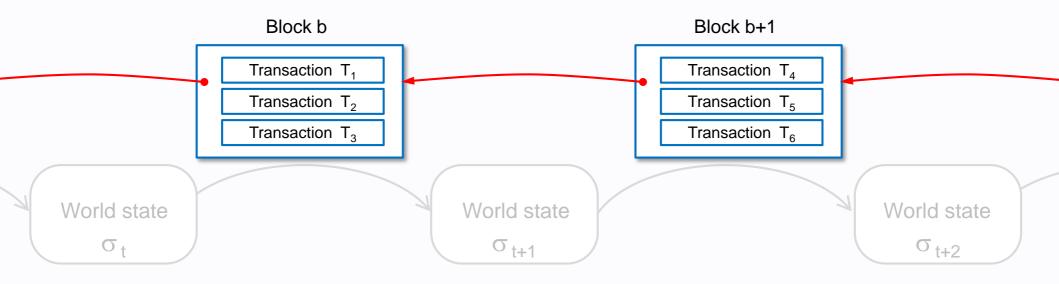
Transactions are collated into blocks. A block is a package of data.

#### Chain of states



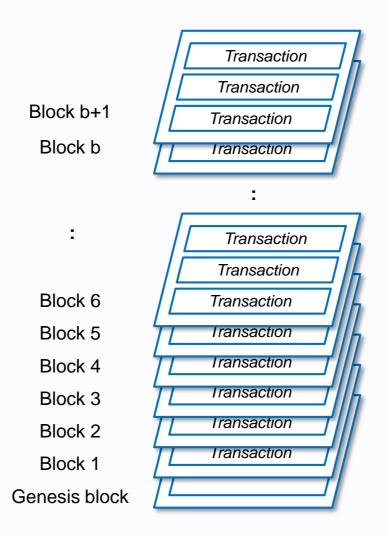
From the viewpoint of the states, Ethereum can be seen as a state chain.

#### Chain of blocks: Blockchain



From the viewpoint of the implementation, Ethereum can also be seen as a chain of blocks, so it is `BLOCKCHAIN`.

## Stack of transactions: Ledger



From the viewpoint of the ledger, Ethereum can also be seen as a stack of transactions.

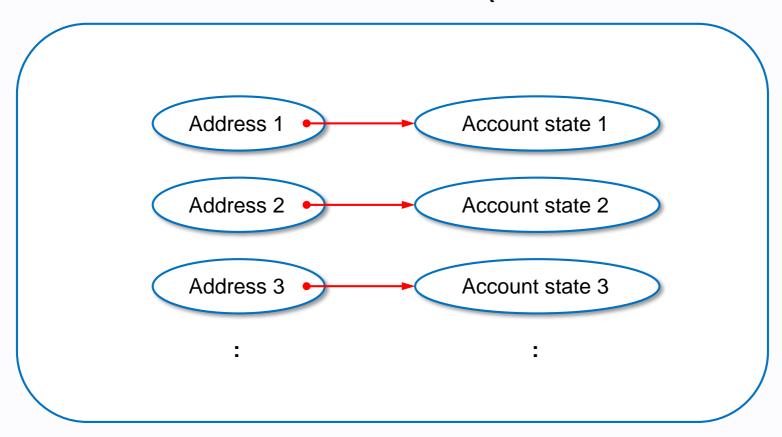
References: [E1] Ch.2, Ch.4, [E2], [E3], [W3]

# 1. Introduction

World state

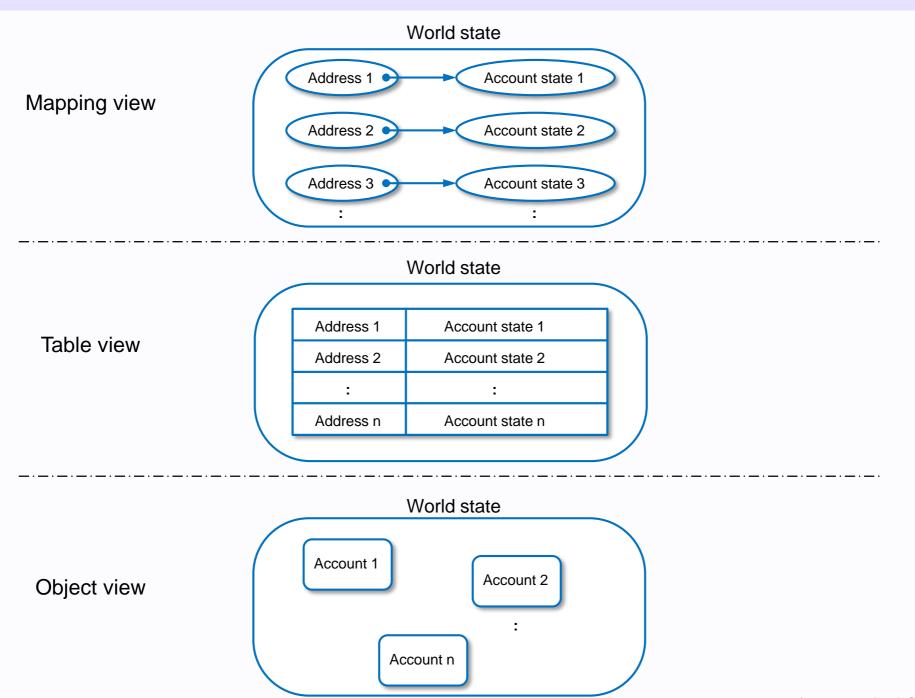
#### World state

### World state $\sigma_t$



The world state is a mapping between address and account state.

#### Several views of world state



References: [E1] Ch.4

# 1. Introduction

Account

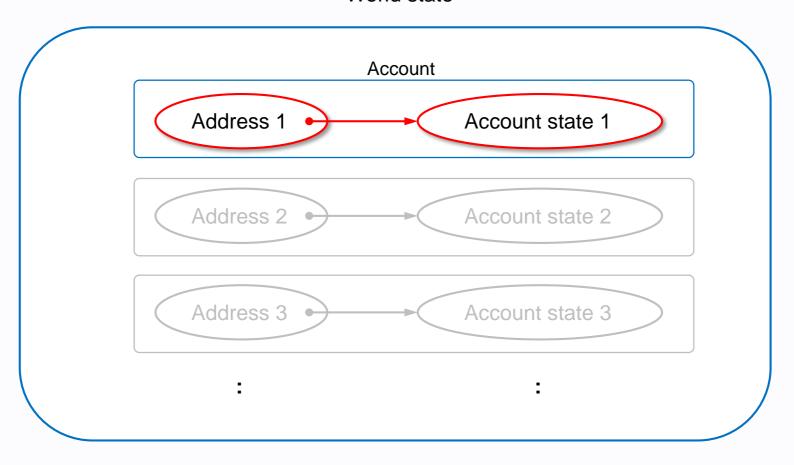
### Account



An account is an object in the world state.

#### Account

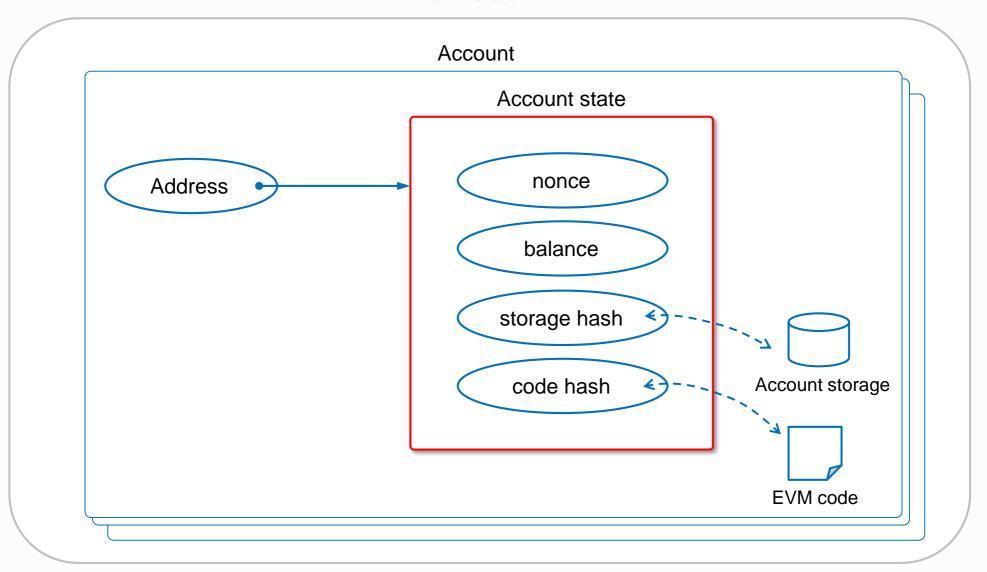
#### World state



An account is a mapping between address and account state.

#### Account state

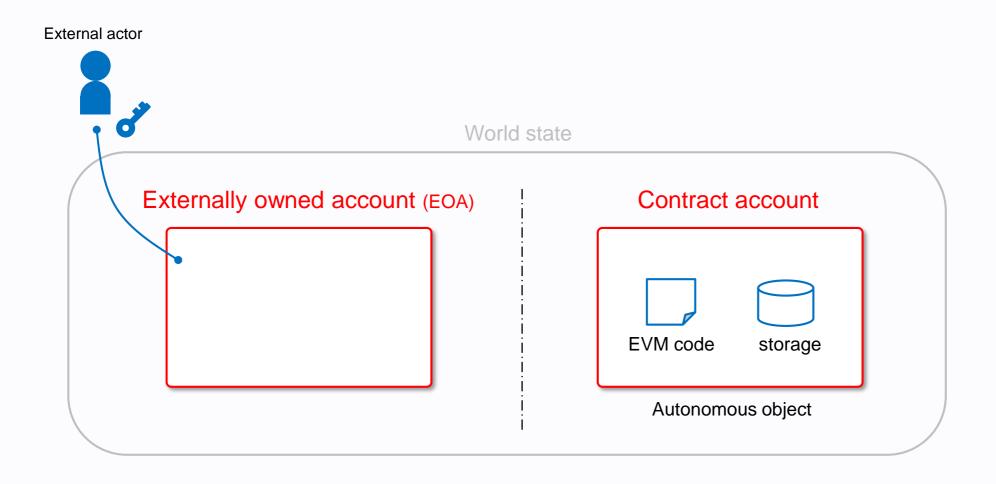
World state



An account state could contain EVM code and storage.

References: [E1] Ch.4

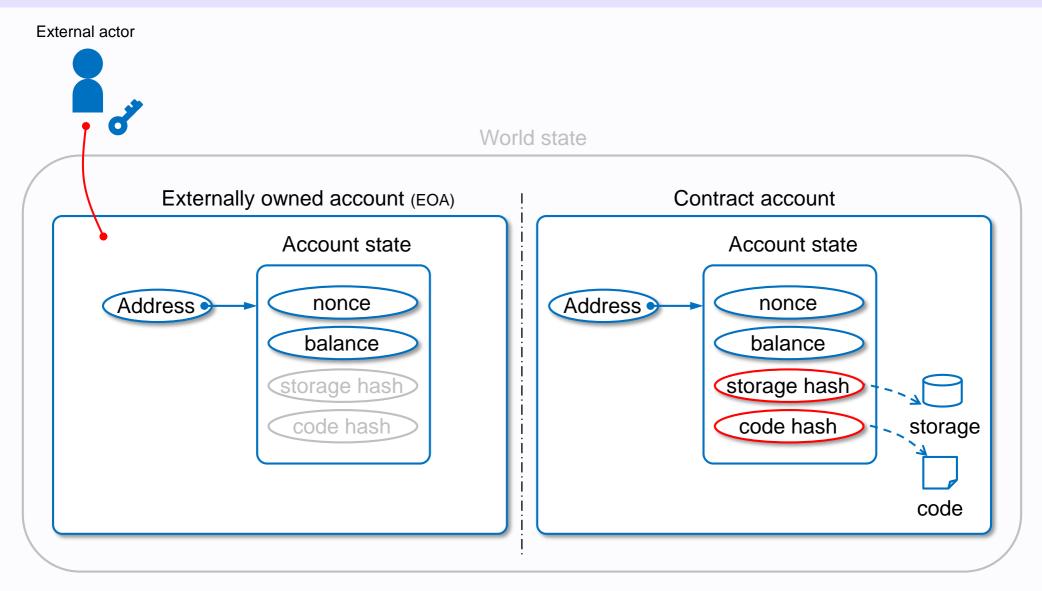
### Two practical types of account



EOA is controlled by a private key.

Contract account contains EVM code.

## Two practical types of account



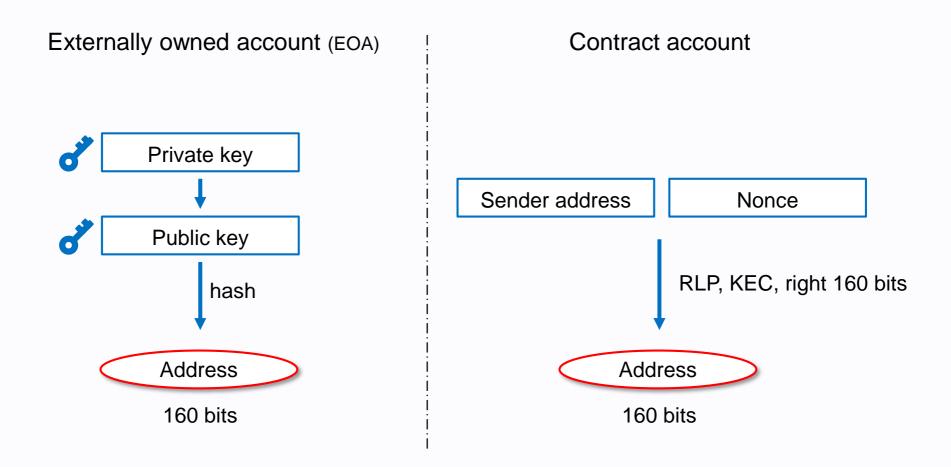
EOA is controlled by a private key. EOA cannot contain EVM code.

Contract contains EVM code.

Contract is controlled by EVM code.

References: [E1] Ch.4

#### Address of account

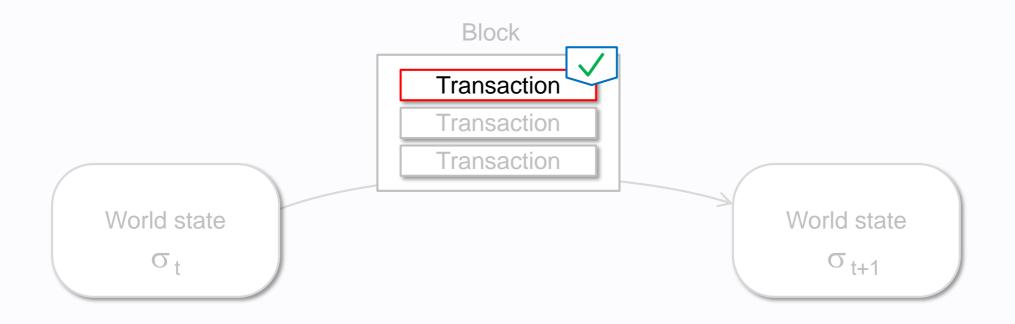


A 160-bit code used for identifying accounts.

# 1. Introduction

**Transaction** 

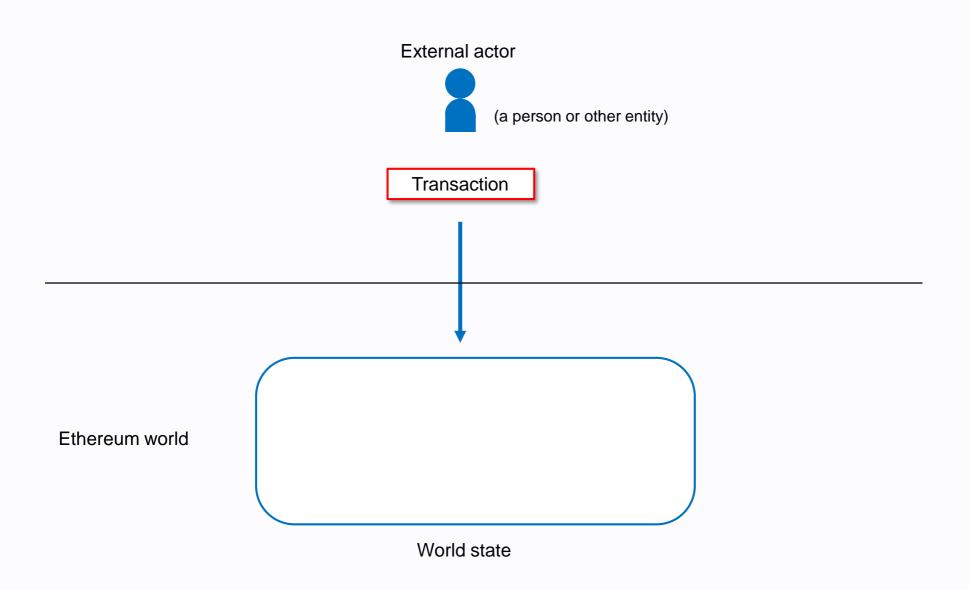
#### A transaction



A transaction is a single cryptographically-signed instruction.

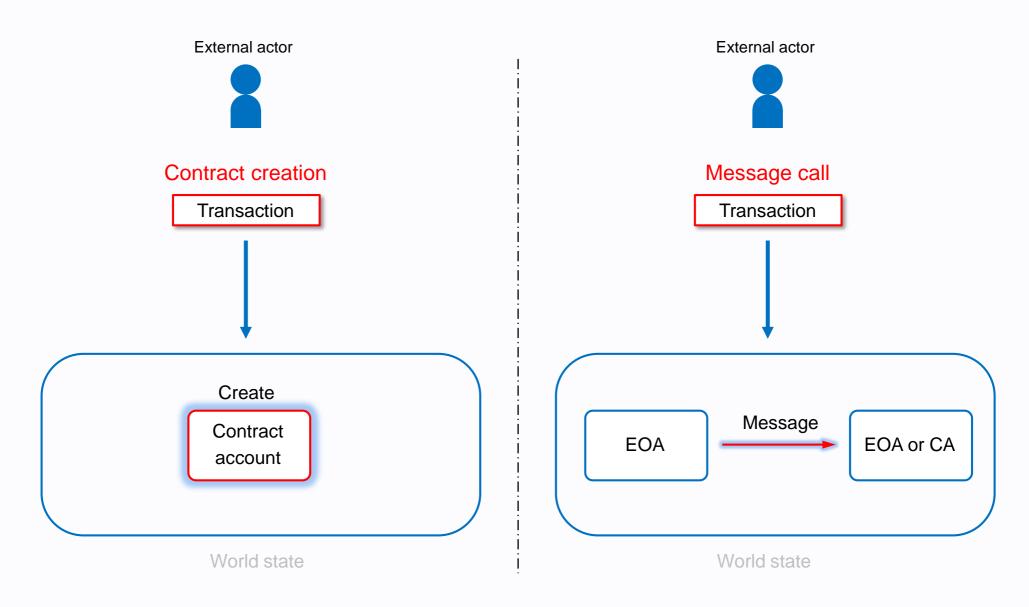
References: [E1] Ch.2, Ch.4, [E2]

#### A transaction to world state



A transaction is submitted by external actor.

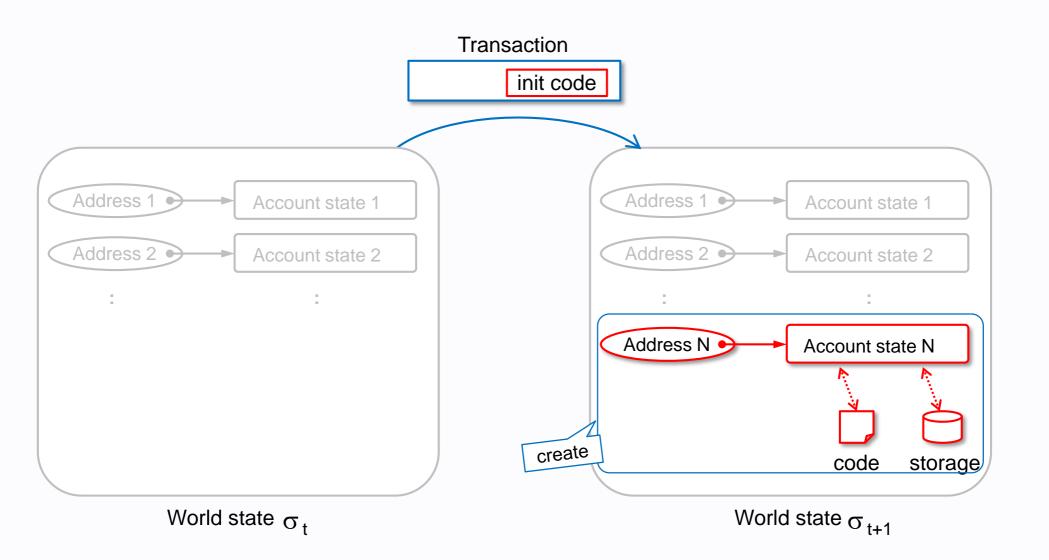
## Two practical types of transaction



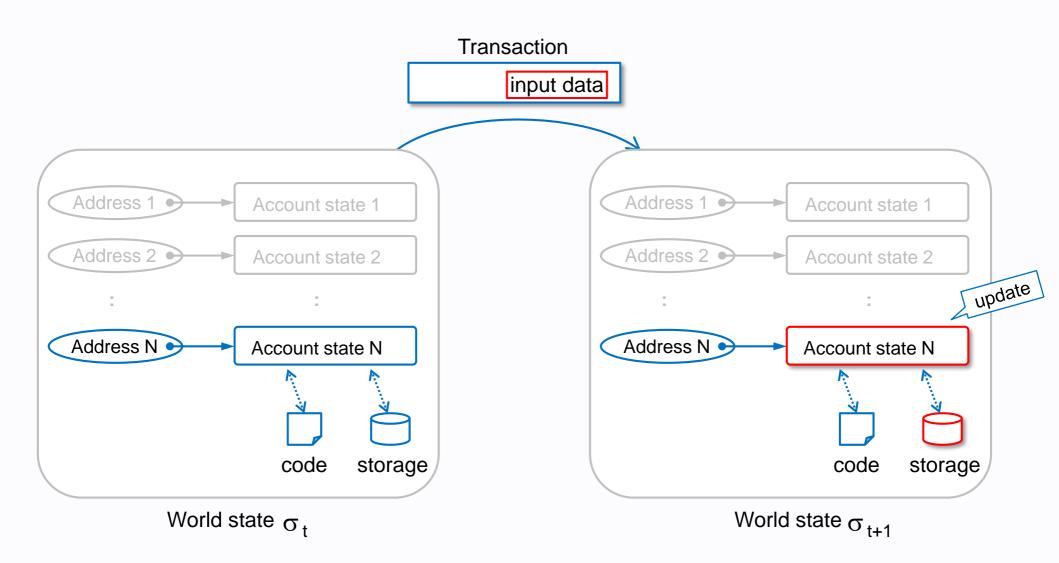
There are two practical types of transaction, contract creation and message call.

References: [E1] Ch.4

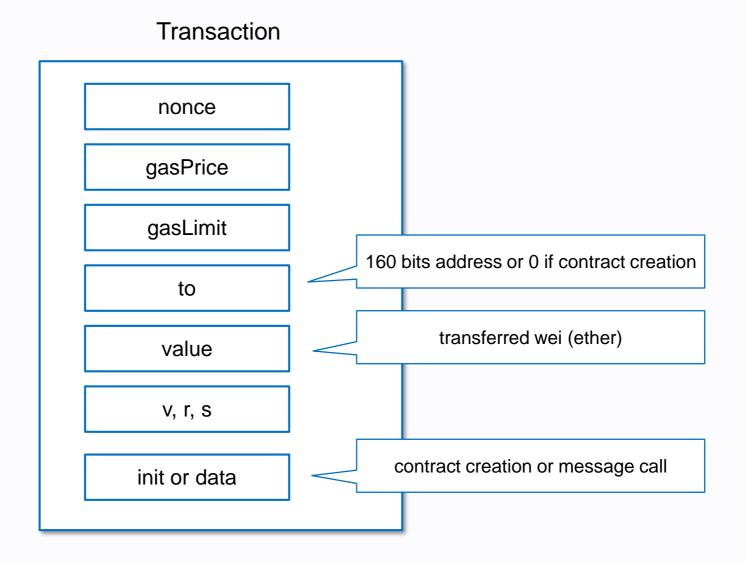
#### **Contract creation**



## Message call



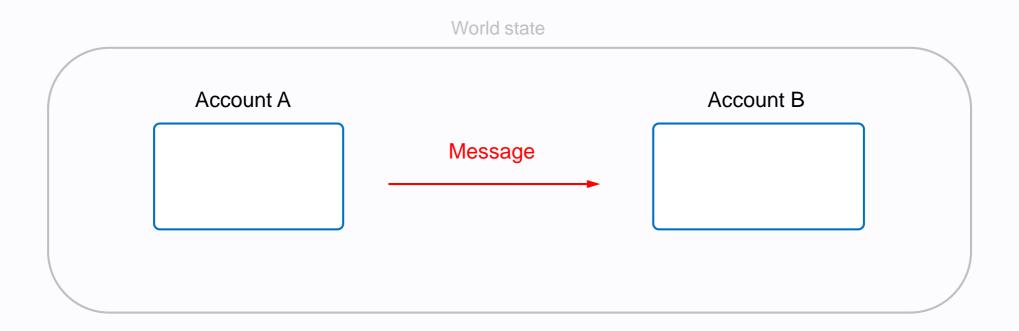
### Field of a transaction



## 1. Introduction

Message

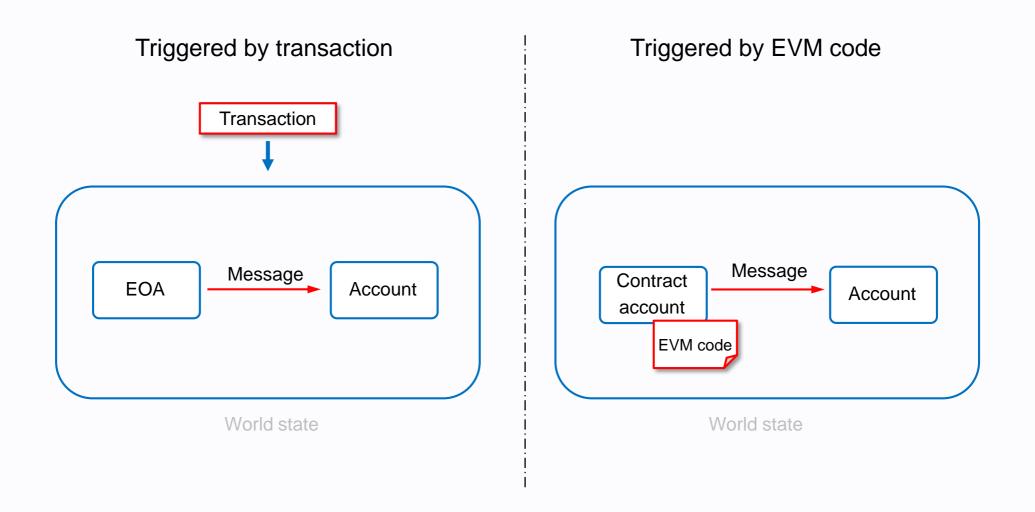
## Message



Message is passed between two Accounts.

Message is Data (as a set of bytes) and Value (specified as Ether) .

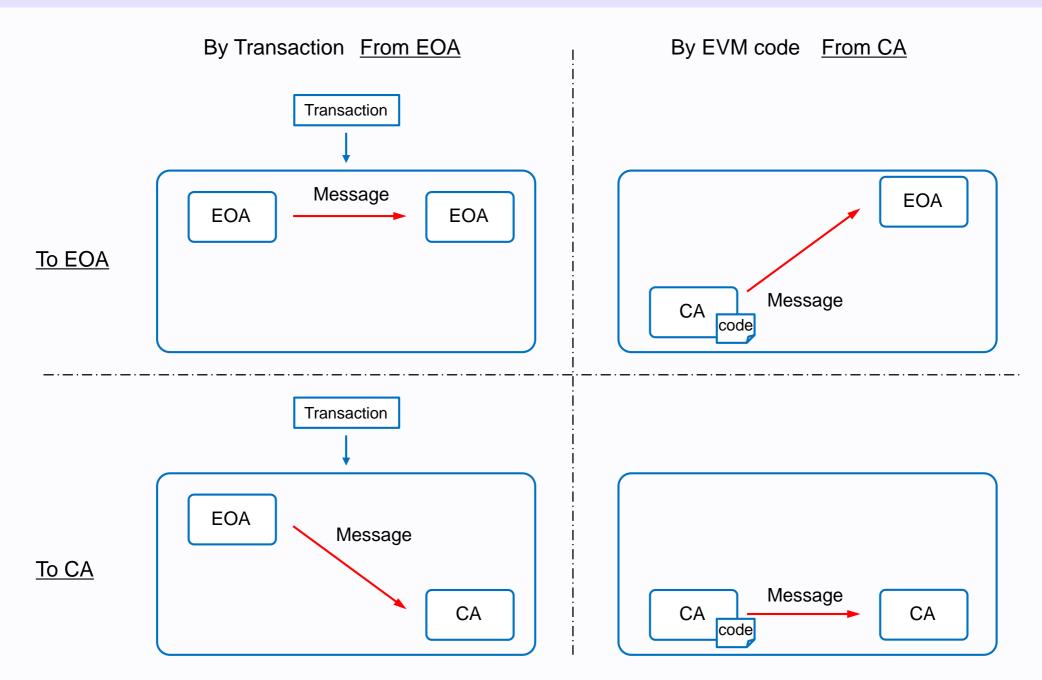
### Message



Transaction triggers an associated message.

EVM can also send a message.

## Four cases of message

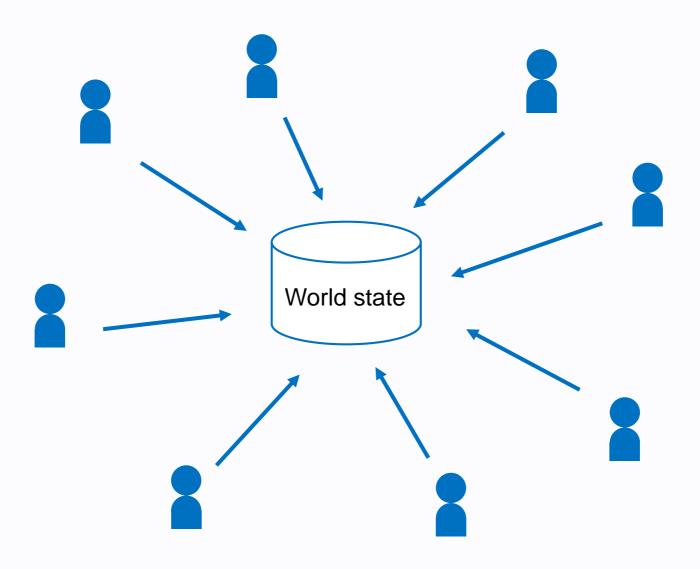


References: [E1] Ch.8

## 1. Introduction

## Decentralised database

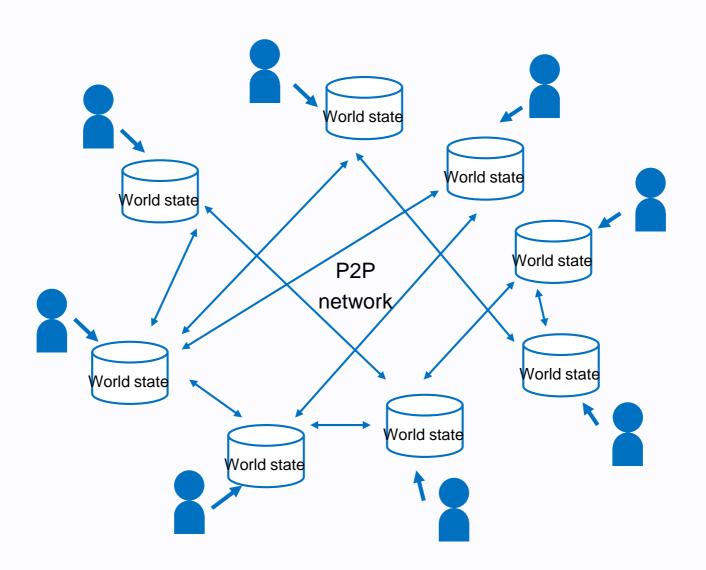
## Globally shared, transactional database



A blockchain is a globally shared, transactional database.

References: [E3], [E7] Ch.7

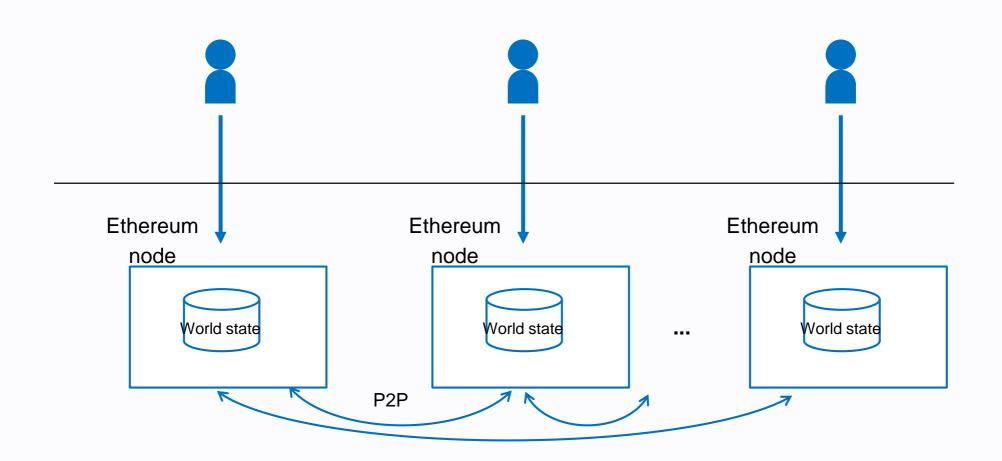
#### Decentralised database



A blockchain is a globally shared, decentralised, transactional database.

References: [E3], [E7] Ch.7

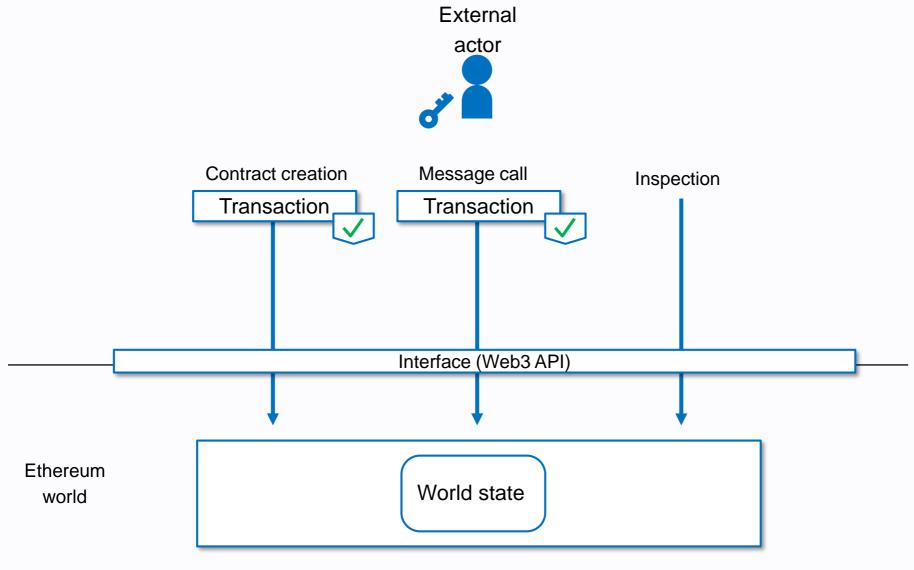
#### P2P network inter nodes



Decentralised nodes constitute Ethereum P2P network.

References: [E3],

#### Interface to a node



Ethereum node (Geth, Parity, ...)

External actors access the Ethereum world through Ethereum nodes.

References: [E1] Appendix A, Ch.4, Ch.7, Ch.8

## 1. Introduction

Atomicity and order

### Atomicity of transaction

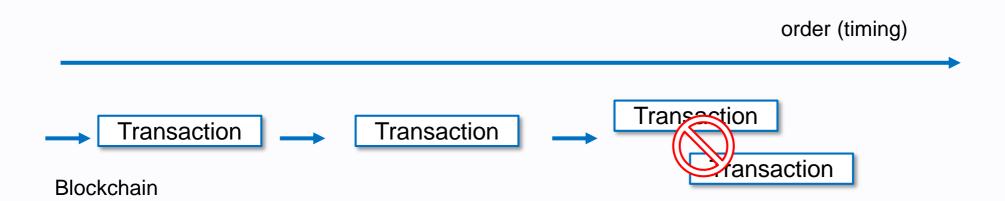


A transaction is an atomic operation. Can't divide or interrupt.

Transaction or Transaction

That is, All (complete done) or Nothing (zero effect).

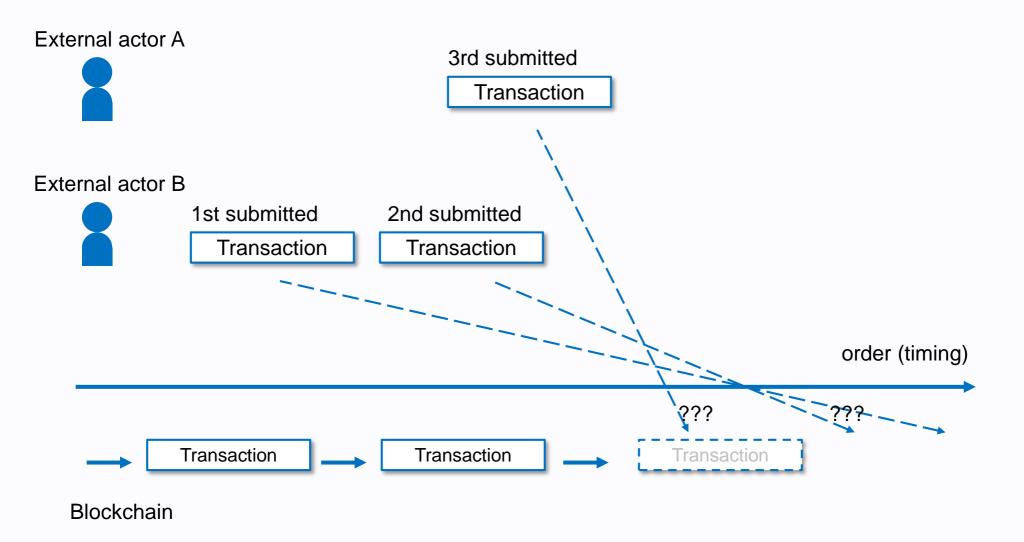
#### Order of transactions



Transactions cannot be overlapped.

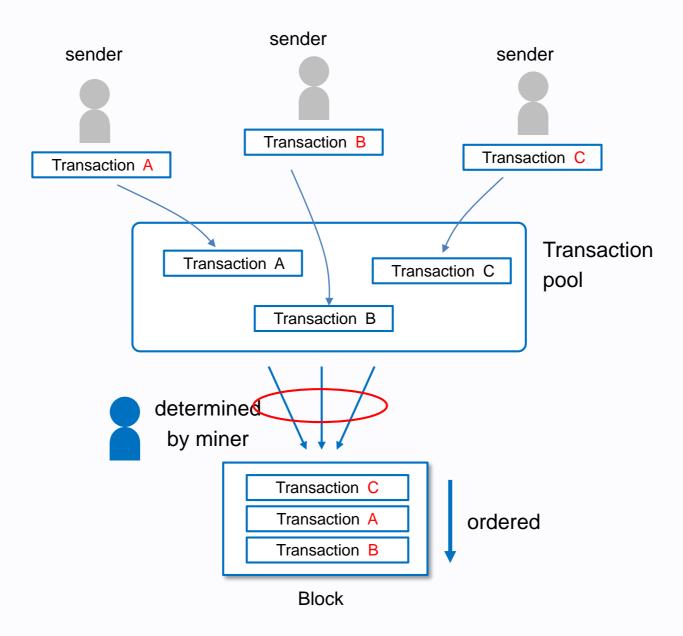
Transactions must be executed sequentially.

#### Order of transactions



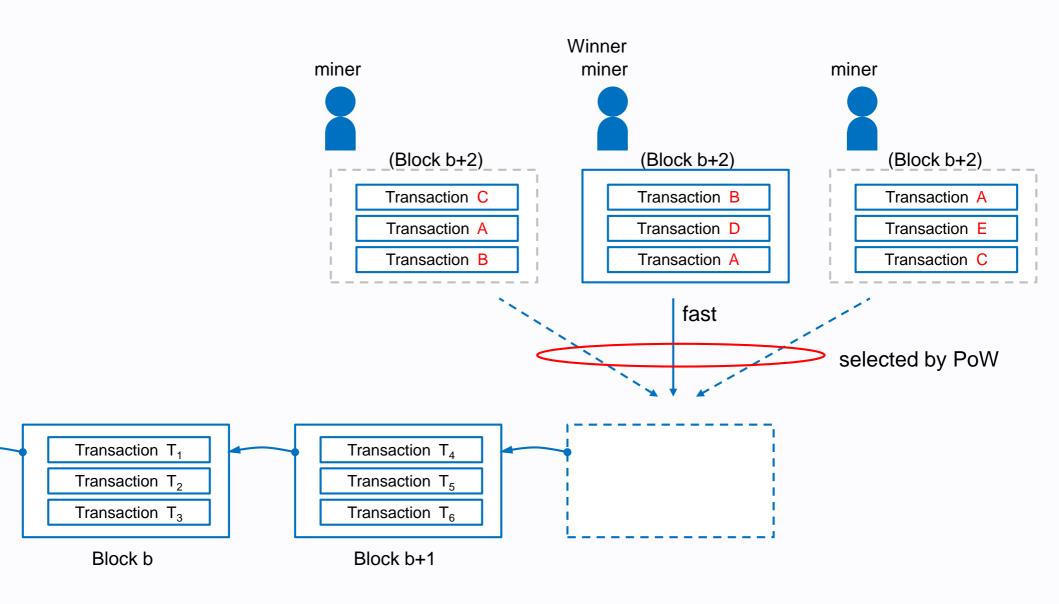
Transaction order is not guaranteed.

### Ordering inner block



Miner can determine the order of transactions in a block.

#### Ordering inter blocks

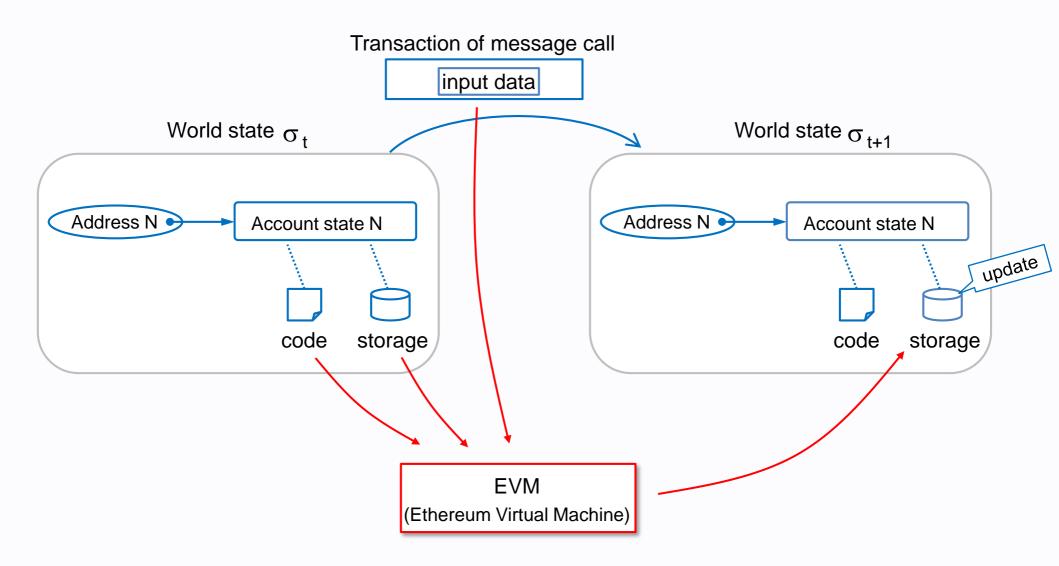


The order between blocks is determined by a consensus algorithm such as PoW.

References: [E1] Ch.2, Ch.4

Ethereum virtual machine (EVM)

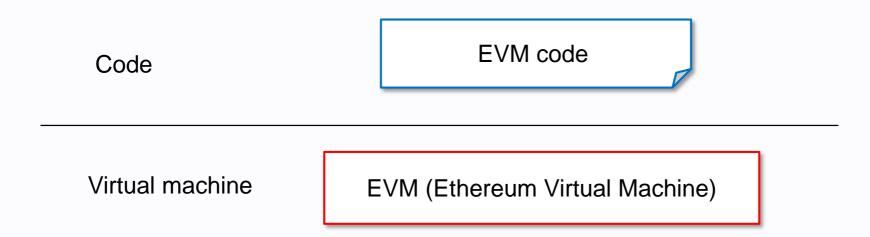
#### Ethereum virtual machine



EVM code is executed on Ethereum Virtual Machine (EVM).

References: [E1] Ch.9, Appendix H

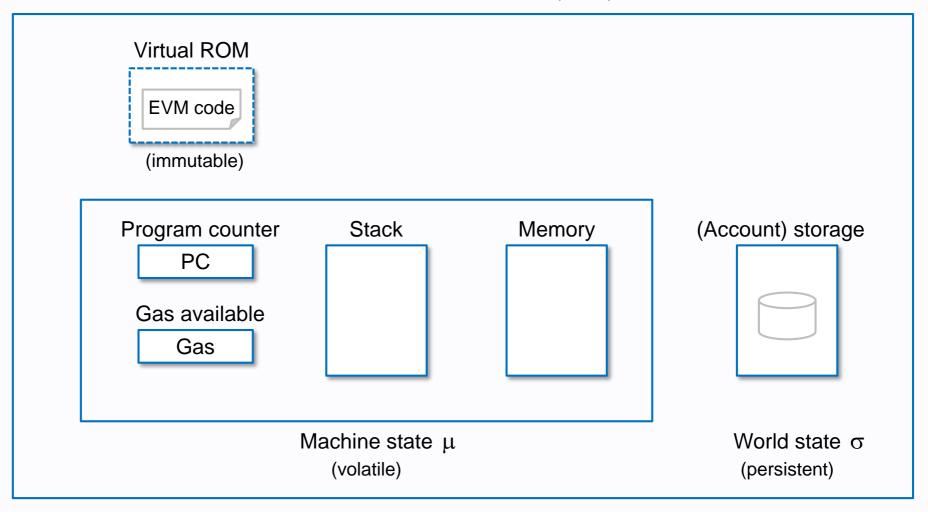
#### Ethereum virtual machine



The Ethereum Virtual Machine is the runtime environment for smart contracts in Ethreum.

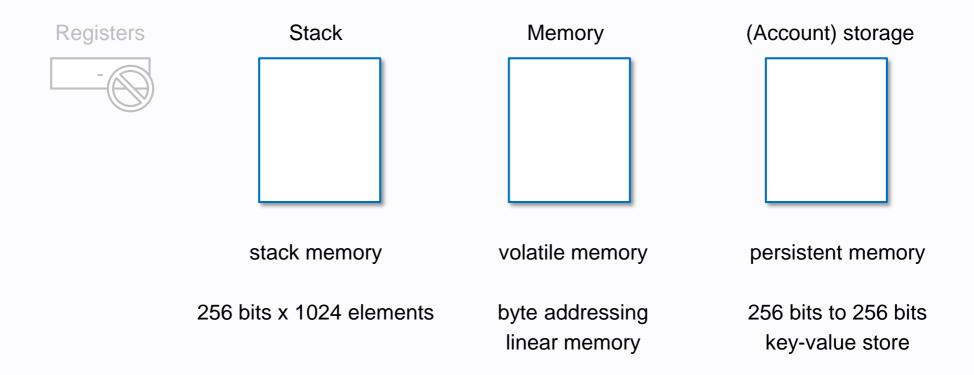
#### **EVM** architecture

#### Ethereum Virtual Machine (EVM)



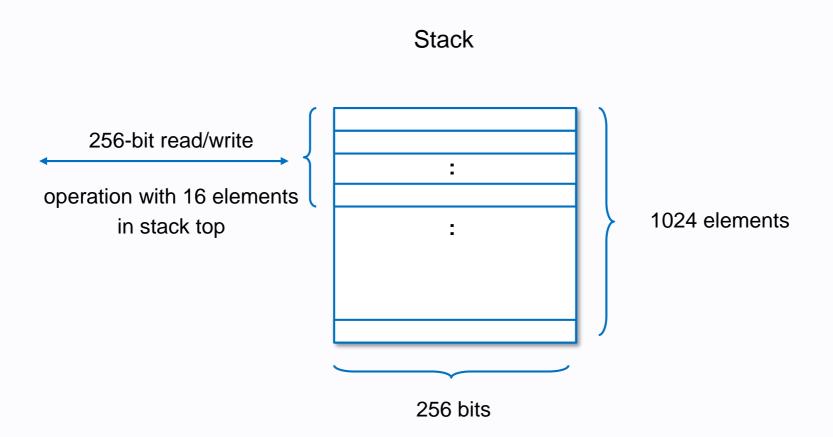
The EVM is a simple stack-based architecture.

### Machine space of EVM



There are several resources as space.

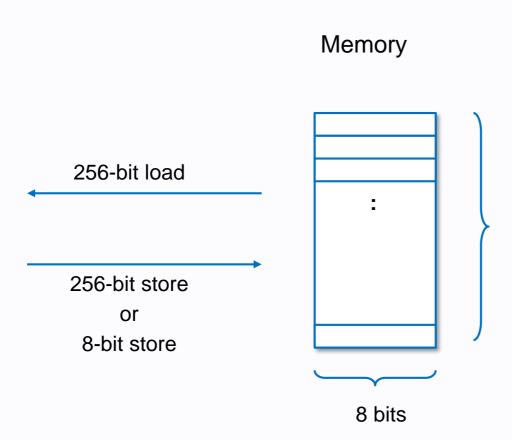
#### Stack



All operation are performed on the stack.

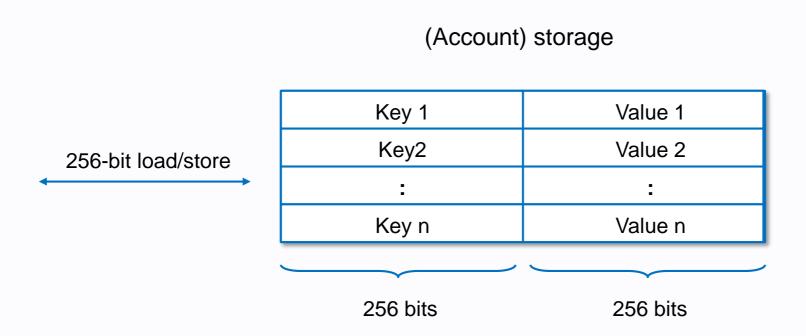
Access with many instructions such as PUSH/POP/COPY/SWAP, ...

### Memory



Memory is linear and can be addressed at byte level. Access with MSTORE/MSTORE8/MLOAD instructions. All locations in memory are well-defined initially as zero.

#### Account storage



Storage is a key-value store that maps 256-bit words to 256-bit words.

Access with SSTORE/SLOAD instructions.

All locations in storage are well-defined initially as zero.

#### **EVM** code

Assembly view

PUSH1 e0 PUSH1 02 EXP PUSH1 00

CALLDATALOAD

:

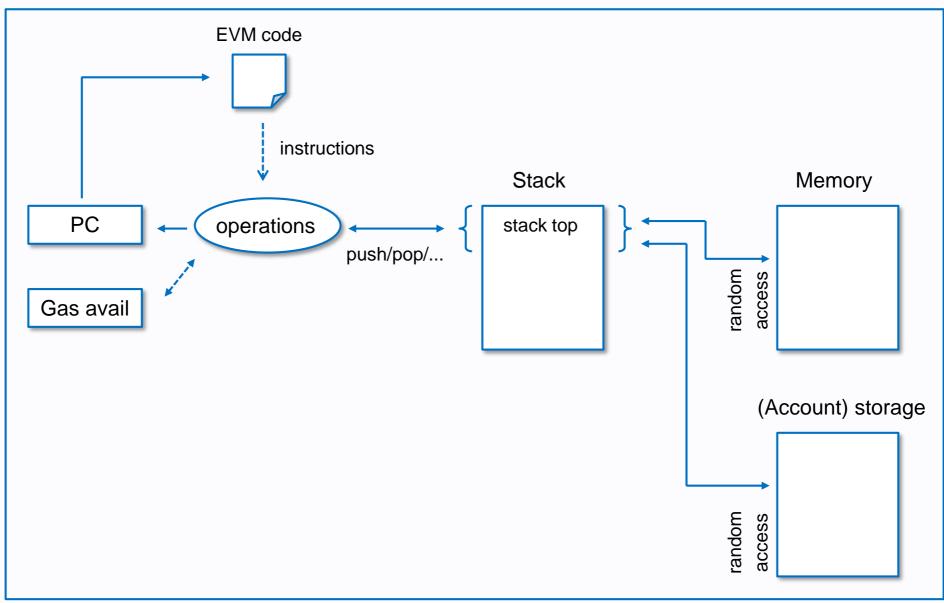
Bytecode view

0x60e060020a600035...

EVM Code is the bytecode that the EVM can natively execute.

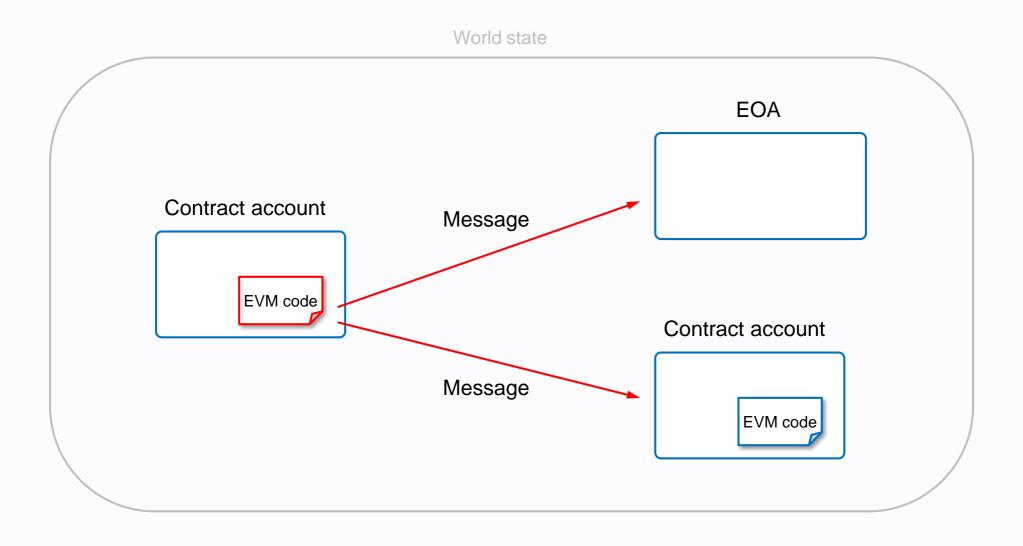
#### **Execution model**

**EVM** 



Message call

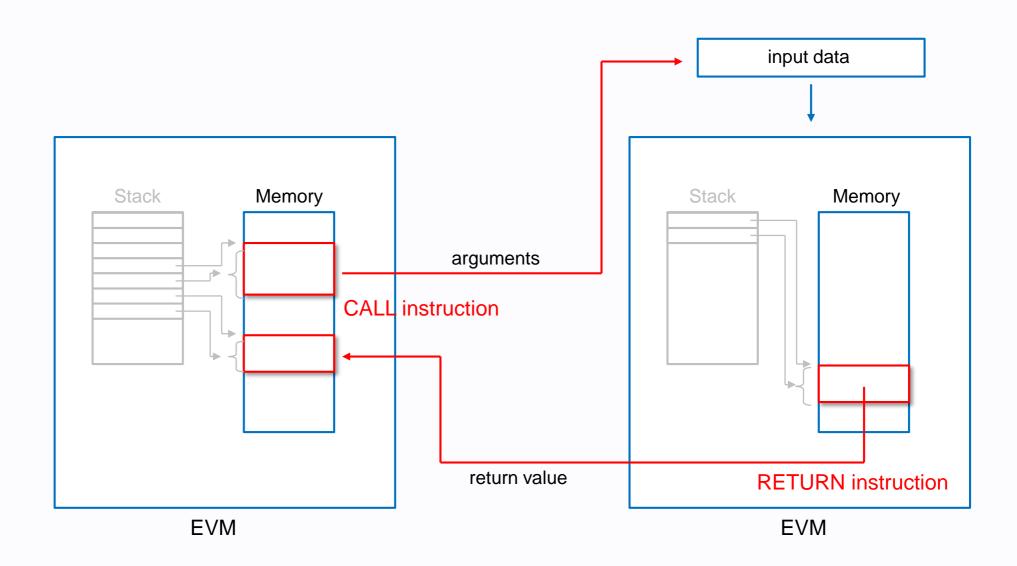
### Message call



EVM can send a message to other account.

The depth of message call is limited to less than 1024 levels.

### Instructions for Message call



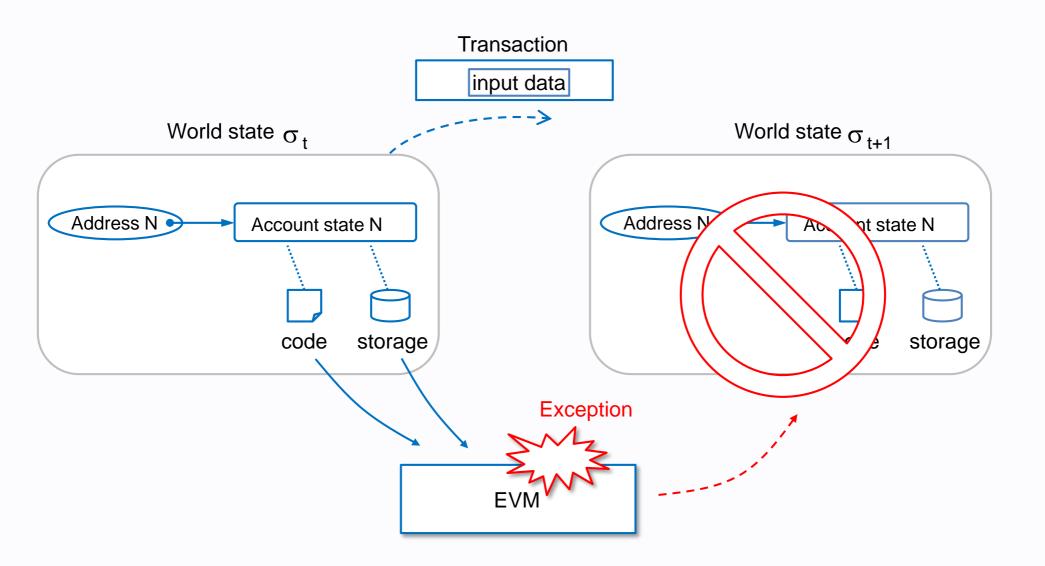
Message call is triggered by CALL instruction.

Arguments and return values are passed using memory.

References: [E1] Ch.8, Ch.9

**Exception** 

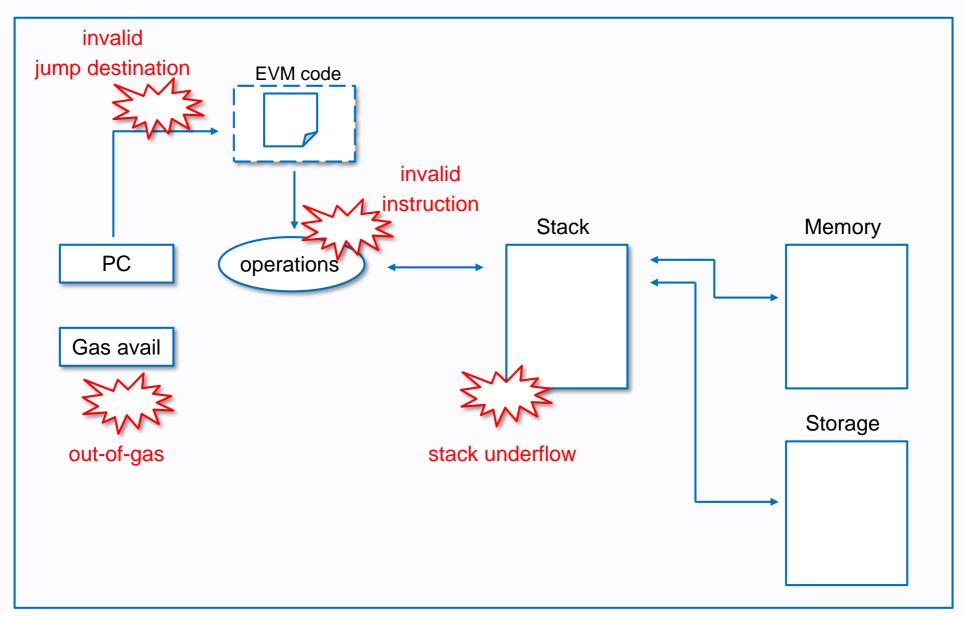
### Exception



If an exception occurs in the EVM, the state is not updated.

## **Exception**

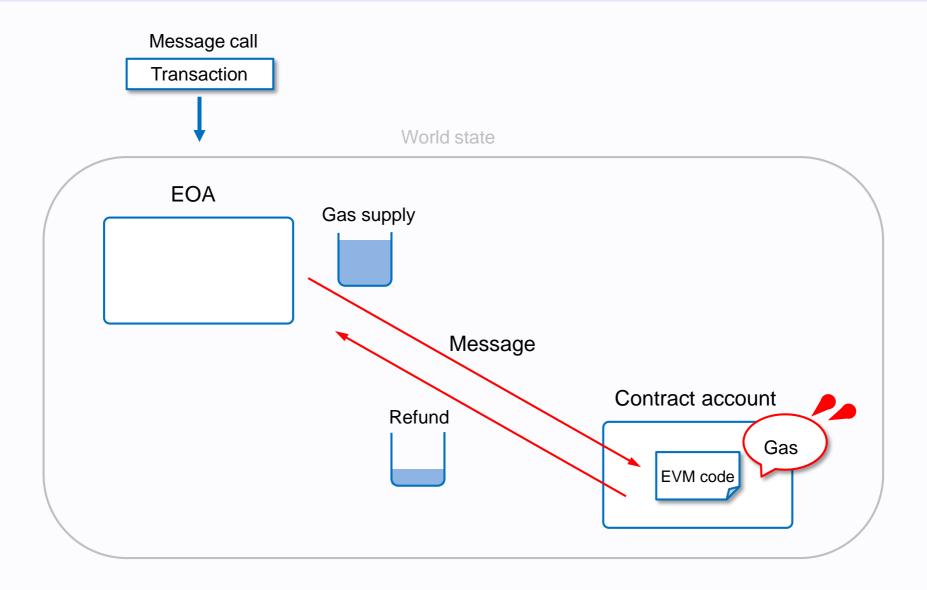
#### **EVM**



References: [E1] Ch.9, Appendix H

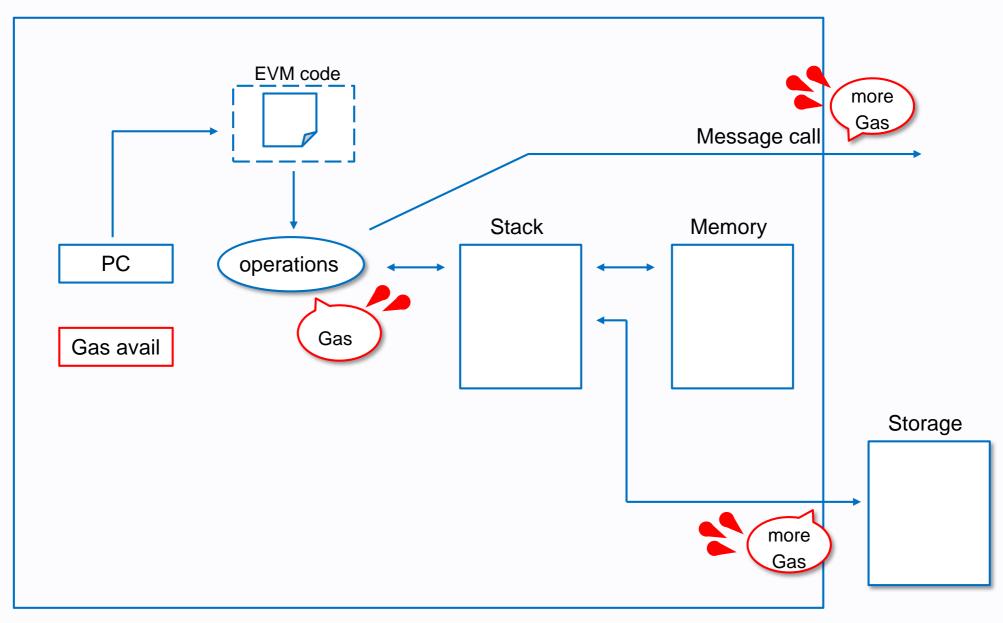
Gas and fee

#### Gas and fee



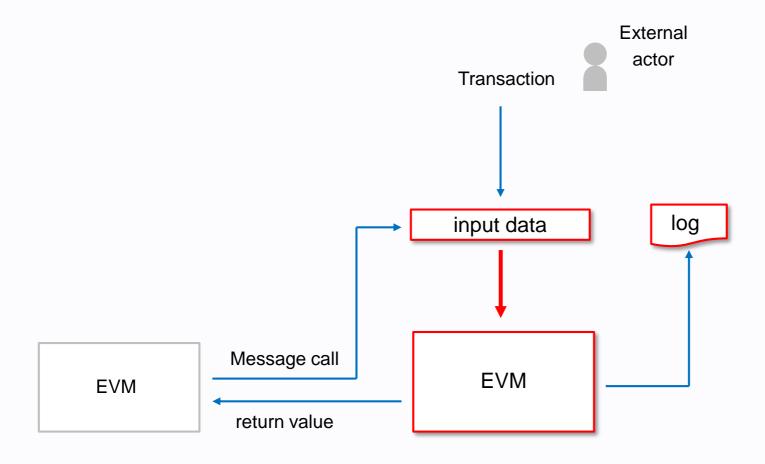
All programmable computation in Ethereum is subject to fees (denominated in gas).

EVM



Input and output

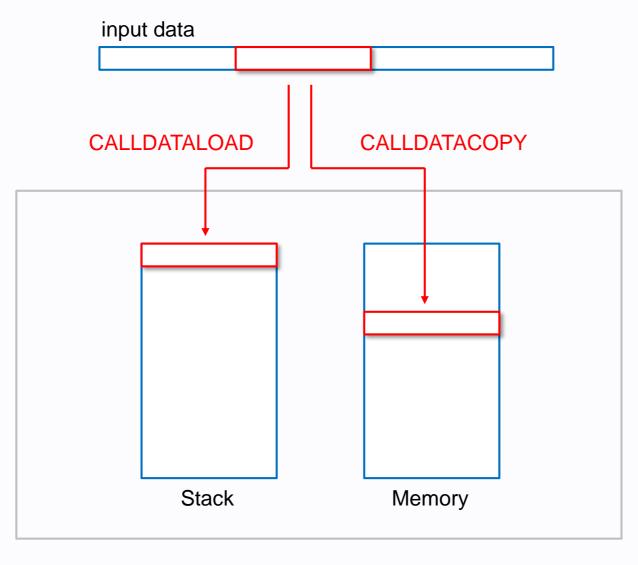
### Input and Output of EVM



EVM can input external data from a message call. EVM can output log. EVM can also return values to Caller EVM.

References: [E1] Ch.9, Appendix H

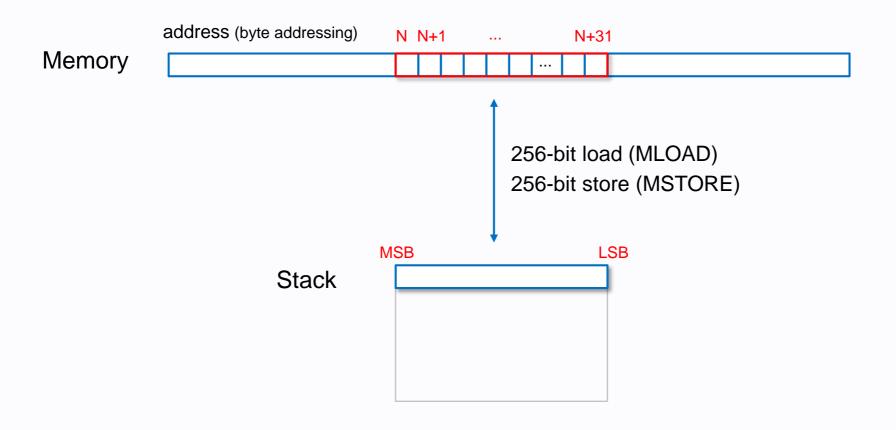
## Instructions for input data



EVM

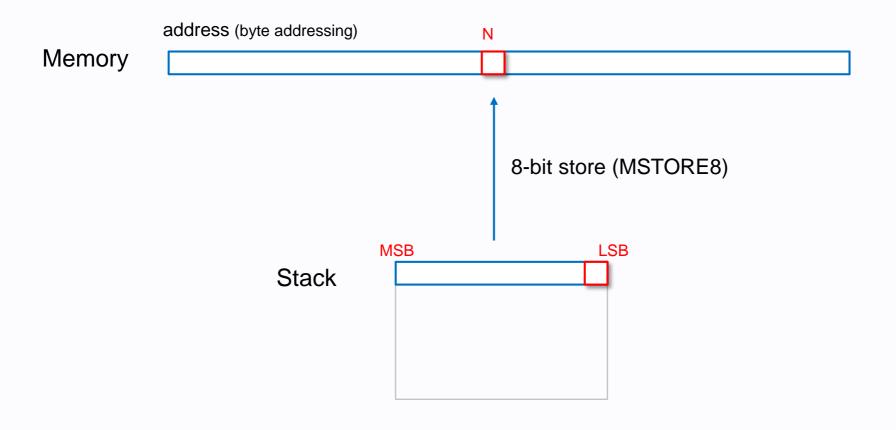
Byte order

### **Endian for Memory**



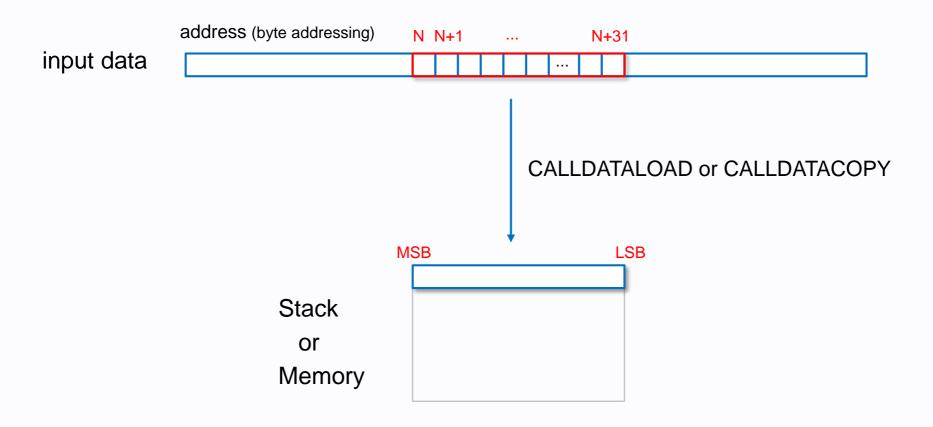
EVM is big endian order (network byte order).

### **Endian for Memory**



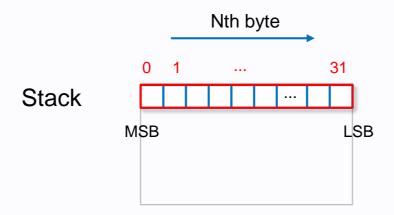
EVM is big endian order (network byte order).

### Endian for input data



EVM is big endian order (network byte order).

### Byte order of BYTE and SIGNEXTEND instruction



BYTE instruction counts from MSB.

Stack

Nth byte

31 ... 1 0

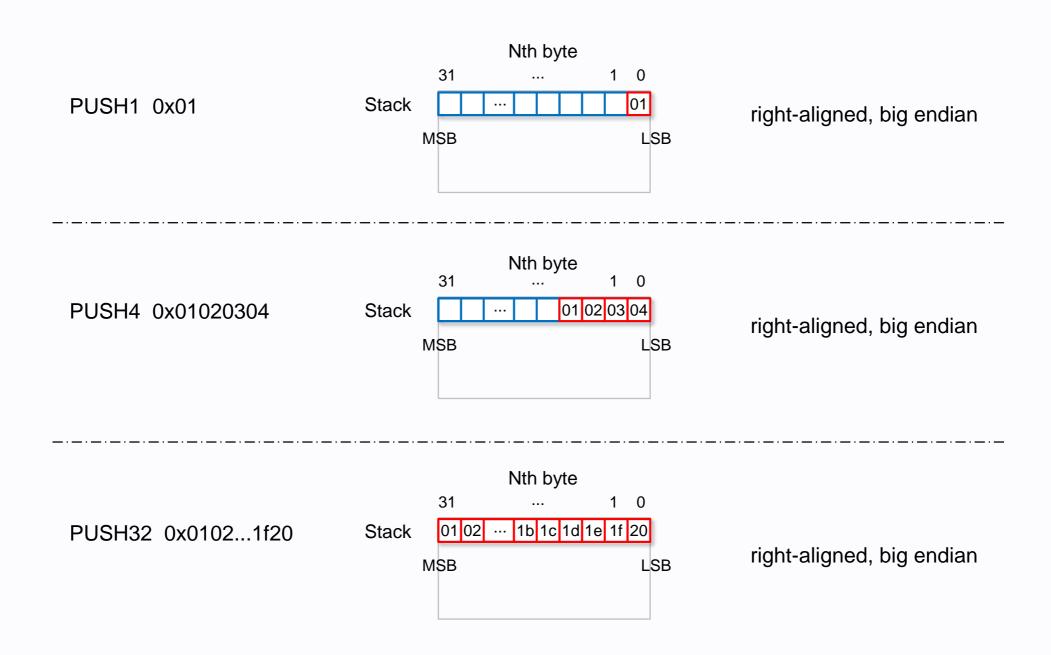
Stack

MSB

LSB

SIGNEXTEND instruction counts from LSB.

### Byte order of PUSH instructions



# 2. Virtual machine

Instruction set

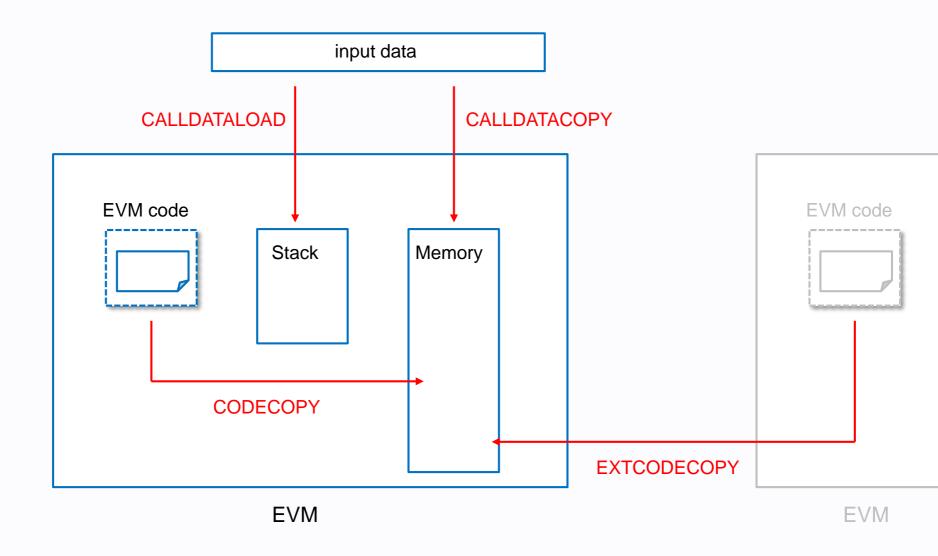
### Instruction set

- \* Basically, 256-bit operation.
- \* Contract creation and destruct
  - \* CREATE, DELEGATECALL
- \* Hash
  - \* SHA3
- \* Shift operation
  - \* using MUL or DIV, SDIV
- \* Div operation
  - \* without zero divisional exception

\*

WIP

# Copy of code and input data



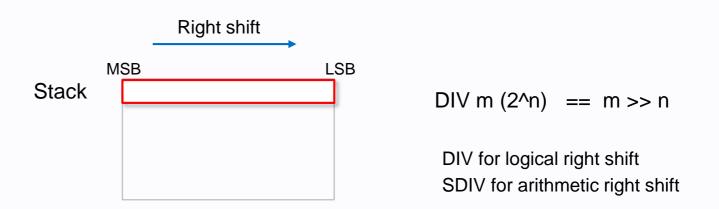
There are several copy instructions for inter spaces.

References: [E1] Ch.8, Ch.9

## Shift by MUL, DIV and SDIV



Left shift is represented by MUL instruction.



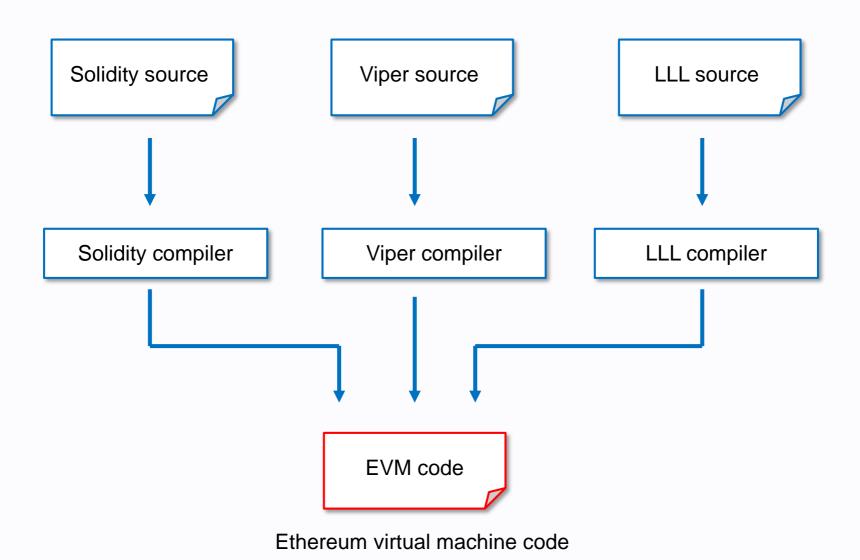
Right shift is represented by DIV and SDIV instruction.

References: [E1] Ch.9, Appendix H, [E7], [W2]

# 2. Virtual machine

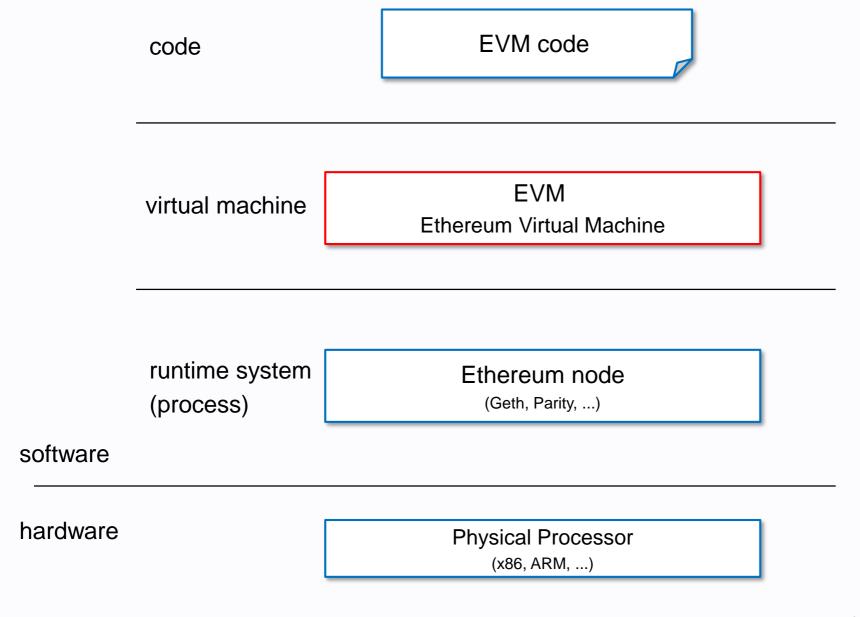
# Miscellaneous

# **EVM** code generation



References: [E7]

## Ethereum virtual machine layer



References: [E1] Ch.9

## eWASM

The eWASM is next generation VM.



# Appendix A

# Appendix A

# Source code in Geth

#### Block header

(go-ethereum version 1.8)

#### [core/types/block.go]

```
Block header
type Header struct {
                                       `json:"parentHash"
                                                                  gencodec:"required"`
         ParentHash
                      common. Hash
                                       `json:"sha3Uncles"
                                                                                    ≥d"`
         UncleHash
                      common. Hash
                                                                    Root of State
                      common.Address
                                       `ison:"miner"
                                                                  <del>geneouec. requirb</del>d"`
         Coinbase
                                       `json:"stateRoot"
                                                                  gencodec:"required"`
         Root
                      common. Hash
                                       `json:"transactionsRoot"
                                                                  gencodec:"required"`
         TxHash
                      common. Hash
                                                                                    ≱d"`
         ReceiptHash common.Hash
                                       `json:"receiptsRoot"
                                                                  Root of Transaction
         Bloom
                                       `json:"logsBloom"
                      Bloom
         Difficulty
                      *big.Int
                                       `json:"difficulty"
                                                                  gencodec: "required" `
         Number
                      *big.Int
                                       `ison:"number"
                                                                  gencodec:"required"`
                                       `json:"gasLimit"
                                                                  gencodec: "required" `
         GasLimit
                      uint64
         GasUsed
                      uint64
                                       `json:"gasUsed"
                                                                  gencodec:"required"`
         Time
                      *big.Int
                                       `json:"timestamp"
                                                                  gencodec: "required" `
                      []byte
                                       `json:"extraData"
                                                                  gencodec:"required"`
         Extra
         MixDigest
                      common. Hash
                                       `json:"mixHash"
                                                                  gencodec:"required"`
                      BlockNonce
                                       `json:"nonce"
                                                                  gencodec:"required"`
         Nonce
}
```

### **Transaction**

(go-ethereum version 1.8)

#### [core/types/transaction.go]

```
Transaction
type txdata struct {
         AccountNonce uint64
                                       `json:"nonce"
                                                         gencodec:"required"`
                      *big.Int
                                       `json:"gasPrice"
         Price
                                                         gen
                                                                  to address
         GasLimit
                      uint64
                                       `json:"gas"
         Recipient
                      *common.Address `json:"to"
                                                         rlp:"nil"`
                                                       // nil means contract creation
                                       `json:"value"
                      *big.Int
         Amount
                                                         gen
                                                                 value (Wei)
         Payload
                                       `json:"input"
                       []byte
                                                                  input data
         // Signature values
         V *big.Int `json:"v" gencodec:"required"`
         R *big.Int `json:"r" gencodec:"required"`
         S *big.Int `json:"s" gencodec:"required"`
         // This is only used when marshaling to JSON.
         Hash *common.Hash `json:"hash" rlp:"-"`
}
```

### World state

(go-ethereum version 1.8)

#### [core/state/statedb.go]

```
World state
type StateDB struct {
         db
              Database
         trie Trie
                                                                   Mapping for
                                                                    Address to Account state
         stateObjects
                           map[common.Address]*stateObject
         stateObjectsDirty map[common.Address]struct{}
         dbErr error
         refund uint64
         thash, bhash common. Hash
         txIndex
                      int
         logs
                      map[common.Hash][]*types.Log
         logSize
                      uint
        preimages map[common.Hash][]byte
```

### Account object (state object)

(go-ethereum version 1.8)

#### [core/state/state\_object.go]

```
type stateObject struct {
                                          Address
        address common.Address
        addrHash common. Hash
                                        Account state
        data
                 Account
        db
                 *StateDB
        dbErr error
        trie Trie // storage trie, which becomes non-nil on first access
        code Code // contract bytecode, which gets set when code is loaded
        cachedStorage Storage // Storage entry cache to avoid duplicate reads
        dirtyStorage Storage // Storage entries that need to be flushed to disk
        dirtyCode bool // true if the code was updated
        suicided bool
        touched bool
        deleted bool
        onDirty func(addr common.Address)
}
```

## Account state, Code and Storage

(go-ethereum version 1.8)

[core/state/state\_object.go]

```
type Account struct {

Nonce uint64

Balance *big.Int

Root common.Hash // merkle root of the storage trie

CodeHash []byte
}

type Code []byte

EVM code

type Storage map[common.Hash] common.Hash

Account storage
```

## Stack and Memory

(go-ethereum version 1.8)

#### [core/vm/stack.go]

```
type Stack struct {
    data []*big.Int
}

func newstack() *Stack {
    return &Stack{data: make([]*big.Int, 0, 1024)}
}
```

#### [core/vm/memory.go]

```
type Memory struct {
    store []byte
    lastGasCost uint64
}

func NewMemory() *Memory {
    return &Memory{}
}
```

### Instruction operation (arithmetic and stack)

(go-ethereum version 1.8)

#### [core/vm/instruction.go]

```
Arithmetic operation
func opAdd(pc *uint64, evm *EVM, contract *Contract, memory *Memory, stack *Stack)
([]byte, error) {
        x, y := stack.pop(), stack.pop()
         stack.push(math.U256(x.Add(x, y)))
        evm.interpreter.intPool.put(y)
        return nil, nil
               Stack operation
func opPop(pc *uint64, evm *EVM, contract *Contract, memory *Memory, stack *Stack)
([]byte, error) {
        evm.interpreter.intPool.put(stack.pop())
        return nil, nil
```

### Instruction operation (memory and storage)

(go-ethereum version 1.8)

#### [core/vm/instruction.go]

```
Memory operation
func opMload(pc *uint64, evm *EVM, contract *Contract, memory *Memory, stack
*Stack) ([]byte, error) {
        offset := stack.pop()
        val := new(big.Int).SetBytes(memory.Get(offset.Int64(), 32))
        stack.push(val)
        evm.interpreter.intPool.put(offset)
        return nil, nil
               Storage operation
func opSload(pc *uint64, evm *EVM, contract *Contract, memory *Memory, stack
*Stack) ([]byte, error) {
        loc := common.BigToHash(stack.pop())
        val := evm.StateDB.GetState(contract.Address(), loc).Big()
        stack.push(val)
        return nil, nil
}
```

### Instruction operation (call)

(go-ethereum version 1.8)

#### [core/vm/instruction.go]

```
Flow operation
func opCall(pc *uint64, evm *EVM, contract *Contract, memory *Memory, stack *Stack)
([]byte, error) {
       // Pop gas. The actual gas in in evm.callGasTemp.
        evm.interpreter.intPool.put(stack.pop())
       gas := evm.callGasTemp
       // Pop other call parameters.
        addr, value, inOffset, inSize, retOffset, retSize := stack.pop(),
            stack.pop(), stack.pop(), stack.pop(), stack.pop()
        toAddr := common.BigToAddress(addr)
       value = math.U256(value)
        // Get the arguments from the memory.
        args := memory.Get(inOffset.Int64(), inSize.Int64())
        if value.Sign() != 0 {
               gas += params.CallStipend
        ret, returnGas, err := evm.Call(contract, toAddr, args, gas, value)
       if err != nil {
```

(go-ethereum version 1.8)

#### [core/vm/gas.go]

```
const (
                                           \mathsf{G}_{\mathsf{base}}
                          uint64 = 2
         GasQuickStep
                                           G_{\text{verylow}}
         GasFastestStep uint64 = 3
                         uint64 = 5
         GasFastStep
         GasMidStep
                       uint64 = 8
         GasSlowStep
                        uint64 = 10
                          uint64 = 20
         GasExtStep
         GasReturn
                           uint64 = 0
         GasStop
                           uint64 = 0
         GasContractByte uint64 = 200
```

#### [core/vm/gas\_table.go]

```
func gasSStore(gt params.GasTable, evm *EVM, contract *Contract, stack *Stack, mem
 *Memory, memorySize uint64) (uint64, error) {
     var (
          y, x = stack.Back(1), stack.Back(0)
          val = evm.StateDB.GetState(contract.Address(),
          :
```

#### [core/vm/interpreter.go]

```
func (in *Interpreter) Run (contract *Contract, input []byte) (ret []byte, err
error) {
        // Increment the call depth which is restricted to 1024
                                                                increment call depth
        in.evm.depth++
        defer func() { in.evm.depth-- }()
        in.returnData = nil
        if len(contract.Code) == 0 {
                return nil, nil
        }
        codehash := contract.CodeHash // codehash is used when doing jump dest caching
        if codehash == (common.Hash{}) {
                 codehash = crypto.Keccak256Hash(contract.Code)
        }
        var (
                                      // current opcode
                       OpCode
                 qo
                                                                  create Memory
                       = NewMemory() // bound memory
                 stack = newstack() // local stack
                                                                   create Stack
```

## **ApplyTransaction**

(go-ethereum version 1.8)

[core/state\_processor.go]

```
func ApplyTransaction(config *params.ChainConfig, bc *BlockChain, author
*common.Address, gp *GasPool, statedb *state.StateDB, header *types.Header, tx
*types.Transaction, usedGas *uint64, cfg vm.Config) (*types.Receipt, uint64, error)
{
        msq, err := tx.AsMessage(types.MakeSigner(config, header.Number))
        if err != nil {
                return nil, 0, err
        // Create a new context to be used in the EVM environment
        context := NewEVMContext(msg, header, bc, author)
        // Create a new environment which holds all relevant information
        // about the transaction and calling mechanisms.
                                                                         create EVM
        vmenv := vm.NewEVM(context, statedb, config, cfg)
        // Apply the transaction to the current state (included in the env)
        , gas, failed, err := ApplyMessage(vmenv, msg, gp)
        if err != nil {
                return nil, 0, err
        // Update the state with pending changes
        var root []byte
        if config.IsByzantium(header.Number) {
```

#### Version of EVM instruction set

(go-ethereum version 1.8)

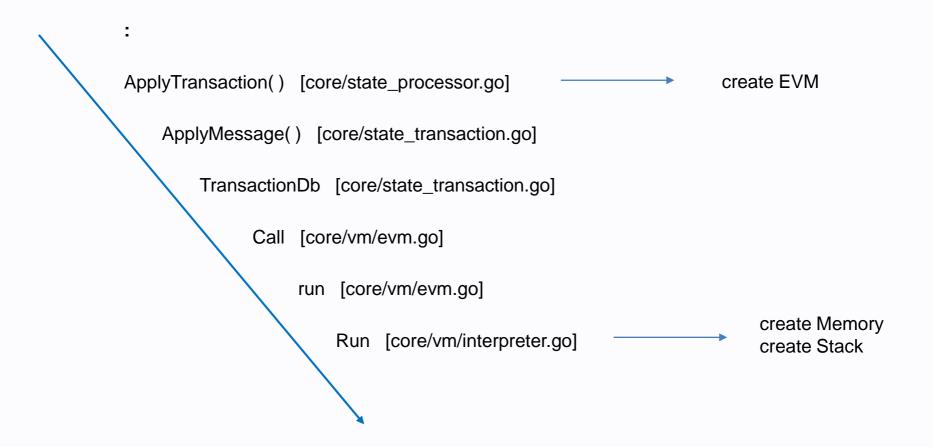
#### [core/vm/interpreter.go]

#### [core/config.go]

```
var (
MainnetChainConfig = &ChainConfig{
                ChainId:
                                big.NewInt(1),
                HomesteadBlock: big.NewInt(1150000),
                DAOForkBlock:
                                big.NewInt(1920000),
                DAOForkSupport: true,
                EIP150Block:
                                big.NewInt(2463000),
                EIP150Hash: common.HexToHash("0x2086799aeebeae135c246c65021c82b4e15a2c451340993a
                EIP155Block:
                                big.NewInt(2675000),
                EIP158Block:
                                big.NewInt(2675000),
                ByzantiumBlock: big.NewInt(4370000),
```

## Bootstrap of EVM in Geth

(go-ethereum version 1.8)



# Appendix A

EVM developer utility

## Example of evm command

(go-ethereum version 1.8)

The go-ethereum project provides evm utility command.

#### Compile EVM assembly code

```
$ cat sample.asm
push 0x1
push 0x2
add

$ evm compile sample.asm
6001600201
```

#### Disassemble EVM bytecode

```
$ cat sample.bin
6001600201

$ evm disasm sample.bin
000000: PUSH1 0x01
000002: PUSH1 0x02
000004: ADD
```

## Example of evm command

(go-ethereum version 1.8)

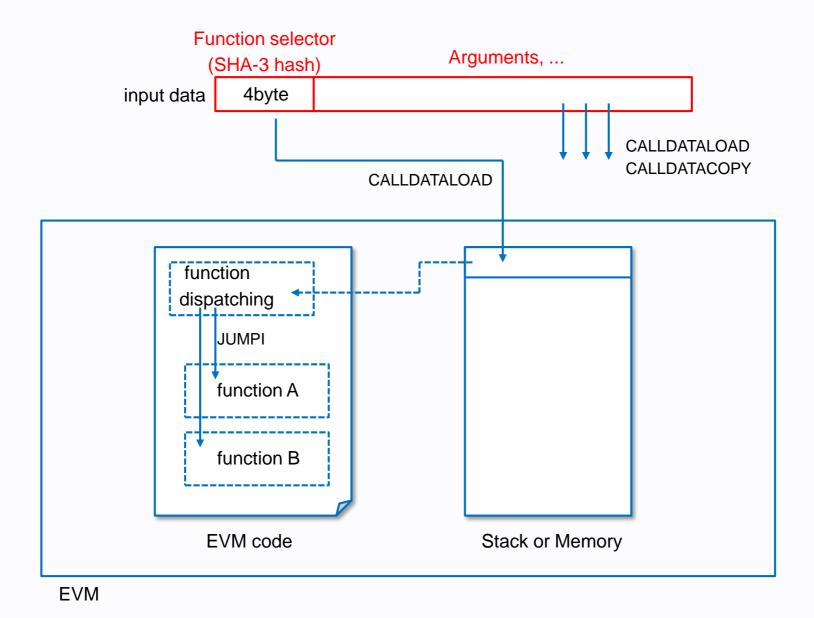
#### Run EVM assembly code

```
$ evm --debug run sample.asm
#### TRACE ####
PUSH1
        pc=00000000 gas=1000000000 cost=3
PUSH1
        pc=00000002 gas=999999997 cost=3
Stack:
00000000
     pc=00000004 gas=9999999994 cost=3
ADD
Stack:
0000000
     0000001
     pc=00000005 gas=999999991 cost=0
STOP
Stack:
0000000
     #### LOGS ####
```

# Appendix A

Solidity ABI

# Solidity Application Binary Interface



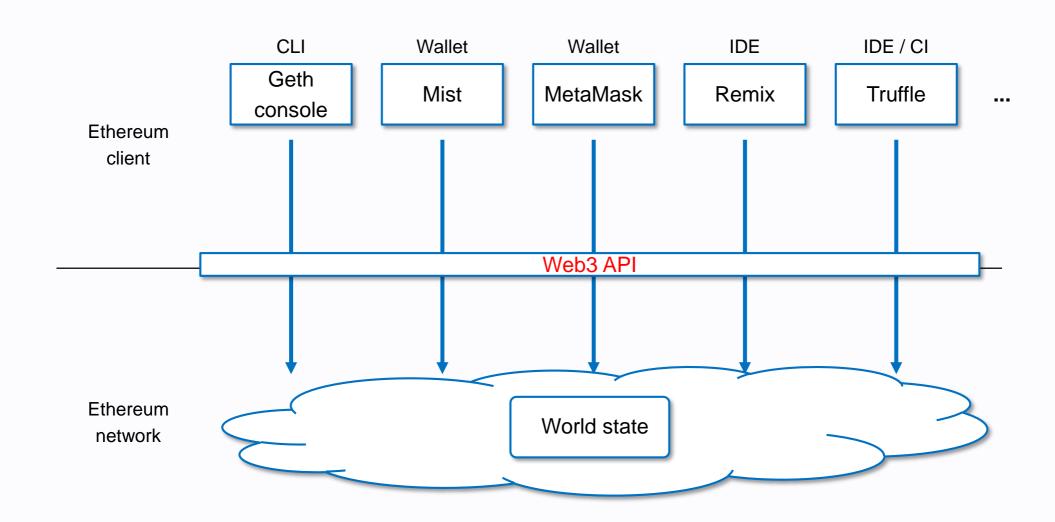
References: [E7]. Ch.7, [E1] Ch.9, Appendix H, [W4], [W2]

# Appendix B

# Appendix B

Web3 API

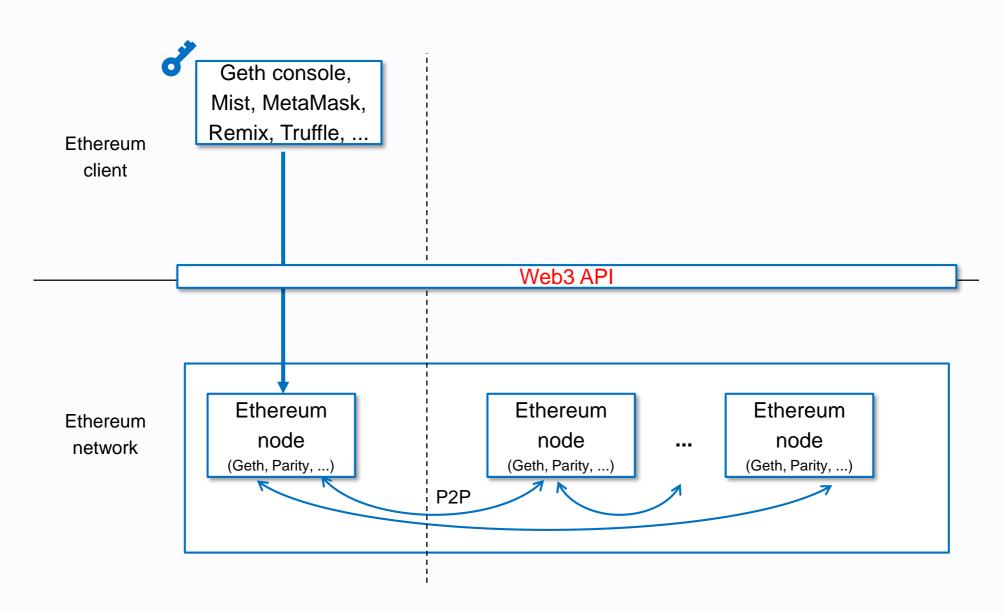
### Web3 API and client



Ethereum clients access to Ethereum network via Web3 API.

References: [E8], [C1], [C3], [C4], [C5], [C6]

### Web3 API and client



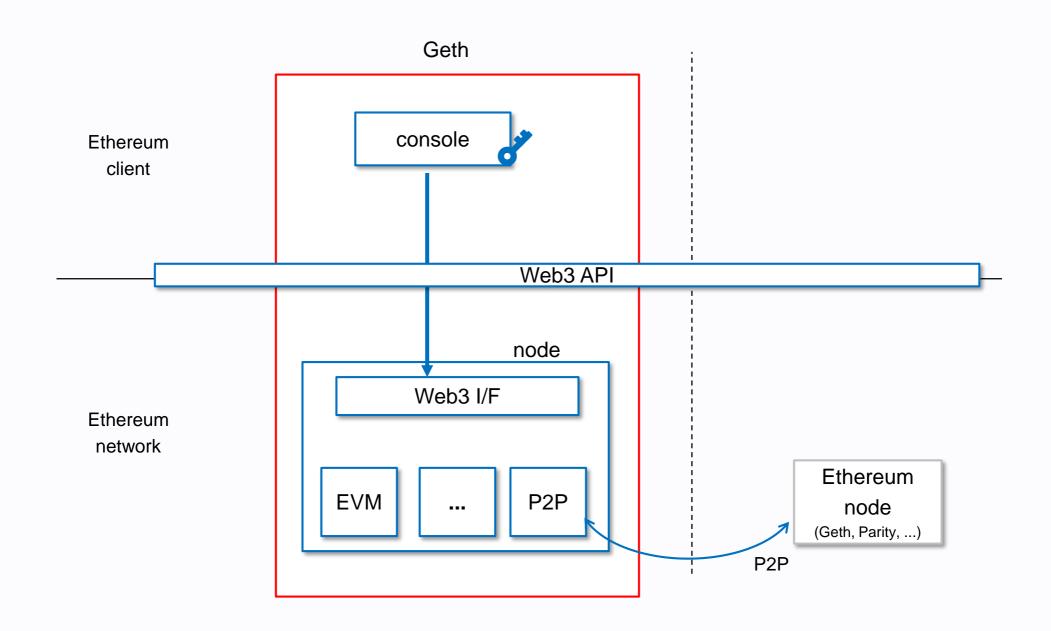
Ethereum clients access to Ethereum network via Web3 API.

References: [E8], [C1], [C3], [C4], [C5], [C6]

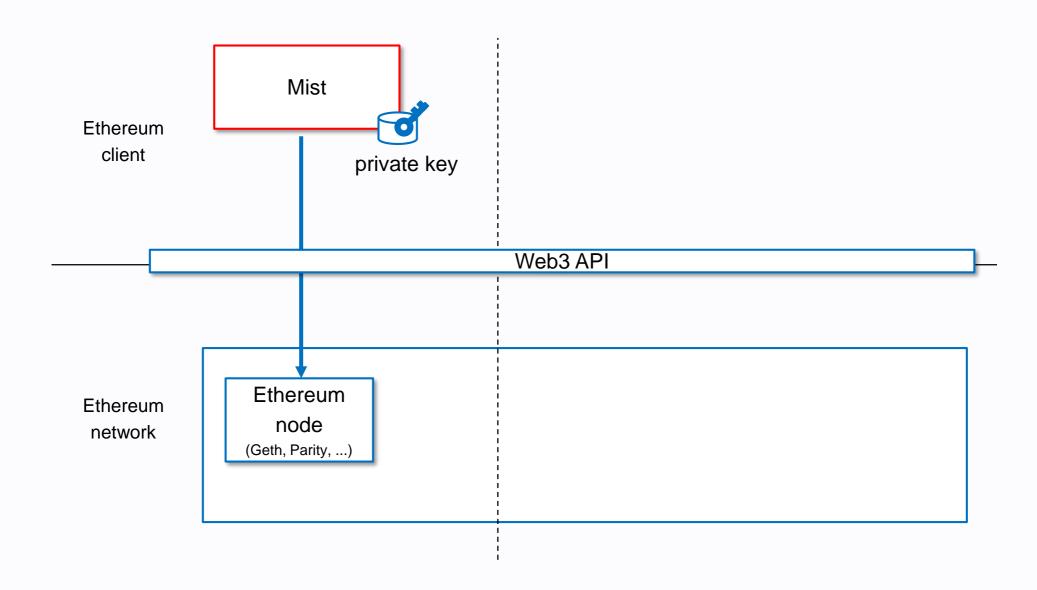
# Appendix B

Geth, Mist, Solc, Remix, Truffle, ...

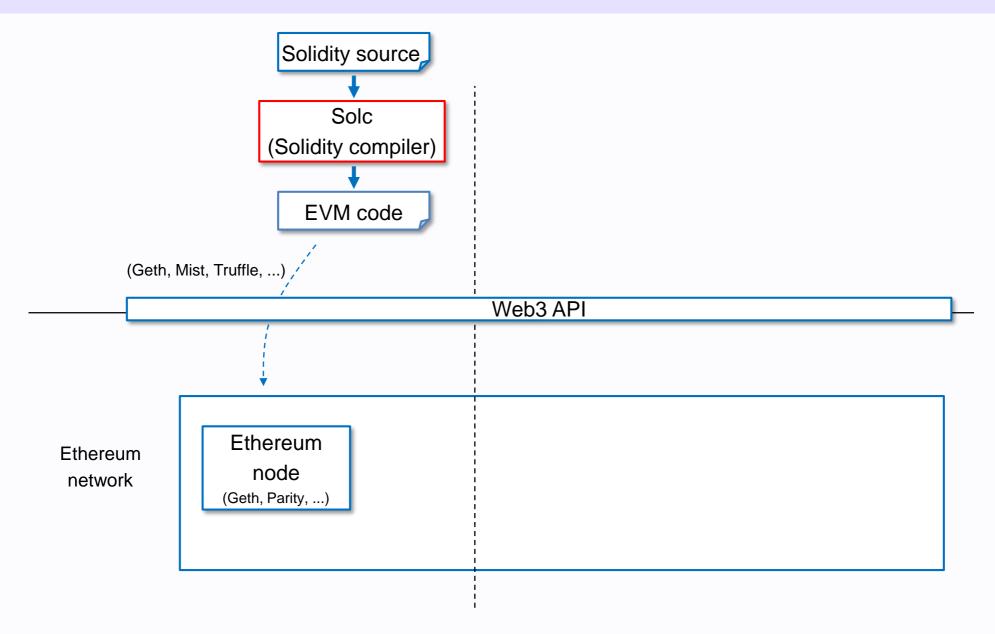
# Geth



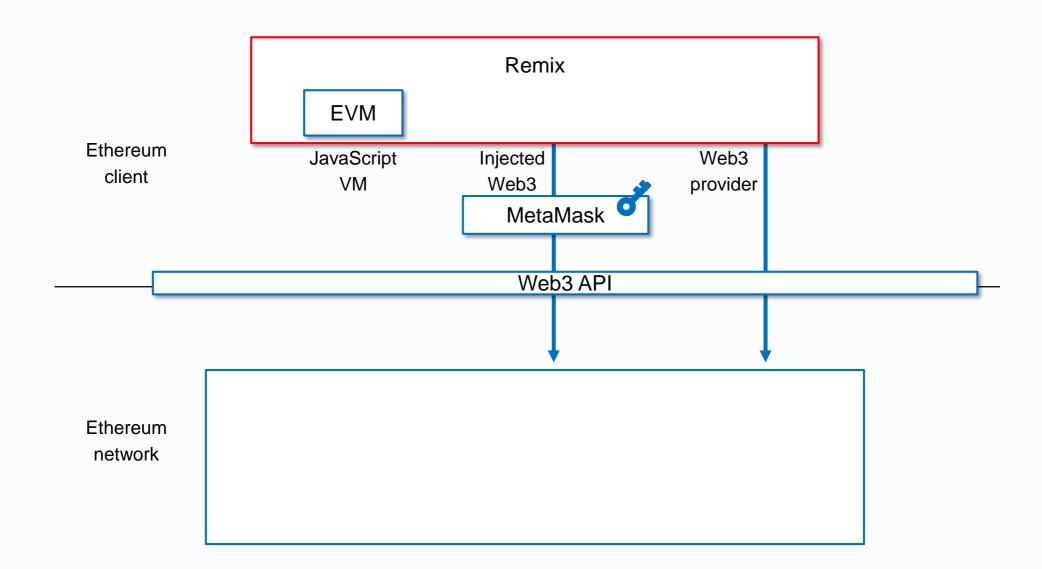
## Mist



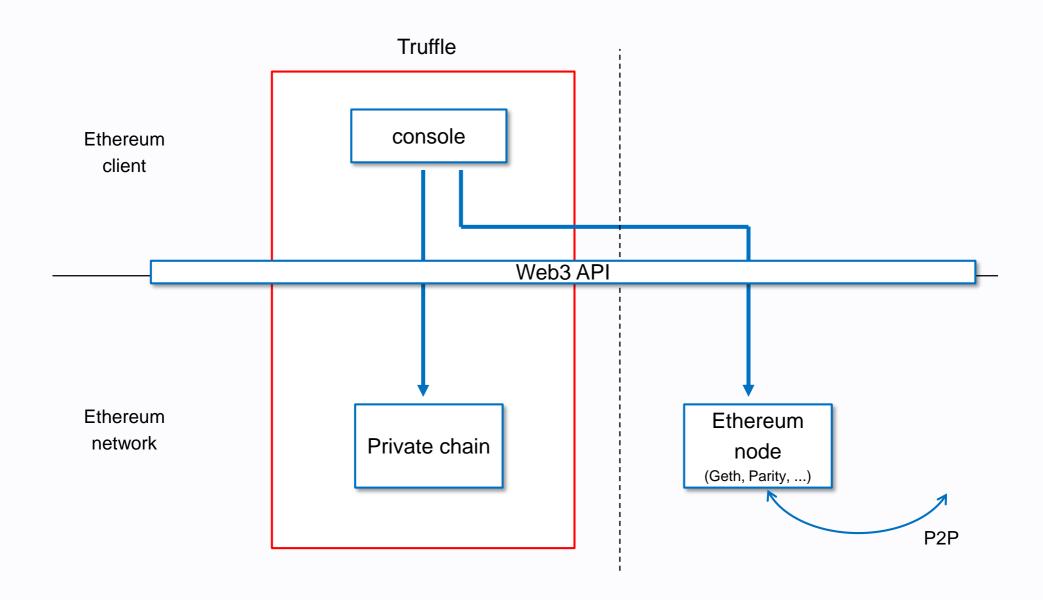
## Solc



## Remix



## Truffle



[E1]	Ethereum Yellow Paper ETHEREUM: A SECURE DECENTRALISED GENERALISED TRANSACTION LEDGER https://ethereum.github.io/yellowpaper/paper.pdf
[E2]	Glossary https://github.com/ethereum/wiki/wiki/Glossary
[E3]	White Paper A Next-Generation Smart Contract and Decentralized Application Platform https://github.com/ethereum/wiki/wiki/White-Paper
[E4]	Design Rationale https://github.com/ethereum/wiki/wiki/Design-Rationale
[E5]	Ethereum Development Tutorial https://github.com/ethereum/wiki/wiki/Ethereum-Development-Tutorial
[E6]	Ethereum Introduction https://github.com/ethereum/wiki/wiki/Ethereum-introduction
[E7]	Solidity Documentation https://media.readthedocs.org/pdf/solidity/develop/solidity.pdf https://solidity.readthedocs.io/en/develop/
[E8]	Web3 JavaScript app API for 0.2x.x https://github.com/ethereum/wiki/wiki/JavaScript-API

- [W1] Awesome Ethereum Virtual Machine https://github.com/pirapira/awesome-ethereum-virtual-machine
- [W2] Diving Into The Ethereum VM https://blog.qtum.org/diving-into-the-ethereum-vm-6e8d5d2f3c30
- [W3] Stack Exchange: Ethereum block architecture https://ethereum.stackexchange.com/questions/268/ethereum-block-architecture/6413
- [W4] Porosity https://www.comae.io/reports/dc25-msuiche-Porosity-Decompiling-Ethereum-Smart-Contracts.pdf

[C1]	Go Ethereum https://github.com/ethereum/go-ethereum
[C2]	Solc (Solidity compiler) https://github.com/ethereum/solidity
[C3]	Mist (Ethereum Wallet) https://github.com/ethereum/mist
[C4]	MetaMask https://github.com/MetaMask/metamask-extension
[C5]	Remix https://github.com/ethereum/browser-solidity
[C6]	Truffle https://github.com/trufflesuite/truffle