TLDR

- Smart contract chains are great for running an exchange: They commoditize trust, make custody, fees, prices and settlement transparent to everyone, and allow anyone to be a market maker.
- However, decentralized exchanges still lag behind CEX volume for good reason: Prices aren't competitive; execution is susceptible to abuse (MEV); and being an LP is not profitable enough.
- However, good solutions exist and will make DEXs more attractive: Pricing with oracles; slower, batch and post-tx settlements; concentrating and automated liquidity; and cheaper blockspace on L2s. We also give some examples of teams who work in this space and try to solve the issues we highlight.

Introduction

Decentralized exchanges are one of the main use cases for smart contract blockchains. People criticize them for being shitcoin casinos, but mechanisms to permissionlessly create markets for any asset hold substantial value, irrespective of whether they result in speculation or more productive activities.

Today, the crypto exchange landscape is composed of OTC, CLOB (central limit order books), RFQ, and AMM (automated market makers).

AMMs have grown to be particularly useful for creating markets on pairs where there is insufficient demand for professional market makers to participate. But CLOBs remain the venue of choice for trading volume on high demand pairs (DEXs are only ~16% of CEX volume according to Defillama, and > 60% of the DEX volume is still driven by MEV, including CEX-DEX arbitrage, see Alastor, page 16).

In this article, we identify what makes a good exchange, we highlight where DEXs are currently lacking, and propose some pathways to improve DEX designs.

What makes a good exchange?

As a trader, a good exchange should give me:

- Trust: Custody risks before, during and after my transaction should be transparent and as minimal as possible.
- Best prices: I want to trust that I get the best price on this exchange every time, or close enough to it – so that I don't need to worry about finding better prices elsewhere.
- Fairness: I don't want my orders abused. And without good reason, others shouldn't get a better price or pay lower fees than me.
- Speed and availability: Waiting on trades to go through or exchanges to open isn't fun.
- Information: The exchange helps me make informed choices and monitor my orders. I can see the prices at which my trade will likely settle, and get good limit order price and slippage suggestions. I can also see open, settled or canceled orders.
- Deep Liquidity and wide asset coverage: Seeing liquidity on many asset pairs can give me more confidence that I will get a good price on this exchange.

And liquidity providers and market makers (MMs) care about:

- Yield: Profits that justify the risk and opportunity cost of capital.

Best risk-adjusted yield is what matters for MMs; other metrics are just a means to this end. High-volume, low competition, high-spreads, good rebates, little toxic flow, last look, speedbumps, lower custody risk – can all be helpful to improve risk-adjusted yield.

Blockchains are great places to run an exchange and already give traders and market makers much of what they want: Decentralization, open source settlement mechanisms and open exchange histories are very strong foundations for trust, security, transparency and fairness.

But decentralized exchanges still struggle:

- To provide reliably good prices;
- To offer good yields for LPs;
- To prevent MEV, which breaks the promise of fair execution.

Why AMMs are so widespread

AMMs capture >95% of all DEX volume and dominate the market. Here are the most important reasons that AMMs hold the lead so far over alternative designs like traditional limit order books or RFQs:

- Low liquidity requirements: AMMs (Univ2 style) always provide a price, even with little liquidity.

- Passive liquidity: Your liquidity is managed for you in an AMM. So it’s easy to LP and anyone, not only market makers, can earn fees.
- Simplicity: AMMs require less compute and storage than order-book exchanges, so they consume less gas;
- No gatekeepers: Market maker and exchange listing fees can be prohibitively high, and centralized exchanges can delist tokens at any time. AMMs make it easy for any project to list and provide or incentivise liquidity themselves.

When the first decentralized exchanges launched, there was little liquidity, few market makers, few trades, and high gas costs. Running a CLOB under these conditions was impossible, so AMMs were a great fit. They are fairly straightforward, so more simple to build, simulate and audit.

But defi looks different now. Order volume is picking up, professional market makers quote a wide range of assets, gas is much cheaper on L2s, and everyone knows more about the weaknesses of CFMMs (i.e., most AMMs today).

AMMs are still your best bet for some markets (e.g., long-tail coins). But they lag behind centralized exchanges in key areas.

The problems of AMMs

High gas fees

Trading on-chain is still expensive. AMM pool fees (0.01–0.3%) are comparable to CEX spreads, but gas fees can easily cost you 1–10% on small trades (< $1000), even on L2s!

Stale prices

AMMs often don't give you the best price. Prices on AMMs only move through trades. So you need to rely on arbitrage traders to ensure AMM prices are up-to-date with current market prices. However, arbitrage traders are also limited by pool fees and gas, in addition to holding risk. So lower liquidity AMM pools can easily drift 1–5% from the best quote on other exchanges.

Loss-vs-rebalancing (LVR)

AMMs are passive, so if prices for the asset are decided somewhere else (e.g., on Binance), prices on the AMM will always trail behind. If prices go up (on Binance), then the AMM will sell the token to arbitrageurs too cheaply. If prices go down, the AMM will buy the token for too high a price from the arbitrageurs.

Over time the AMM, and specifically its LPs, will continue to accumulate a loss. This is a price the LPs pay to the arbitrageurs to move the price back to the market price.

In contrast, active market makers on limit order books will try to move their quotes immediately whenever prices change. And then rebalance their portfolio at market price. Hence the name, “loss-vs-rebalancing.” This is the loss that passive LPs suffer from selling at wrong prices to arbitrageurs, versus rebalancing their assets at the current market price.

This thread from Ankit gives a great example of LVR.

- LVR is permanent: Loss from LVR is not recovered if prices return to previous levels, as opposed to impermanent loss, or loss-vs-holding.
- LVR increases with volatility: The greater the price jumps, the greater the loss for LPs. In fact, loss scales quadratically with volatility.
- LVR depends on where prices are discovered: What matters is how much worse the price at which you sell to the arbitrage trader is than the current weighted mean market price.

Since you as an LP also earn fees on each trade, this acts in your favor if your pool has a big enough weight in the market. As soon as LVR < fees, these trades actually make you a profit and not a loss. Smaller pools with less weight in the market will experience higher LVR, and they will tend to pay the profits that LPs in the bigger pools make.

Extractable Value

Traders and LPs are vulnerable to value extraction on AMMs:

- Traders: Searchers can frontrun, sandwich or block your trade, and worsen your price.
- Passive LPs: More sophisticated and active LPs can provide just-in-time liquidity to take the majority of your trading fees.

Fragmented liquidity

On CFMMs, the same tokens are often paired with multiple different tokens (e.g., USDC-WBTC, DAI-WBTC, ETH-WBTC) and even multiple fee tiers for the same pair. This fragments liquidity (in this case WBTC) over multiple pools and leads to less fees for LPs, and lower depth and worse prices for traders. Much of the liquidity is not used (e.g., in Univ2 designs) for trades and even on range order AMMs the price often moves away from where liquidity is concentrated.

On centralized exchanges there is usually only a single quote asset (e.g., USD), and market makers actively keep most liquidity around the current market price This leads to far fewer pools, deeper order books, better returns for market makers, and more depth and less price impact for traders.

Further issues

The aforementioned weaknesses also give rise to more drawbacks for CFMMs:

- Price and inclusion uncertainty: Trades often fail or have slippage from trades that move the market against them.
- Fixed spread: AMMs charge a fixed spread on orders. This makes them vulnerable in highly volatile markets, and less competitive in less volatile markets.
- Difficulty attracting liquidity: Loss-vs-rebalancing and liquidity fragmentation make LPing on AMMs less profitable, and attracting liquidity more difficult. Protocols therefore often need to subsidize LPs with liquidity-mining incentives to attract adequate liquidity.
- Fragmented liquidity: On DEXs tokens are often multiple pairs for a token, even multiple fee tiers for the same pair. Much of the liquidity is not used (e.g., in Univ2 designs) for trades and even on range order AMMs the price often moves away from where liquidity is concentrated. This leads to less fees for LPs, and lower depth and worse prices for traders.

But all these issues don’t mean AMMs are doomed. Research and blockchain technology has advanced significantly and enabled new building blocks that can fix these drawbacks.

Building blocks for better on-chain exchanges

Several methods have been developed, or proposed, to fix the issues of bad prices, MEV, loss-vs-rebalancing and liquidity fragmentation. Let’s summarize the most important methods and also propose a few new ones.

Fixing high gas fees

Cheaper blockspace

Costs on L2s are one or two orders of magnitude cheaper. Tx costs therefore are now less of a bottleneck. This means more compute-intensive protocol designs, like order books, are starting to become possible. But to compete with CEXs on small swaps, gas costs probably need to come down another order of magnitude.

CoWs

Coincidence of wants (CoWs) are basically assets swapped P2P between traders who are trading complementary pairs at the same time. This improves prices as traders don't pay AMM swap fees and pay less gas (just a transfer). For them to work, however, you need good Oracles of the current best bids and offers.

CowSwap fully supports coincidence of wants (full, partial and multi-party ring trades) – settling many trades that don’t need to pay DEX fees.

Compute off-chain; verify on-chain

More sophisticated features are possible if you keep compute-intensive parts off-chain, and only use the chain for custody, settlement and verification. For example, tracking and matching limit orders off-chain, but holding custody of funds and settling trades on-chain.

Fixing stale prices

Request for Quote (RFQ)

Through an RFQ you can buy directly from a market maker. Since market makers can trade on all venues (off-chain and other chains), through them, you also have access to the prices and liquidity of those venues, even if you stay on just one chain. RFQ orders are also more gas efficient (just a transfer and signature verification, instead of routing through a pool).

Hashflow and Airswap offer easy on-chain access to RFQs.

Just-in-time (JIT) liquidity

To compensate for the risk of toxic flow, market makers don’t quote as tight and deep as they could on exchanges. Effectively normal users pay a tax to market makers to subsidize the toxic flow (arbitrageurs).

However, if you turn it around, and let market makers set the price after the user submits a trade, then the market makers can quote better prices because they take on less risk. This gives normal users a better price, and makes life harder for arbitrageurs.

This idea came from the design of ChainFlip’s JIT AMM model.

Lower DEX fees

One reason fees are high is to protect LPs against LVR. However, if a DEX can protect itself against LVR (see below) then it can also set lower fees. Lower fees keep a pool closer to other pools through arbitrage.

One way to keep prices up-to-date and also protect against unprofitable arbitrage is to set prices with Oracles.

Fixing LVR

Oracle-based pricing

As long as AMMs set their price passively, they are probably vulnerable to toxic flow. One way to avoid this is to actively update the price on the AMM – before an arbitrage trader comes in.

The oracle needs to be fast and accurate enough to leave no toxic arbitrage opportunities. An arbitrage is unprofitable (toxic) whenever fees made from the trade < price difference to market price. So to avoid toxic flow the accuracy of the oracle price needs to be smaller than the pool's swap fee.

The AMM could even set the price after the user signs the trade. This protects the LPs even more against offering outdated prices – and hence against the risk of arbitrage.

Swaap uses oracle-guided pricing to drastically reduce LVR for LPs.

Incentivized delay

Numerous problems would be solved if AMMs could differentiate between informed (potentially very unprofitable) and uninformed (on average profitable) order flow, and only keep the uninformed flow.

Trading signals decay quickly, so a long negative delay on the oracles would make it much harder for informed traders to catch the AMM off-guard. This is how it could work:

- Slow settlement is cheap: It's cheap to swap (e.g., 0.1% fee), if you can wait 5 mins for your trade to settle. The trade will settle at the price the Oracle will have in 5 mins. Uninformed traders won't mind this option, since they save on fees and waiting 5 mins costs little.
- Fast settlement is expensive: It's expensive (e.g., 0.4%) to settle at the current oracle price. A larger fee reduces the chances that an informed trader's signal advantage is big enough to be unprofitable for the AMM. And, this still gives a fast settlement option to users who are willing to pay for it.

A delay allows the DEX to separate toxic from non-toxic flows and adjust guardrails (fees) accordingly, or a DEX could simply prohibit fast settlement altogether. To be an effective block against toxic order flow, fast settlement fees have to take the pairs' market volatility into account.

An exciting example in this direction is this recent vote, where Balancer decided to reduce swap fees by 50-75% for all orderflow coming from CowSwap. CowSwap runs a batch auction, and batch auctions introduce a delay that makes them unattractive for toxic flow, so Balancer can safely reduce its fees for it, increasing the profit for it's LPs.

Active liquidity management

Concentrated liquidity positions (Uni v3) allow LPs to direct their liquidity in a specific price range. This makes it possible that LPs, or third parties, keep the liquidity around the current market price, and drastically increase capital efficiency for LPs.

Active liquidity management can even protect LPs against some of the LVR.

With a reliable oracle, an AMM can even set the liquidity around the current oracle price by itself, and therefore need no active LP management.

Maverick is successfully using this strategy to drastically improve capital efficiency for it's LPs.

Dynamic spread & volatility oracles

Since the loss for an AMM depends on the size of an arbitrageur's signal advantage, toxic order flow is more likely on more volatile pairs. In traditional order books, market makers increase their spread when markets are more volatile. AMMs could do the same and adjust their fees dynamically based on current market volatility.

Uniswap v3 already has a crude version of this by offering different fee tiers for the same pair and letting LPs choose the fee tier that fits the price volatility of the pair.

Market makers also adjust their spreads in order to rebalance their positions to their target inventory – AMMs could do something similar for their LPs.

Fixing vulnerable settlement (MEV)

Private submission

Privacy RPCs that circumvent the public mempool are one way to effectively protect txs against frontrunning and sandwiching.

Batch auctions

Batch auctions are a great way to make prices fair: You batch orders together over a period of time and all trades on the same pair execute at the same price. This reduces the chances of your trade being frontrun or sandwiched. Batch auctions also add a latency that discourages toxic order flow. Although, like negative latency oracles, batch auctions are less composable.

They also greatly improve pricing, available liquidity and routing of swaps. This mostly eliminates opportunities for backruns.

As mentioned earlier, CowSwap already runs batch auctions that lead to fairer and safer settlements for traders.

Dynamic slippage tolerance

Setting slippage is not easy. If the pair price is volatile, too little slippage can make your trade fail, and too much slippage makes you vulnerable to sandwiches. So, to avoid failed txs, DEXs often have high default slippage tolerances.

However, with volatility and depth oracles, DEX UIs can do much better and predict the right slippage for each trade. This helps the user avoid getting sandwiched or having a trade fail.

1inch already runs logic to set slippage dynamically.

Make all LPs JIT LPs

There’s also a way you can mitigate Just-in-Time (JIT) liquidity attacks: As above with the “last look” for LPs, if you change the model to determine prices after a user signs the txs, then you can allow everyone to submit their offers JIT and level the playing field. However, this only works with LPs willing to run an active strategy and able to respond to every trade individually.

Structurally, batch auctions are also just-in-time liquidity exchanges – since solvers find liquidity and prices after the user submits their tx.

Summary

While chains are an excellent infrastructure for exchanges, DEXs don’t yet handle the majority of exchange volume. However, there are good reasons why trading volume and market makers have not yet moved fully on-chain: Prices are uncompetitive, user experience is bad for traders, yields are too low and execution is unsafe. Fortunately, good fixes already exist for all of them. Together, they might be enough to bring the majority of trading volume on-chain.