

The End of Float: Settlement Latency, Money Velocity, and the Economics of Programmable Cash

ACM Global Tech

Abstract

Float—the value held in transit while a payment or trade clears—is not a feature of money. It is an artifact of settlement latency. Wherever the legal finality of a transfer lags its initiation, value accumulates in the gap, and that gap is simultaneously a source of intermediary revenue and a reservoir of counterparty and systemic risk. This paper treats settlement latency as the independent variable. We begin from Fisher’s equation of exchange, $MV = PT$, and argue that the velocity term V has historically been bounded not by the willingness of agents to transact but by the operational and legal latency of the rails beneath them. As settlement latency approaches zero and money becomes programmable, two things happen at once: effective velocity rises, and velocity ceases to be a passive aggregate and becomes a *design parameter*—an issuer or treasurer can specify the conditions under which value moves. We then examine how atomic settlement and tokenized bank liabilities reshape the money multiplier and the “singleness of money,” how float-based business models give way to intraday liquidity management, and what this implies for monetary transmission. We connect the argument to the Bank for International Settlements’ unified-ledger thesis and close with a brief, vendor-light note on where instantly-settled, programmable rails fit. The contribution is analytical: a framework for pricing the end of float, not a forecast of when it arrives.

1 Float Is an Artifact of Settlement Latency

Consider what it means for a payment to “clear.” An instruction is given at time t_0 ; legal finality—the moment the transfer becomes irrevocable and the recipient holds good funds—arrives at t_1 . The interval $\Delta = t_1 - t_0$ is settlement latency. During Δ , the underlying value belongs, economically, to no one and to everyone: the payer has parted with it, the payee cannot yet use it, and an intermediary typically holds the corresponding balance. That held balance is *float*.

Float is therefore a direct function of latency. If $\Delta = 0$ —if initiation and finality coincide—there is no interval in which value can be held in transit, and float is identically zero. Every dollar of float in the financial system can be traced to some non-zero Δ somewhere in the chain of intermediation. This reframing matters because float is usually discussed as a balance-sheet line item or a revenue stream, as though it were intrinsic to payments. It is not. It is the monetized shadow of a delay.

Three habitats of float illustrate the point.

Card networks.

Between authorization and settlement, an acquirer, network, and issuer sit in sequence, each with its own clearing cycle. Merchant funds are typically available on a $T + 1$ or $T + 2$ basis. The interval funds float through the network, and the economics of the system depend on it.

Correspondent banking.

A cross-border payment routed through correspondent and nostro/vostro relationships may traverse several institutions across time zones. Each leg introduces latency and a pre-funded balance held in advance of settlement. The aggregate of these pre-funded nostro balances is float held precisely because settlement is not atomic across the chain.

Securities settlement.

Major equity markets settle on a $T + 1$ basis; many other instruments and cross-border trades settle on $T + 2$ or longer. In the gap between trade and settlement, the buyer owes cash it has not paid and the seller owes securities it has not delivered. The exposure created by that gap is the reason central counterparties post margin and is the channel through which a single default can propagate.

The securities case makes the dual nature of float explicit. The settlement interval is simultaneously *revenue* for those who earn on balances and fees within it, and *risk*—specifically counterparty and replacement-cost risk—borne by everyone exposed to a default before finality. Float is not free money; it is money rented from a queue, and the rent is paid in systemic fragility. Any analysis that books the revenue without booking the risk has mispriced the position.

2 Velocity as a Design Parameter

Fisher’s equation of exchange states that

$$M V = P T, \tag{1}$$

where M is the quantity of money, V its velocity of circulation, P the price level, and T the volume of transactions over a period [1]. Read as an identity, the equation says only that total spending equals total receipts. Read as a relationship, it invites the question that this section pursues: *what bounds V ?*

Velocity is, dimensionally, transactions per unit of money per unit of time. A single unit of money can be spent, received, and respent; the faster each holder can pass it on, the higher V . The classical literature treats V as governed by institutional habits—pay cycles, the demand for money, the structure of intermediation. We add a sharper constraint: *the maximum velocity of a unit of money is bounded by the latency of the rails on which it settles*. A dollar cannot be respent until the recipient holds it with finality. If finality takes Δ to arrive, then no holder can recirculate that dollar more often than once per Δ , and the system-wide velocity is capped accordingly.

To make the bound concrete, consider a single unit confined to one rail with settlement latency Δ and incremental holder dwell time τ (the time a holder chooses to keep the funds before spending again). The minimum time between successive respendings is $\Delta + \tau$, so the unit’s annualized turnover satisfies

$$V_{\text{unit}} \leq \frac{1}{\Delta + \tau} \text{ (periods per year)}. \tag{2}$$

When dwell time is short relative to settlement latency ($\tau \ll \Delta$), velocity is *latency-bound*: the rail, not the willingness to transact, sets the ceiling. As $\Delta \rightarrow 0$, the ceiling rises to $1/\tau$, and velocity becomes *preference-bound*—governed by how long agents choose to hold money, which is an economic decision rather than a plumbing constraint.

The more consequential change is qualitative. When money is *programmable*—when transfers can carry conditions enforced at settlement—velocity stops being a passive aggregate that we measure after the fact and becomes a parameter that an issuer or treasurer can *set*. A programmable

transfer can be made to execute only when a delivery is confirmed, only against simultaneous receipt of another asset (atomic delivery-versus-payment), only within a time window, only up to a rate limit, or only to a counterparty that satisfies an on-chain compliance check. Each of these is a condition on *when* and *whether* value moves. Collectively they let the holder shape the realized velocity of a balance rather than inherit it from the rail.

This is the substantive thesis of the paper: as $\Delta \rightarrow 0$, velocity rises toward its preference-bound ceiling *and* becomes *conditionable and schedulable*. Two treasurers holding identical balances on the same instant-rail can run them at deliberately different velocities—one sweeping idle cash every minute against tightly scoped conditions, the other holding it static—because the conditions of movement are now under their control. Velocity, in a programmable-cash regime, is less a property of the money supply than a policy chosen by each holder, aggregated up.

3 The Money Multiplier Under Tokenized Deposits vs. Stablecoins

Compressing Δ does not merely speed up existing money; it changes how money is created and what counts as money. Two tokenized forms must be distinguished because they sit on opposite sides of the bank balance sheet.

A *tokenized deposit* is a commercial-bank liability represented on a programmable ledger. It remains a claim on the issuing bank, fungible at par with that bank’s other deposits and within the deposit-insurance and prudential-supervision perimeter. A *stablecoin*, by contrast, is typically a claim on a non-bank issuer, backed by reserve assets held off the issuer’s balance sheet relative to the banking system, and outside the deposit framework. The distinction governs how each interacts with credit creation.

In the textbook fractional-reserve account, banks create deposits by lending, and the multiplier links base money to broad money through the reserve and currency ratios. Tokenization on a shared ledger does not, by itself, abolish this mechanism: a tokenized deposit is still a deposit, and a bank issuing it can still extend credit against it. What atomic, instant settlement changes is the *velocity of reserves* and the precision of liquidity management. When interbank settlement is atomic, banks need hold smaller precautionary buffers against settlement timing risk, and reserves can be redeployed intraday with a granularity that the daily settlement cycle never permitted. The multiplier is not repealed; its frictional slack is squeezed.

Stablecoins interact differently. To the extent a fully reserve-backed stablecoin draws balances out of fractional-reserve deposits and parks them in non-lent reserve assets, it shifts funds from an instrument that supports credit creation to one that does not, with consequences for the supply of bank- intermediated credit. The recent regulatory regimes sharpen this. The U.S. GENIUS Act establishes a federal framework for payment stablecoins with reserve and disclosure requirements [3], and the EU’s Markets in Crypto-Assets Regulation (MiCA) imposes reserve, redemption, and authorization requirements on asset-referenced and e-money tokens [4]. Both regimes, by design, push stablecoins toward high-quality liquid reserves rather than maturity transformation—which is prudent for the holder and precisely why the instrument does not multiply credit the way a deposit does.

This is where the *singleness of money*—the principle that a dollar is a dollar regardless of which institution issues the claim, so that all forms exchange at par on demand—becomes the binding concern. A monetary system works because par convertibility is taken for granted: a deposit at one bank, a banknote, and a balance at another bank are mutually interchangeable at 1:1. A proliferation of privately issued tokens that do *not* reliably trade at par —because their

backing, redemption rights, or issuer credit differ—fragments this singleness and reintroduces the wildcat-banking problem of discounting one issuer’s money against another’s. The BIS frames the resolution as anchoring tokenized money in central-bank money and supervised bank liabilities so that settlement occurs in a common, par-preserving unit [2]. The analytical upshot: tokenized *deposits* preserve singleness because they remain bank liabilities settling in central-bank money; *stablecoins* preserve it only insofar as regulation enforces par redemption robustly enough to make the par assumption credible under stress.

4 The Collapse of Float-Based Models and the Rise of Intraday Liquidity

If float is the monetized shadow of latency, then driving $\Delta \rightarrow 0$ extinguishes the shadow—and with it any business model that capitalized the delay. Revenue earned on funds in transit, on the spread between when a payer is debited and a payee is credited, and on the pre-funded balances that correspondent chains require, all compress toward zero as settlement becomes instant and atomic. This is not a marginal efficiency gain for incumbents; for some intermediaries it is the removal of a primary revenue line.

The mirror image of vanishing float is the rise of *intraday liquidity management* as a first-order discipline. Under daily batch settlement, liquidity is managed at end-of-day granularity, and the float in the system silently absorbs intraday timing mismatches. When settlement is continuous and final, those mismatches surface in real time: an institution must have the right funds, in the right place, at the right second, because there is no settlement lag to borrow against. Liquidity that was previously held as a static buffer becomes a flow to be actively scheduled. Programmability is what makes this tractable—conditional, time-windowed, and rate-limited transfers are the instruments with which a treasurer choreographs intraday flows rather than merely funding an end-of-day position.

The implications for *monetary transmission* follow from the same shift. Transmission—the passage of a policy-rate change through money markets into broad financing conditions—operates through the repricing and reallocation of short-term funds. If those funds move and settle instantly and conditionally, the operational lags in transmission shorten, and the central bank’s footprint in intraday liquidity provision plausibly grows, because the system’s tolerance for timing mismatch falls as float disappears.

5 Policy and Infrastructure Implications

The framework above is rail-agnostic, but it points to a concrete infrastructure question: on what substrate can settlement latency collapse without sacrificing the singleness of money or the supervisory perimeter? The BIS answer is the *unified ledger*—a common programmable platform on which tokenized central-bank money, tokenized commercial-bank deposits, and tokenized assets coexist, so that conditional, atomic settlement (delivery-versus-payment and payment-versus-payment) can occur natively while par convertibility is anchored in central-bank money [2]. In that design, programmability and singleness are not in tension: tokenized deposits remain supervised bank liabilities, and final settlement still occurs in the common unit. The policy task is to extend the existing institutional guarantees—par convertibility, deposit insurance, supervision, and a stablecoin regime such as GENIUS [3] and MiCA [4]—onto rails where Δ is small, rather than to let those guarantees lapse simply because settlement got faster.

ACM Global Tech operates at this layer. ACM provides compliant, instantly-settled, programmable settlement rails—built on the Lux Network’s regulated ledger together with the Hanzo and Zen infrastructure—on which tokenized deposits and regulated digital assets can settle atomically under on-chain compliance conditions [5]. The relevance to this paper is narrow and deliberate: such rails are one instantiation of the near-zero- Δ , programmable substrate the analysis describes. The argument stands independently of any particular vendor; the rails are an existence proof, not the thesis.

Closing. Float was never a property of money. It was the price the economy paid for slow settlement, collected as intermediary revenue and charged back as systemic risk. Compress settlement latency to zero and that price goes to zero with it—ending the float, raising the ceiling on velocity, and, for the first time, making velocity something a holder can *set* rather than merely suffer. The open questions are not whether this is technically possible but how monetary institutions choose to preserve the singleness of money and the transmission of policy in a world where the queue—and the rent it charged—has disappeared.

References

- [1] I. Fisher. *The Purchasing Power of Money: Its Determination and Relation to Credit, Interest and Crises*. The Macmillan Company, New York, 1911. (Statement and analysis of the equation of exchange $MV = PT$.)
- [2] Bank for International Settlements. *Blueprint for the Future Monetary System: Improving the Old, Enabling the New*. BIS Annual Economic Report 2023, Chapter III. Bank for International Settlements, Basel, 2023.
- [3] United States Congress. *Guiding and Establishing National Innovation for U.S. Stablecoins Act* (GENIUS Act). Federal framework for the issuance and reserve backing of payment stablecoins.
- [4] European Union. *Regulation (EU) 2023/1114 of the European Parliament and of the Council on Markets in Crypto-Assets* (MiCA), and amending Regulations (EU) No 1093/2010 and (EU) No 1095/2010 and Directives 2013/36/EU and (EU) 2019/1937. Official Journal of the European Union, 2023.
- [5] Z. Kelling. *The Lux Network: A Regulated, Post-Quantum Blockchain for Digital Securities, Real-World Assets, and Compliant Global Payments*. Lux Industries, Inc. Source: `lux:lux-whitepaper/lux-whitepaper.tex`.