

Drafting in Self-Timed Circuits

Christopher Cowan^{1b}, *Member, IEEE*

Abstract—Intervals between data items propagating in self-timed circuits are controlled by handshake signals rather than by a clock. In many self-timed designs, a trailing data item will catch up with a leading item or token, even when it trails by thousands of gate delays. This effect, called “drafting,” can be seen in many of the self-timed designs, e.g., GasP, Mousetrap, Click, and Micropipeline. Drafting occurs because the delay of a trailing token through a self-timed stage depends on how much earlier the leading token departed. Contrary to earlier work, we find the cause of drafting to be charge stored on an isolated node between two series transistors. This mechanism occurs in many decision gates that implement a logical AND. The charge on the floating internal node can drift between actions and thereby change the delay of the gate. Drafting behavior may be modulated by controlling the internal node of the GasP NOR gate. This offers possibilities for using self-timed circuits in applications where the interval between data items carries information, for instance, spiking neural networks, security, or real-time signal processing.

Index Terms—Click, drafting, GasP, Micropipeline, Mouse-trap, physically unclonable devices (PUF), self-timed circuits, spiking neural network.

I. INTRODUCTION

SELF-TIMED circuits move data through pipelines using handshake signals rather than a global clock. Data and the handshake usually move together as bundled data. In our investigation, we are concerned only with the handshake signals. The sequence of handshakes along a chain of stages, or “FIFO,” may be regarded as the movement of “tokens” along the chain. The presence of a token between stages is indicated by a logic high on the state wire connecting them. A token is absent if the state wire is logic low. A token will advance through a stage if its input or “predecessor” state wire is high, indicating presence of a token, and its output or “successor” state wire is low, indicating a space or “vacancy.” This condition produces a pulse on the fire signal that advances the token [1], [2]. Tokens can change their spacing but not their sequence. If the FIFO is closed into a ring, tokens can recirculate continuously for testing purposes [3]. Except for experiments like ours, FIFOs are rarely closed into rings.

Because token movement is controlled by local handshake signals rather than a global clock, intervals between tokens

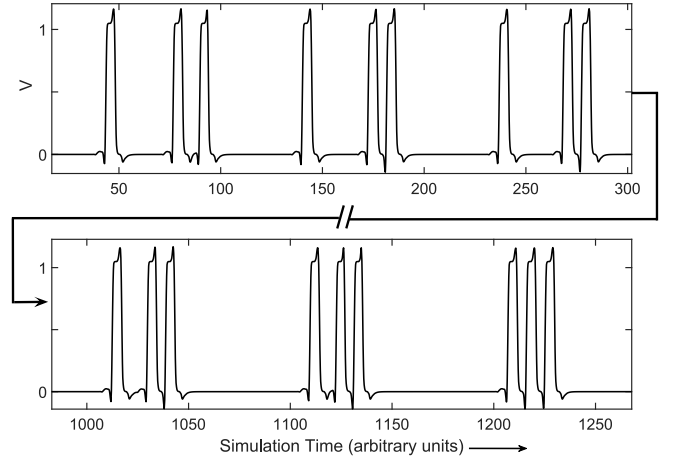


Fig. 1. Example of drafting behavior in a circular GasP-based FIFO. One state wire is observed as tokens pass. Three tokens start at arbitrary intervals and after time 1200 they become and remain tightly clustered.

can vary. When following tokens tend to catch up with leading tokens (Fig. 1) the effect is called “drafting” after the technique bicyclists use to make cross-country cycling easier. The opposite effect, where tokens are pushed apart, is called “negative drafting” or “antidrafting.” For drafting (antidrafting), in each stage, a lead token creates a memory condition where a following token propagates through the stage faster (slower). Total drafting is an accumulation of many small decrements in token intervals by each stage, so the amount of drafting by a following token depends on the number of stages traversed by the token as well as proximity of the preceding token. The drafting phenomenon is well-known and easily observed in circuit simulations such as SPICE and in silicon. Therefore, it is not necessary to invoke device physics beyond effects already embodied in SPICE models to observe the phenomenon and elucidate the circuit mechanism. In spite of this, there is controversy about exactly where the memory condition resides and how it works at the circuit level. The purpose of this paper is to reveal the circuit mechanism in several kinds of self-timed FIFO stages.

For many applications of self-timed circuits, drafting is unimportant because only the sequence of tokens matters and not the intervals between them. But there are new applications, such as spiking neural networks [4] or time-of-arrival measurements [5], for which the intervals carry meaning. The use of self-timed circuits in these applications will require control of drafting. The drafting effect we investigate here is absent from synchronous circuits precisely because a global clock controls the arrival times of data tokens.

Manuscript received October 24, 2018; accepted November 16, 2018. Date of publication January 11, 2019; date of current version March 20, 2019. This work was supported in part by DARPA, “Flexible Specification, Analysis, & Implementation of Self-Timed Circuits,” under Award UTA17-000001 and in part by Portland State University Foundation.

The author is with the Electrical and Computer Engineering Department, Portland State University, Portland, OR 97207 USA (e-mail: clcowan@cecs.pdx.edu).

Digital Object Identifier 10.1109/TVLSI.2018.2884881

Two explanations for token spacing in a self-timed FIFO have been given [6], [7]. The first is the added delay in the two-input decision gate when the inputs are nearly simultaneous [8], also described as the “Charlie effect” in [9]. The Charlie effect is significant only when tokens are nearly minimally spaced which occurs near maximum FIFO throughput. The minimum spacing of tokens in a GasP FIFO is the result of the cycle time of GasP and is only slightly modified by the Charlie effect. Applications that use intervals between tokens for data cannot operate with minimum token spacing; the data would be lost because the intervals would be replaced by fixed circuit-dependent intervals. It is remarkable that drafting occurs even for widely spaced tokens well beyond the range of the Charlie effect.

The other explanation for token spacing is state-wire capacitive effects [10]. Here, drafting is explained as the result of incomplete state-wire charging or discharging which speeds up transitions when tokens are more closely spaced. The combination of state-wire effects and the Charlie effect is frequently used to describe token position in self-timed ring oscillators [11]–[16]. Contrary to this former work, we demonstrate a drafting effect on tokens spaced by over 4000 gate delays. It is very likely that state wires would be fully charged or discharged in this time frame and SPICE measurements fail to show any incomplete transitions. Clearly, there must be a dominant memory effect elsewhere in the circuit.

We find that drafting is the result of the behavior of the internal node between two series transistors in the decision gate. In GasP, the decision gate is a NOR implementing the logical AND. We confirm this by measuring NOR gate behavior for different token spacings. We also demonstrate a test circuit to control the drafting effect which suggests the possibility of using such control for other purposes.

Section II describes the experimental design, Section III describes the results, Section IV offers a discussion of our results, and Section V gives our concluding remarks.

II. STRATEGY AND DESIGN OF EXPERIMENT

A first test circuit is a 17-stage GasP circular FIFO shown in Fig. 2. We selected a simple GasP design [17] because of our familiarity with it. We used a prime number of stages to avoid any synchronization artifacts. Tokens arrive on the predecessor (Pred) state wire and exit on the successor (Succ) state wire. The inputs to the NOR are $\overline{\text{Pred}}$ and Succ signals.

There are two NOR configurations that can implement the AND logic function. If the NOR input that controls the rail-connected pMOS transistor is connected to the predecessor side of the GasP unit we call this the “Rail Pred” configuration. The other NOR input is then connected to the successor side which controls the output-connected pMOS. If the inputs are reversed and the rail-connected pMOS transistor is connected to the successor side we call this the “Rail Succ” configuration.

The event of a token passing through a GasP stage is marked by the fire signal which is derived from the NOR output. Fire is typically used to strobe data through latches, but in our experiments, we omit latches and concentrate only on the handshake events. Many GasP circuits use “get-out-of-the-way” keepers that are disabled when the state wires are changed. But the

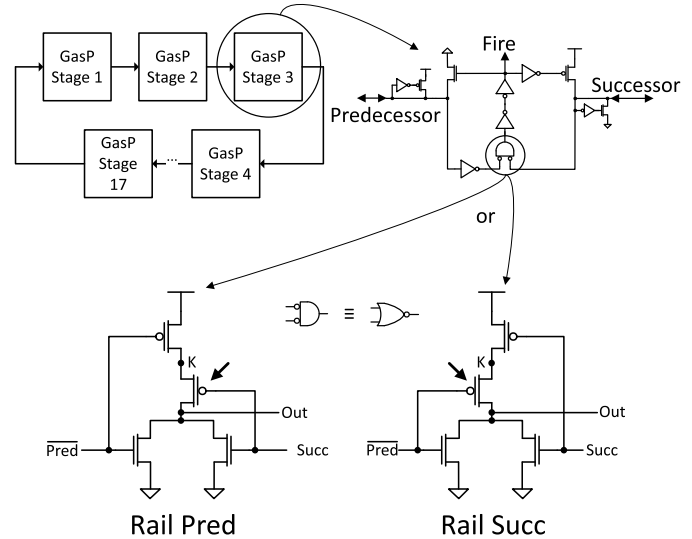


Fig. 2. 17-stage circular GasP FIFO test circuit. Each stage is a simple GasP circuit. Each GasP stage can have a Rail Pred or a Rail Succ NOR configuration. The K-node and the pMOS transistor (arrows) that discharges it are critical to understanding drafting. The AND function of the decision gate is implemented as the De Morgan's equivalent NOR.

experiments reported here use always-on over-driven keepers for simplicity. Our GasP design is, therefore, simple but shows all the drafting behavior of interest.

A second test circuit is a NOR gate isolated in a test bench consisting of FO4 input drivers and an FO4 NOR load. This affords accurate timing measurements because the NOR inputs are fully controlled.

We use a generic 32-nm Synopsys model in HSPICE for simulation at nominal temperature (25 °C) and supply voltage (1.0 V). Logical effort was used to size the gates in GasP so that each gate has the same delay and near-optimal circuit performance. Timing is reported in gate delays (gd) which is the nominal propagation through an FO4 inverter; in this technology 1 gd \simeq 10 ps.

A FIFO containing “sparse tokens” means the tokens are free to move and are not fully drafted. A FIFO containing “sparse vacancies” means the FIFO contains many tokens minimally spaced and there are vacancies between some tokens which are free to move and are not fully drafted. Vacancies can draft like tokens but in a retrograde fashion. Once fully drafted, vacancies or tokens maintain a minimal spacing which is the cycle time of one GasP stage.

A schematic timing diagram, Fig. 3, shows two cases of a pair of tokens passing through one Rail Pred configured GasP stage in an initially quiescent FIFO. In either case, the tokens are taken to be separated more widely than minimum spacing so that they are not yet fully drafted. That is, $TI > 10$ nominal gate delays. The interval between tokens, TI , is measured between rising edges of Pred. Also, the cycle time of 10 nominal gate delays is required for the logical processing to move a token across a GasP stage and reset the stage. The NOR interval (NI) is the interval between inputs to the NOR gate; that is, between the falling edge of Succ due to the leading token, and the falling edge of $\overline{\text{Pred}}$ due to the following token (1). NI is part of TI , and TI is 10 nominal gate delays

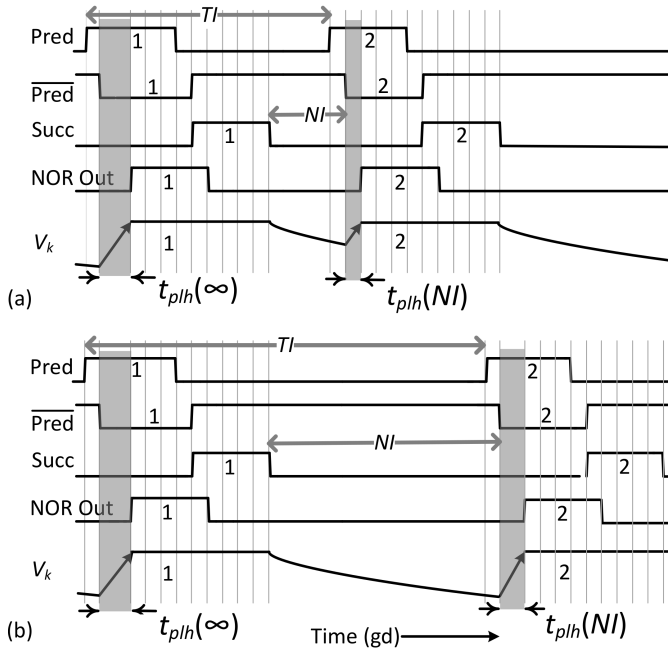


Fig. 3. Example GasP timing diagram for two tokens passing through one GasP stage with a Rail Pred NOR. Token 2 follows token 1, K-node voltage V_K decays between the fall of Succ and the fall of Pred and alters t_{plh} . (a) Short TI and a short NI resulting in a short NOR t_{plh} . Token 2 propagates through the GasP stage with short delay. (b) Longer TI and NI resulting in a longer t_{plh} and token 2 takes longer to propagate.

longer than NI . One of the 10 nominal gate delays is the delay through the NOR gate, t_{plh} , timed from the controlling falling input of the NOR

$$NI = t(\overline{\text{Pred}} \downarrow) - t(\text{Succ} \downarrow). \quad (1)$$

$$t_{plh} = t(\text{NOR out} \uparrow) - \max[t(\overline{\text{Pred}} \downarrow), t(\text{Succ} \downarrow)]. \quad (2)$$

During the NOR interval, the K-node is isolated from active drive so the voltage V_K on that node decays slowly. Comparison of Fig. 3(a) and (b) shows that when NI is shorter, the decay of V_K is less complete so the subsequent NOR t_{plh} transition is shorter because it starts from a potential closer to V_{DD} . That is, for shorter NI , the following token is passed more quickly through the GasP stage than for longer NI . In the two-token example, the leading token transits the GasP with maximum delay corresponding to $t_{plh}(NI = \infty)$, so the following token always transits the GasP more quickly than the leading token. The following token, therefore, catches up with the leading token (reducing TI and NI), with the rate of catch up increasing as the following token closes in on the leading token. When $TI \simeq 10$ nominal gate delays ($NI \simeq 0$), the following token stops catching up and moves at the same speed as the leading token.

We performed four simulation experiments. The first experiment surveyed the drafting behavior of tokens and vacancies using GasPs all configured first with Rail Pred, and then with Rail Succ.

A second experiment used a one-at-a-time replacement of components in all the GasP stages with ideal elements to isolate the cause of the drafting effect. An ideal SPICE switch was packaged with an ideal one-gate delay to create an

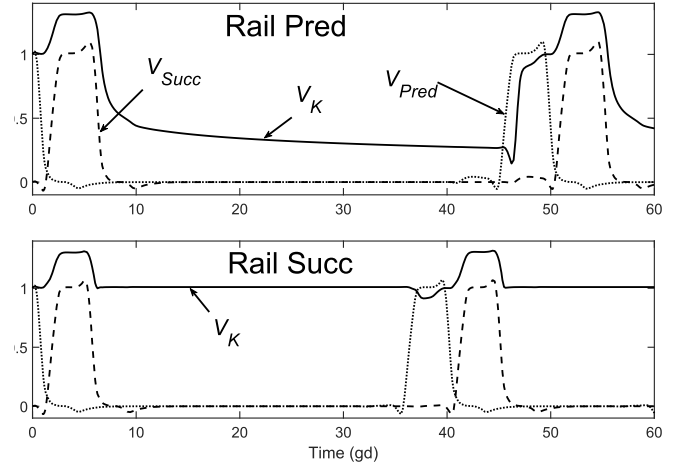


Fig. 4. Difference between the Rail Pred and the Rail Succ NOR configurations. For Rail Pred, V_K decays between tokens, while for Rail Succ it does not. Decay of the K-node residual charge affects NOR t_{plh} , and ultimately delay through the GasP stage.

ideal transistor. Ideal transistors in nMOS and pMOS versions were combined to form ideal inverters and an ideal NOR gate. First, we replaced the state-wire drivers, followed by other gates until all were replaced with ideal components.

In a third experiment, measurements were performed on a NOR gate isolated in a test bench in order to determine the relationship between NOR t_{plh} , K-node voltage (V_K), and NI .

In a fourth experiment, a control circuit was added to every GasP stage of a circular FIFO to prevent drafting or introduce antidrafting using control inputs.

III. RESULTS

A. Experiment 1: Drafting Survey

A GasP FIFO containing sparse tokens was observed to draft by the mechanism shown in Fig. 3 for the NORs of the GasPs in the Rail Pred configuration, but the same FIFO using the Rail Succ NOR configuration would not draft. Fig. 4 shows the behavior of the internal K-node in the Rail Pred configuration (top) and in the Rail Succ configuration (bottom). For Rail Pred, V_K decays between tokens [Figs. 3 and 4 (top)], while for the Rail Succ V_K does not decay between tokens [Fig. 4 (bottom)]. Since V_K does not decay between tokens for the Rail Succ configuration, there is no drafting for that configuration. This is because, unlike the Rail Pred case in Fig. 3, t_{plh} does not depend on NI . These observations implicate the key role of the K-node in the drafting mechanism.

B. Experiment 2: Cause of Drafting (V_K Decay)

The effect of state-wire capacitance on drafting was studied by replacing state-wire drivers with ideal transistors in a simulation of GasP elements in a FIFO showing drafting. Ideal transistors eliminate state-wire capacitance charging delays by instantly fully charging or fully discharging the state wires. As a worst case test, each state wire was loaded with a 1 pF capacitor. No change of drafting behavior was seen when ideal transistors were used, eliminating the state-wire capacitance mechanism as the cause of drafting. The experiments continued by systematically replacing other gates with ideal gates.

No change in drafting behavior was seen until replacement of the NOR with an ideal NOR turned off the drafting effect. Experiments continued further by systematically replacing transistors within the NOR with ideal transistors. Only when either transistor in the pMOS stack was replaced, did drafting cease. When all transistors in the GasP were returned to normal, except keeping either pMOS transistor ideal, drafting remained disabled. This implicates the pMOS stack in the NOR and the behavior of the common K-node as the key factor in the drafting mechanism. An ideal rail-connected pMOS disables drafting by pulling the K-node high quickly, regardless of the value of V_K . More fundamental to the “memory effect” of the drafting mechanism is the way an ideal output-connected pMOS transistor disables the slow decay of V_K prior to when the $\overline{\text{Pred}}$ input goes low.

The K-node capacitance includes the diffusion capacitance at the junction of the drain of the upper rail-connected pMOS and the source capacitance of the lower output-connected pMOS transistor. This node was mentioned by previous studies but was not fully investigated [15], [16]. The NOR gate is the only combinational circuit in GasP that contains a node which is sometimes isolated from active drive and can, therefore, drift over a very long time between NOR actions.

C. Experiment 3: Measure V_K and t_{plh} of the NOR Gate

We measured the timing characteristics of an isolated Rail Pred NOR gate in a SPICE. The interval between NOR inputs was varied. This afforded more accurate measurements than using a running FIFO. Both inputs to the NOR, the NOR output, and V_K were probed. NI was measured as the time difference between the 50% falling edges of Succ and $\overline{\text{Pred}}$ using (1). The NOR propagation delay t_{plh} was measured from the 50% falling edge of $\overline{\text{Pred}}$ to the 50% rising edge of NOR output, which is the typical condition for a FIFO with sparse tokens, using (2). The measurements of V_K , NI , and t_{plh} are shown in Fig. 5. Fig. 5(a) shows that voltage V_K decreases as NI increases. Fig. 5(b) shows that NOR t_{plh} increases as V_K decreases. Elimination of V_K between the top two graphs gives the bottom one which shows t_{plh} increasing with NI , Fig. 5(c). This quantifies the effect shown schematically in Fig. 3.

NOR t_{plh} is approximately linearly related to V_K [Fig. 5(b)] because pMOS transistors approximate constant current sources when charging node K and the NOR output node. Therefore, the time to charge up to the threshold for generating a logic high at the NOR output will also be linear with time. With short intervals ($NI < 1$ gd), node K remains at or above V_{DD} and t_{plh} is nearly constant so there is only slight antidrafting due to the Charlie effect. In this technology, V_K is initially boosted above V_{DD} during NOR output high due to prominent Miller feedthrough (Fig. 4). The drafting effect extends to 4000 gd, the escape interval (EI). For $NI > EI$, a trailing token is unaffected by the preceding token and there is no drafting.

D. Experiment 4: A Control Circuit

Because the decay of V_K between tokens causes token drafting, one expects that a rising V_K between tokens should generate antidrafting. Also, because a Rail Succ configuration

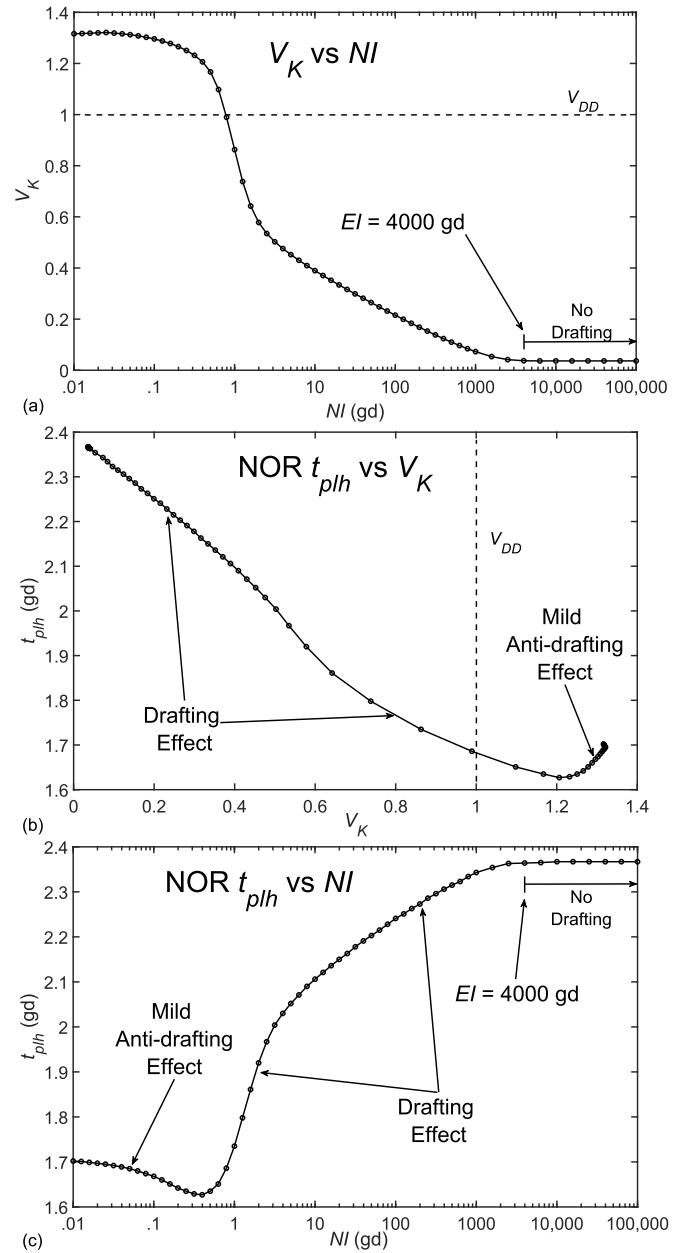


Fig. 5. Measurements of NOR in a test bench. (a) Voltage V_K decreases with NI . (b) t_{plh} decreases approximately linearly with V_K . (c) Elimination of V_K in (a) and (b) results in t_{plh} increasing with NI . Note the slight antidrafting effect at very short NI (less than one-gate delay).

clamps V_K to V_{DD} preventing drafting, holding V_K constant between tokens in a Rail Pred configuration is expected to prevent drafting or antidrafting regardless of the token separation. Each time the GasP-stage fires and moves a token to the output state wire, node K is reset. Fig. 6(a) schematically shows a circuit to control the voltage on the K-node in a GasP stage. Circuit details are in the Appendix. For drafting, the control circuit does not affect the normal reset to V_{DD} . For antidrafting the control circuit forces a reset to 0 V, overriding the normal NOR reset. To prevent drafting or antidrafting, node K is reset to $V_{DD}/2$ and clamped there.

Each GasP stage in our test FIFO has a copy of the control circuit that connects directly to the Rail Pred NOR K-node in that stage and can alter its voltage over time, overriding the

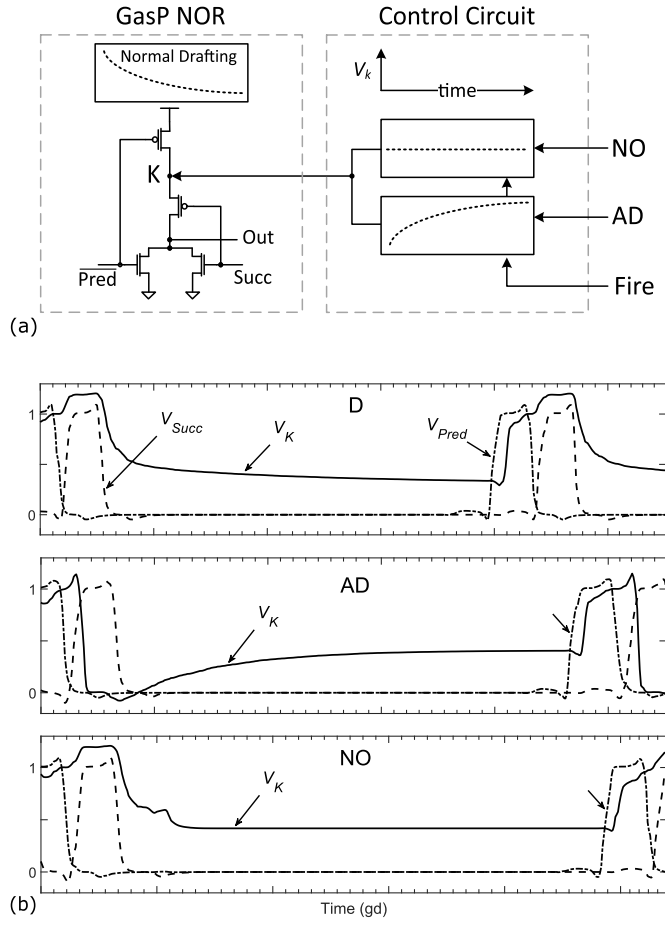


Fig. 6. (a) Control circuit, detailed in the Appendix and Fig. 17, selects a desired V_K versus time characteristic. The circuit overrides the natural drafting effect (D) of a Rail Pred NOR gate by asserting either NO, which clamps K to a fixed voltage, or by asserting AD which generates a rising V_K function. (b) Waveforms show how V_K varies between the fall of V_{Succ} and the rise of V_{Pred} for each of the control circuit configurations (D, NO, AD).

normal decay of V_K . The control circuit uses that stage's fire signal to modify V_K at the correct time in order to preserve the logical function of the NOR. Node K is controlled only when it is not actively driven which is when the GasP stage is empty and waiting for a new token to arrive. This condition occurs between the falling edge of Succ and falling edge of Pred. The control inputs to each copy of the control circuit are NO (no drafting or antidrafting) and AD (antidrafting). These inputs are mutually exclusive and are asserted with a logic high. Each stage could be designed for independent control but for our design, all stages in the FIFO are controlled in parallel by common NO and AD inputs.

The desired waveforms during control of node K are shown schematically in Fig. 6(a). Fig. 6(b) shows the measured waveforms at one stage in the FIFO during simulation. During normal drafting (D), the control circuit is inactive and there is the usual decaying function of V_K associated with drafting. During AD, the V_K waveform rises which gives the inverse action on the NOR t_{plh} and antidrafting behavior. During NO, V_K remains constant between firings and there is neither drafting nor antidrafting. These measurements, Fig. 6(b), show that the control circuit creates the desired waveforms for V_K .

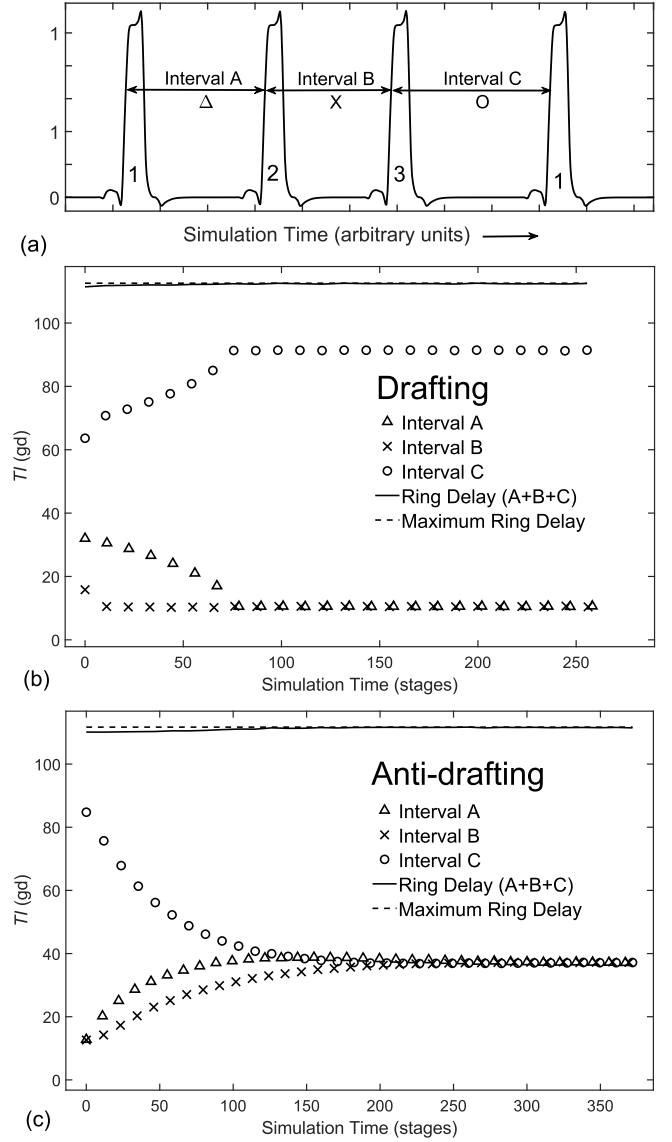


Fig. 7. Evolution of token intervals (TI) measured on a single Pred state wire in a Rail Pred circular FIFO with three tokens. (a) Monitoring V_{Pred} detects tokens 1, 2, and 3 with token intervals A, B, and C as they pass. (b) Evolution of drafting. (c) Evolution of antidrafting. Full drafting results in two small token intervals and one large token interval. Full antidrafting evolves to equal token intervals. Cycle time through the FIFO (ring delay) changes very slightly as drafting or antidrafting evolves.

The control circuit was tested on a circular 17-GasP-stage FIFO with three tokens circulating. Token intervals, defined as the time between rising edges of Pred [Fig. 7(a)], are shown for the drafting [Fig. 7(b)] and the antidrafting [Fig. 7(c)] settings of the control circuit. For drafting, the initial configuration of tokens had three different intervals which evolved into two minimal intervals ($TI = 10$ nominal gate delays) and one long interval so that the three tokens are grouped together and move as one unit, Fig. 7(b). For the antidrafting case, three fully drafted tokens evolved into three equally spaced tokens, shown in Fig. 7(c). The total ring delay, which is the sum of intervals, is seen to be insensitive to token spacing [top curve, Fig. 7(b) and (c)], and is nearly equal to the maximum ring delay corresponding to the maximum t_{plh} for the leading

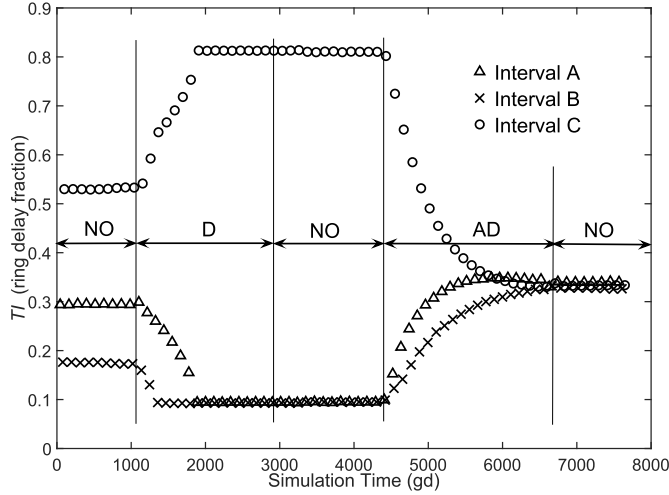


Fig. 8. SPICE simulation using the K control circuit. During NO, tokens remain stable. During D, tokens become fully drafted after 1900 gate delays. During AD, tokens move to equal spacing that is fully antidrafted.

token. Note that drafting and antidrafting is a cumulative effect of small increments requiring many cycles through the FIFO and passage through many stages.

Fig. 8 shows the effect of the control circuit on FIFO operation. Three tokens are placed in a 17-stage Rail Pred circular FIFO with control circuits in each stage. The arbitrary starting positions (stages 1, 4, and 9) gives A, B, and C starting token intervals of (5/17), (3/17), and (9/17) of a total ring delay, respectively. Token intervals defined as in Fig. 7(a) are plotted as fractions of the total ring delay to eliminate the changes in ring delay due to the control circuit. As the tokens circulate, control circuits on all stages were exercised simultaneously with a sequence of commands NO, D, NO, AD, and NO asserted as indicated by arrows in Fig. 8. During NO, no drafting or antidrafting occurs, preserving the token intervals. During D, normal drafting occurs and the tokens become fully drafted after 1900 gate delays. During AD, the tokens become fully antidrafted (equally spaced) after another 1900 gate delays. This demonstrates full control of drafting and antidrafting behavior by manipulation of the V_K profile using the control circuit.

IV. DISCUSSION

It is useful to discuss t_{plh} versus NI characteristic in more detail. Fig. 9 shows t_{plh} as a function of NI for a Rail Pred NOR gate within a single GasP stage. Measurements of NOR input and output edge times were made as token intervals presented to the stage were varied. Substitution of resulting input and output NOR edge times into (1) and (2) gives t_{plh} versus NI characteristic in Fig. 9. For negative NI , t_{plh} is very different from positive NI . In a sparse vacancy FIFO $\overline{\text{Pred}}$ falls before Succ, so NI is negative, t_{plh} is independent of NI , and there is no drafting [Fig. 9 (light-shaded area)]. On the other hand, in a sparse token FIFO, $\overline{\text{Pred}}$ falls after Succ like Fig. 3, so NI is positive, t_{plh} increases strongly with NI , and drafting occurs. The shape of the t_{plh} curve for positive NI induces drafting because shortening the interval between tokens reduces the delay and shortens the interval further.

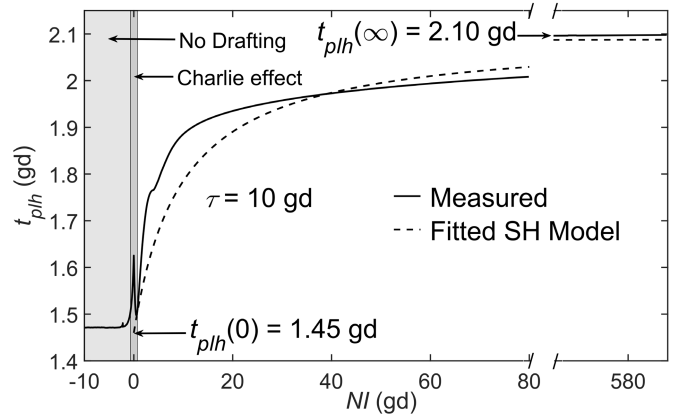


Fig. 9. NOR propagation delay t_{plh} versus NOR input arrival time difference NI for the Rail Pred configuration. t_{plh} is a strong function of NI for $NI \gtrsim 1$ gd but is nearly independent of NI for $NI \lesssim 1$ gd. For $NI \gtrsim 1$ gd, the positive slope of $t_{plh}(NI)$ causes drafting, for which a simple “SH” model [(12), dashed line] can be derived. A more complex behavior, the Charlie effect, occurs for NOR inputs within less than one-gate delay of simultaneity, $|NI| \lesssim 1$ gd. A token will stop drafting when the interval to a preceding token falls to $TI \simeq 11$ gd ($NI \simeq 1$ gd) and the antidrafting Charlie effect stops further drafting.

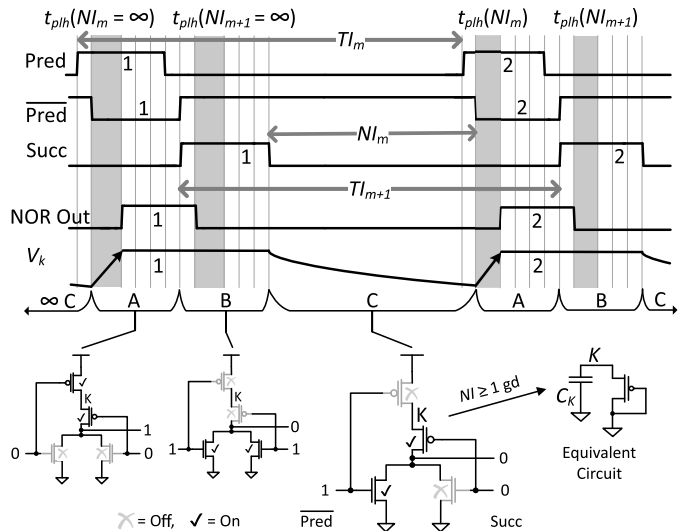


Fig. 10. Suppose two tokens, 1 then 2, traverse the m th GasP in a chain of GasPs that has been idle for a long time. When a token traverses the m th GasP the NOR gate goes from state C to A, then to B before returning to C. State A charges the K-node to V_{DD} , which is preserved by state B. If token 2 follows token 1 such that $NI_m \gtrsim 1$ gd, then in the C-state V_K decays slowly from V_{DD} to V_T through a diode-connected pMOS transistor. C-state decay of V_K is interrupted by token 2 before the K-node is fully discharged, so $t_{plh}(NI_m) < t_{plh}(\infty)$, causing the drafting effect, $TI_{m+1} < TI_m$. TI_{m+1} is identified because $\text{Succ}_m = \text{Pred}_{m+1}$.

Tokens separated by many hundreds of gate delays will draft by this mechanism. When inputs are nearly simultaneous, a different behavior occurs [Fig. 9 (dark-shaded area)]. This behavior, the Charlie effect, causes antidrafting for positive NI and drafting for negative NI when $|NI| \lesssim 1$ gd. This effect only influences tokens propagating at nearly minimum spacing ($TI \simeq 10$ gd) by pulling or pushing them slightly.

Fig. 10 shows the circuit mechanism of V_K decay during drafting (Fig. 3) in more detail. Between tokens (state C), the GasP stage is empty and waiting for a token to arrive at Pred; Succ is logic 0 and $\overline{\text{Pred}}$ is logic 1. Node K is charged to

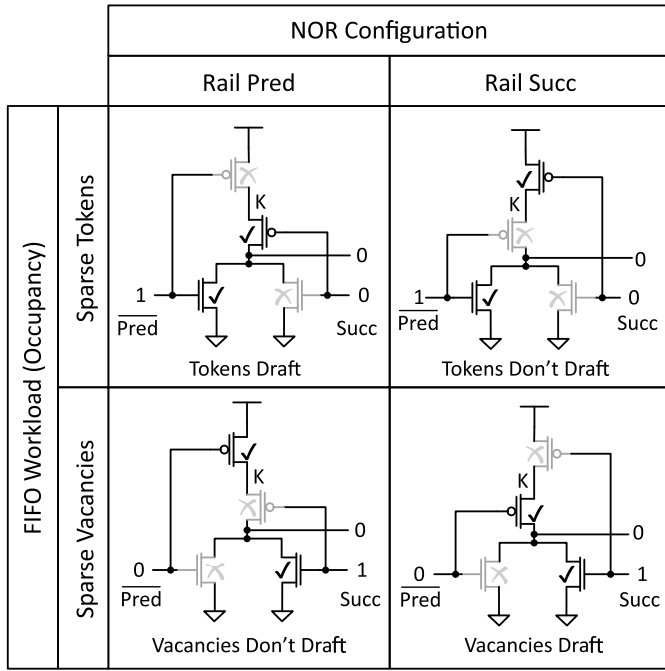


Fig. 11. GasP NOR design configuration determines drafting behavior for sparse token or sparse vacancy workloads. The C-state (time between tokens) for each case is shown here. Drafting occurs when the output-connected pMOS is turned on before the rail pMOS.

V_{DD} and the NOR output is at logic 0. This condition results in a diode-connected pMOS transistor that discharges node K through the Pred side nMOS transistor which is turned on. In order to generate a logic high output, both the K-node and the NOR output node must be charged.

The preceding discussion was for sparsely distributed tokens propagating on a chain of GasPs which have NOR gates in the Rail Pred configuration. The Rail Succ NOR configuration (Fig. 2) and other workload cases shown in Fig. 11 can be analyzed similarly. Drafting of sparse tokens/vacancies occurs when the C-state has the output-connected pMOS on, slowly discharging K. Nondrafting of sparse tokens/vacancies occurs when the C-state has the rail-connected pMOS is on, clamping K to V_{DD} .

A. Simplified Drafting Model

Fig. 10 shows the sequence of states of the NOR gate and timing of signals as two tokens traverse a Rail Pred GasP stage m in a chain of GasPs that has been idle for a long time. The most usual case of widely spaced tokens has token 2 following token 1 by an interval, $TI_m > 10$ logical gate delays, such that $NI_m > 0$. If logical gate delays are assumed to be unity, except that $t_{\text{plh}}(NI)$ depends on the duration NI of the preceding C-state, inspection of Fig. 10 gives

$$TI_m = 1 + t_{\text{plh}}(NI_m = \infty) + 5 \\ + t_{\text{plh}}(NI_{m+1} = \infty) + 3 + NI_m - 1 \quad (3)$$

$$TI_{m+1} = 1 + t_{\text{plh}}(NI_{m+1} = \infty) + 3 \\ + NI_m + t_{\text{plh}}(NI_m) + 4 \quad (4)$$

where (3) and (4) use the fact that Succ of GasP stage m is Pred of GasP stage $m + 1$ so the second gate delay after the reset (rising edge) of Succ of GasP stage m is the NOR delay of GasP stage $m + 1$, $t_{\text{plh}}(NI_{m+1})$. Simplifying, (3) and (4) become

$$TI_m = 8 + NI_m + 2t_{\text{plh}}(\infty) \quad (5)$$

$$TI_{m+1} = 8 + NI_m + t_{\text{plh}}(\infty) + t_{\text{plh}}(NI_m). \quad (6)$$

Replacing m with $m + 1$ in (5) and subtracting from (6) gives

$$NI_{m+1} = NI_m + t_{\text{plh}}(NI_m) - t_{\text{plh}}(\infty). \quad (7)$$

If t_{plh} is known as a function of NI , then (7) may be iterated to find NI_m for $m \geq 1$. TI_m for $m \geq 1$ may be computed via (5).

A simple model of $t_{\text{plh}}(NI)$ may be derived by noticing that in the usual case of widely spaced tokens ($NI > 1$ gd) t_{plh} is controlled by the mechanism of charge on the K-node leaking away through a diode-connected pMOS transistor during the preceding C-state, as shown in Fig. 10. The complex behavior of simultaneous NOR input transitions (Charlie effect) may be modeled as simply a “stop” to drafting when $NI = 0$. For $NI > 1$ gd in the C-state, the condition that the current discharging the K-node capacitance, C_K , flows through the diode-connected transistor is written

$$C_K \frac{dV_K}{dt} = -\kappa(V_K - V_T)^2 \quad (8)$$

where the Shichman–Hodges (SH) model of transistor action, ignoring body effects, is assumed. V_T is the threshold voltage. Equation (8) can be integrated with the boundary condition that $V_K = V_{DD}$ at $t = 0$ to give

$$V_K = \frac{V_{DD} - V_T}{1 + t/\tau} + V_T \quad (9)$$

where

$$\tau = \frac{C_K}{\kappa(V_{DD} - V_T)}. \quad (10)$$

When the C-state duration is indefinitely long, $t \rightarrow \infty$, $V_K \rightarrow V_T$, and t_{plh} reaches a maximum of $t_{\text{plh}}(\infty)$. For the example shown in Fig. 9, $t_{\text{plh}}(\infty) = 2.10$ gd. On the other hand, at $t = 0$, $V_K = V_{DD}$ and the t_{plh} function, ignoring the Charlie effect, will extrapolate to $t_{\text{plh}}(0) = 1.45$ gd. The simplest model of t_{plh} , linear in V_K per Fig. 5(b), with these limits is

$$t_{\text{plh}}(V_K) = (t_{\text{plh}}(0) - t_{\text{plh}}(\infty)) \frac{V_K - V_T}{V_{DD} - V_T} + t_{\text{plh}}(\infty) \quad (11)$$

or, using (9)

$$\begin{aligned} t_{\text{plh}}(t) &= -\Delta t_{\text{plh}} \frac{\tau}{\tau + t} + t_{\text{plh}}(\infty) \\ \Delta t_{\text{plh}} &\triangleq t_{\text{plh}}(\infty) - t_{\text{plh}}(0). \end{aligned} \quad (12)$$

In (12), t is identified as NI in Fig. 9, and a fit to the data gives $\tau = 10$ gd. Substitution of (12) into (7) gives

$$NI_{m+1} = NI_m - \Delta t_{\text{plh}} \frac{\tau}{\tau + NI_m} \quad (13)$$

which may be iterated to find NI_m , $m > 1$.

The continuum limit of (13) is easily integrated to give a closed-form formula. The approximation is good because the

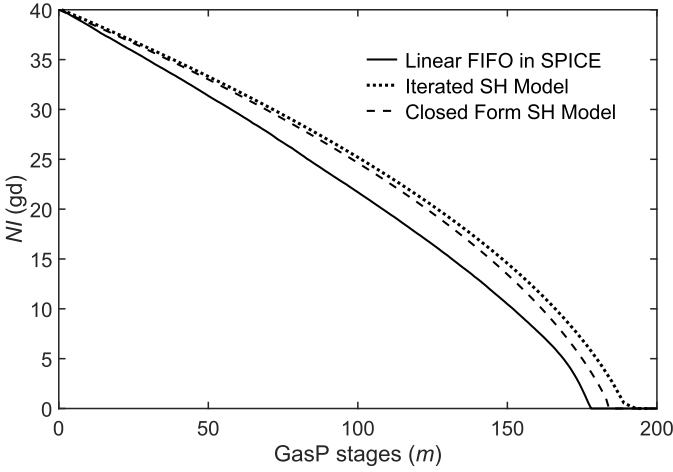


Fig. 12. Iterated SH model and closed-form SH model are compared to a linear FIFO simulated in SPICE. The starting NI is 40 gd for each.

change in NI per GasP is usually small ($NI_m - NI_{m+1} \ll \Delta t_{plh} = 2.1 - 1.45 = 0.65$ gd). The continuum limit is written

$$NI_{m+1} - NI_m \rightarrow \frac{dNI}{dm} = -\Delta t_{plh} \frac{\tau}{\tau + NI} \quad (14)$$

which integrates to a closed form

$$NI_m = ((NI_0 + \tau)^2 - 2\tau \Delta t_{plh} m)^{1/2} - \tau. \quad (15)$$

Fig. 12 shows the NOR input interval NI_m resulting from two tokens traversing GasP stage m in a Rail Pred GasP FIFO, for $m \geq 0$. Token 2 is injected onto GasP element $m = 0$ following token 1 by $TI = 50$ gd so that $NI_0 = 40$ gd. The FIFO uses the same GasP design and technology as the example shown in Fig. 9. SPICE simulation (Fig. 12) shows that the second token catches up with the first and becomes fully drafted at GasP stage $m = 178$. Also, shown in Fig. 12 are the results of using Δt_{plh} and τ from Fig. 9 to iterate the SH model (12) and to evaluate the closed-form approximation (15). The good agreement between the SPICE simulation, the SH model, and closed form shows that the latter are useful approximations. The number of stages, m , required for a pair of tokens separated initially by NI_0 to totally draft is the solution to (15) with $NI_m = 0$. That is

$$m = ((NI_0 + \tau)^2 - \tau^2) / (2\tau \Delta t_{plh}) \quad (16)$$

which, for this example, gives $m = 185$ stages.

B. Behavior of the Simplified Circuit

The K-node behavior of the NOR gate, and two simplifications of the NOR gate, the pMOS Stack and a further simplification (“Simple Circuit”), were simulated in SPICE. Results of V_K versus NI for these circuits are compared in Fig. 13, and good agreement among all is observed. Additionally, the simple SH model (9) is plotted using parameters given in Fig. 9. The deviation of the SH model from the others shows that the body effect on threshold voltage (V_T), ignored in the SH model, is quite significant. I_D does not follow the SH model when the V_K falls below $V_{DD}/2$ because

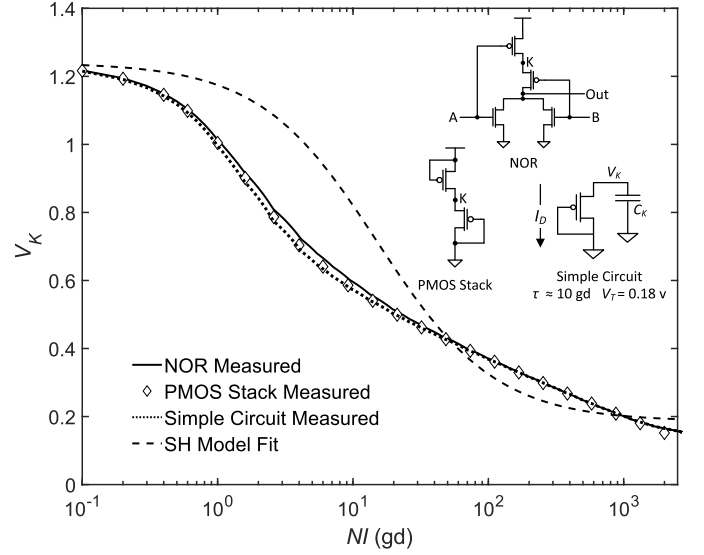


Fig. 13. Fitted SH model for V_K (9) is compared to HSPICE simulations of the NOR and two simplified circuits. Note that the SH model curve departs from the measured data. This is due to the omission of the body effect which raises V_T and causes the pMOS transistor to go into subthreshold conduction.

the diode-connected pMOS transistor moves into subthreshold conduction. The parasitic capacitance C_K also varies (decreases) as V_K drops because of increasing reverse bias across the n-well/source-drain diffusions, but the agreement of the “Simple Circuit” with the NOR and pMOS stack shows that the effect is small. Although the SH model omits the significant effect of subthreshold conduction and body bias effects, it is mathematically simple [see (13), (15), and (16)] and gives qualitatively useful results (Fig. 12).

C. Other Technologies and Other Gate Designs

Fig. 14 (top) shows a comparison of NOR gates and FIFOs in several technology nodes. Fig. 14 (bottom) compares delays in four additional decision gate designs used in self-timed FIFOs (Fig. 15) for the 32-nm technology node. Gates in all examples were driven and loaded by FO4. The gate inputs are A and B. The NOR results, Fig. 14 (top), show that the behavior in all technologies is qualitatively the same, but the variation is greater for the smaller technology nodes. Comparison of all gates across the 32-nm node shows that some degree of drafting can occur in FIFOs based on any of the gate designs. FIFOs were constructed from each of the four additional gate designs, and token propagation was simulated so that drafting effects could be observed. The FIFO designs were Micropipeline, Click, GasP NOR, and Mousetrap.

The Muller C-Element CEL with its associated latch causes drafting in Micropipeline [18]. Fig. 15 shows that there is a discharge path from node K1 to ground through the B-connected pMOS transistor and through an nMOS transistor (*) in the latch.

In Click [19], which we implement from [20], the key decision gate is the NAND. A Click FIFO shows early pairing of tokens followed by drafting of intervals between pairs. The pairing is due to the different propagation delays of a bit 0

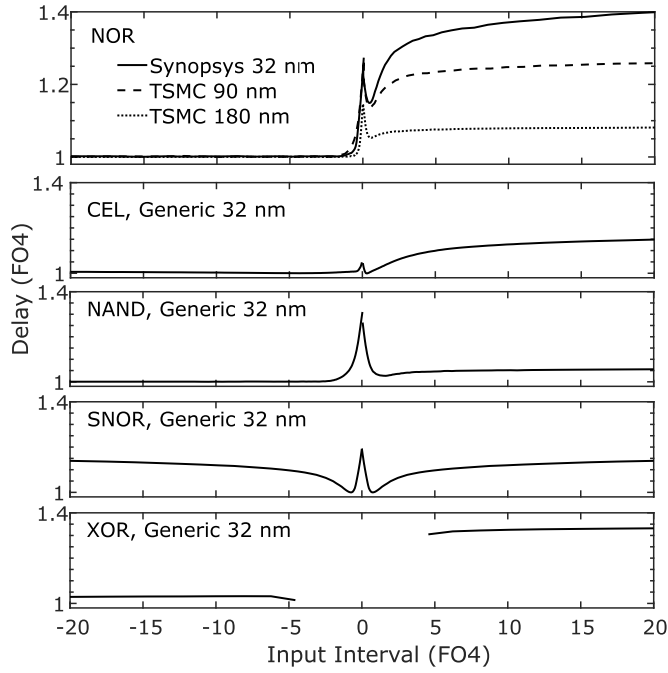


Fig. 14. Three different technologies are shown in the case of the NOR. All three show the same basic shape (top). Propagation delay characteristics for other gate designs for 32-nm technology are shown for comparison (bottom). All of these gates show propagation delay versus interval between inputs A and B that will cause drafting. The intervals are shown in units of FO4 delays for that gate.

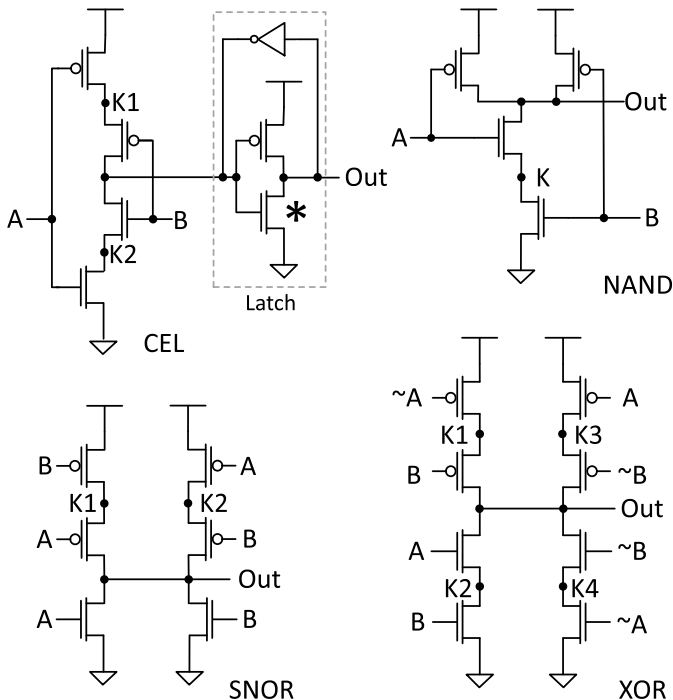


Fig. 15. Four gate designs, with inputs A and B, that have internal K nodes that can cause drafting. These nodes are K1 in CEL (Micropipeline), K in the NAND (Click), both K1 and K2 in SNOR depending on which input is asserted first, and K2 and K4 in XOR (Mousetrap). In CEL, the latch nMOS (*) completes the discharge path from K1.

versus a bit 1 through the one-bit latch in the design. This overwhelms the t_p variations of the XOR and creates longer t_p for odd intervals and shorter t_p for the even intervals.

Eventually, the pairs draft together from the effect of the NAND K-node.

The symmetric NOR (SNOR) does not have a Rail Pred or Rail Succ difference and the two K nodes cause drafting of both tokens and vacancies when used in a GasP FIFO.

In Mousetrap [21], as realized in [1], the XOR variability is slight but present. In our design, the XOR only operates with positive input intervals greater than four FO4 gate delays. Like Click, a Mousetrap FIFO also shows early pairing of tokens because it also uses a one-bit latch. Superimposed on this is a gradual drafting of the paired intervals due to an XOR K-node.

A general principle can be derived from the decision gates examined here. Since all CMOS decision gates have at least one pMOS or nMOS stack in the design there will be one or more internal K nodes. When the timing of the two inputs to the gate arrive such that the inner (output-connected) transistor in the stack conducts before the outer (rail-connected) transistor the intervening K-node will drift and there will be drafting. If the output node does not float between tokens there will be no drafting.

V. CONCLUSION

Drafting, in the case of a GasP FIFO, is due to the behavior of the internal node, called “K” in the NOR decision gate. In a Rail Pred NOR FIFO with sparse tokens, this node discharges slowly between NOR actions because it is the only node in GasP that can float or decay with time; all other nodes are constantly driven. When a new token arrives the t_{plh} of the NOR, and therefore the propagation delay through the GasP stage, differs from the preceding token’s t_{plh} depending on the amount of residual charge at K. Short token intervals preserve more charge at K and shorten t_{plh} . Longer token intervals lengthen t_{plh} . This results in shortening of token intervals over time which is seen as drafting. Similarly, in a Rail Succ NOR FIFO with sparse vacancies, this node also discharges between vacancies resulting in the retrograde drafting of vacancies. By controlling the charge on K between tokens in a Rail Pred configuration tokens can be made to draft, antidraft or, if the voltage on node K is kept fixed, neither.

The two earlier explanations for drafting fail to explain our observations. We have measured, in simulations, drafting effects persisting for over hundreds of GasP stages or thousands of gate delays even with ideal state-wire drivers. We also observe that the time constant for the decay of V_K matches the time scale of drafting. Measurements of state-wire voltage in SPICE simulations indicate that state wires are fully charged or discharged within the five gate delays that GasP state-wire drivers are driven. Therefore, state-wire capacitance effects cannot explain the long-range drafting effect.

One obvious way to prevent drafting in a FIFO operating with sparse tokens is to use the Rail Succ NOR configuration. A more general way is to replace the NOR with a ratioed NOR [22] that avoids internal K nodes, Fig. 16(a). This prevents drafting of either tokens or vacancies for any FIFO occupancy but with a penalty of extra power consumption. This is similar to the approach taken by Fairbanks in [16]. Another technique is to add an additional, small precharge nMOS transistor [23] which discharges the K-node between

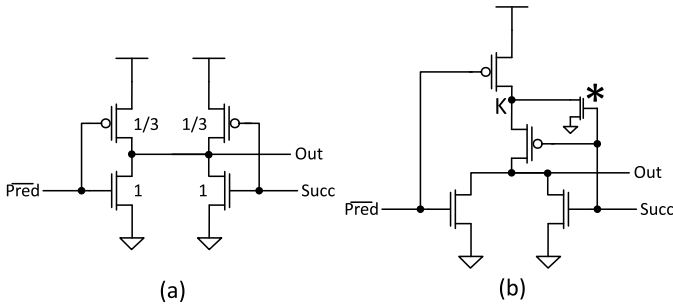


Fig. 16. Two NOR examples to prevent drafting. (a) Ratiod NOR gate which has no internal K nodes. The drive strength of the pMOS transistors is 1/3 that of the nMOS. This gate prevents drafting of either tokens or vacancies for any FIFO occupancy. (b) Adding a predischARGE transistor (*) on the rail input fully discharges the K-node between events and prevents drafting.

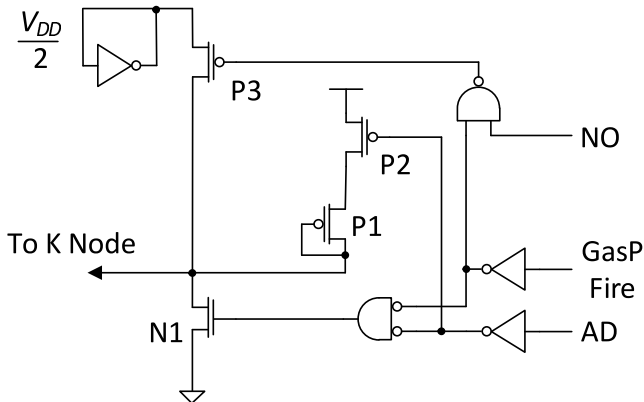


Fig. 17. Detail of the circuit that controls node K. For drafting, node K is unaffected. For antidrafting, node K is discharged by N1 and then charged by P1 and P2. For no drafting or antidrafting, node K is clamped by P3 to $V_{DD}/2$, here generated by the shorted inverter which stabilizes at this voltage.

events, Fig. 16(b). This provides a fixed V_K of 0 V which prevents drafting and consumes less power than the ratiod NOR.

Apart from just preventing drafting a full understanding of the mechanism suggests several new possible applications where intervals between tokens carry the information. Examples include spiking neural networks and computation based on interval encoded data rather than bundled data. Another application may be data security by using controlled amounts of drafting and antidrafting for obfuscation of interval encoded data. We suspect that interval encoding of data would result in significant energy savings over standard bit-encoded data.

Further work will characterize environmental and process variation of the drafting effect because aside from mitigation one can take advantage of manufacturing variability to create a physically unclonable devices [24]. The exact drafting behavior for any specific IC instance could provide unique identification or unique interval-based computation that could be used in cryptography.

APPENDIX

In order to directly affect drafting in a GasP FIFO, we chose to control the internal K-node in each GasP stage. The circuit we used, Fig. 17, is present in each GasP stage of the FIFO

and reaches into that stage's NOR gate to control the K-node. The control inputs AD and NO are mutually exclusive and are asserted logic high. Fire is the GasP Fire signal for that stage.

For drafting (D) neither control input is asserted, the control circuit is inactive and node K discharges normally through the NOR gate; this configuration results in drafting.

For antidrafting, AD is logic high and the Fire signal turns on N1 which discharges node K and turns off the NOR. Because the Fire input propagates through three gates in the control circuit before discharging K, the discharge will happen about the same time as when the Succ line is driven logic high which also turns off the native NOR. Thus, normal NOR logic function will be retained. Between Fire signals node K will then charge through P2 and the diode-connected P1. This yields a long charging time constant similar to the long discharging time constant with drafting.

For no drafting or antidrafting, NO is asserted and node K is clamped to $V_{DD}/2$ by P3 between firings. This voltage was selected because is a good average of the drafting and antidrafting V_K values. Any fixed voltage from 0 V to V_{DD} will work. Depending on which control input is activated, the overall speed of the FIFO changes slightly because the average value of V_K and, therefore, the average NOR t_{plh} varies.

ACKNOWLEDGMENT

The author would like to thank I. Sutherland for motivating this paper and C. G. Shirley for his guidance and assistance. He would also like to thank W. R. Daasch whose extensive contribution to this investigation made this paper possible.

REFERENCES

- [1] M. Roncken, S. M. Gilla, H. Park, N. Jamadagni, C. Cowan, and I. Sutherland, "Naturalized communication and testing," in *Proc. 21st IEEE Int. Symp. Asynchronous Circuits Syst. (ASYNC)*, May 2015, pp. 77–84.
- [2] S. M. Gilla, M. Roncken, and I. Sutherland, "Long-range GasP with charge relaxation," in *Proc. IEEE Symp. Asynchronous Circuits Syst. (ASYNC)*, May 2010, pp. 185–195.
- [3] K. E. Molnar, I. W. Jones, W. S. Coates, and J. K. Lexau, "A FIFO ring performance experiment," in *Proc. IEEE 3rd Int. Symp. Adv. Res. Asynchronous Circuits Syst.*, Apr. 1997, pp. 279–289.
- [4] S. Ghosh-Dastidar and H. Adeli, "Spiking neural networks," *Int. J. Neural Syst.*, vol. 19, no. 4, pp. 295–308, Aug. 2009.
- [5] J. Kalisz, "Review of methods for time interval measurements with picosecond resolution," *Metrologia*, vol. 41, no. 1, p. 17, 2004.
- [6] A. Winstanley and M. Greenstreet, "Temporal properties of self-timed rings," in *Advanced Research Working Conference on Correct Hardware Design and Verification Methods*. Berlin, Germany: Springer-Verlag, 2001, pp. 140–154.
- [7] A. J. Winstanley, A. Garivier, and M. R. Greenstreet, "An event spacing experiment," in *Proc. 8th IEEE Int. Symp. Asynchronous Circuits Syst. (ASYNC)*, Apr. 2002, pp. 47–56.
- [8] V. Chandramouli and K. A. Sakallah, "Modeling the effects of temporal proximity of input transitions on gate propagation delay and transition time," in *Proc. 33rd Design Autom. Conf.*, 1996, pp. 617–622.
- [9] J. C. Ebergen, S. Fairbanks, and I. E. Sutherland, "Predicting performance of micropipelines using Charlie diagrams," in *Proc. 4th Int. Symp. Adv. Res. Asynchronous Circuits Syst.*, 1998, pp. 238–246.
- [10] J. Hamon, L. Fesquet, B. Miscopein, and M. Renaudin, "Constrained asynchronous ring structures for robust digital oscillators," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 17, no. 7, pp. 907–919, Jul. 2009.
- [11] A. Cherkaoui, V. Fischer, A. Aubert, and L. Fesquet, "A self-timed ring based true random number generator," in *Proc. 19th IEEE Int. Symp. Asynchronous Circuits Syst. (ASYNC)*, 2013, pp. 99–106.

- [12] J. Hamon, L. Fesquet, B. Miscopein, and M. Renaudin, "High-level time-accurate model for the design of self-timed ring oscillators," in *Proc. 14th IEEE Int. Symp. Asynchronous Circuits Syst. (ASYNC)*, Apr. 2008, pp. 29–38.
- [13] E. Yahya, O. Elissati, H. Zakaria, L. Fesquet, and M. Renaudin, "Programmable/stoppable oscillator based on self-timed rings," in *Proc. 15th IEEE Int. Symp. Asynchronous Circuits Syst. (ASYNC)*, May 2009, pp. 3–12.
- [14] V. Zebilis and C. P. Sotiriou, "Controlling event spacing in self-timed rings," in *Proc. 11th IEEE Int. Symp. Asynchronous Circuits Syst. (ASYNC)*, Mar. 2005, pp. 109–115.
- [15] S. Fairbanks and S. Moore, "Analog micropipeline rings for high precision timing," in *Proc. 10th IEEE Int. Symp. Asynchronous Circuits Syst.*, Apr. 2004, pp. 41–50.
- [16] S. M. Fairbanks, "High precision timing using self-timed circuits," Ph.D. dissertation, Comput. Lab., Univ. Cambridge, Cambridge, U.K., 2005.
- [17] I. Sutherland and S. Fairbanks, "GasP: A minimal FIFO control," in *Proc. 7th IEEE Int. Symp. Asynchronous Circuits Syst. (ASYNC)*, Mar. 2001, pp. 46–53.
- [18] I. E. Sutherland, "Micropipelines," *Commun. ACM*, vol. 32, no. 6, pp. 720–738, 1989.
- [19] A. Peeters, F. T. Beest, M. de Wit, and W. Mallon, "Click elements: An implementation style for data-driven compilation," in *Proc. IEEE Symp. Asynchronous Circuits Syst. (ASYNC)*, May 2010, pp. 3–14.
- [20] Y. Liu, H. Chen, D. Wang, and A. He, "An asynchronous loop structure based on the click element," in *Proc. Int. Conf. Electron Devices Solid-State Circuits (EDSSC)*, Oct. 2017, pp. 1–2.
- [21] M. Singh and S. M. Nowick, "Mousetrap: Ultra-high-speed transition-signaling asynchronous pipelines," in *Proc. ICCD*, Sep. 2001, pp. 1–9.
- [22] M. G. Johnson, "A symmetric CMOS NOR gate for high-speed applications," *IEEE J. Solid-State Circuits*, vol. SSC-23, no. 5, pp. 1233–1236, Oct. 1988.
- [23] V. G. Oklobdzija and R. K. Montoye, "Design-performance trade-offs in CMOS-domino logic," *IEEE J. Solid-State Circuits*, vol. SSC-21, no. 2, pp. 304–306, Apr. 1986.
- [24] C.-H. Chang, Y. Zheng, and L. Zhang, "A retrospective and a look forward: Fifteen years of physical unclonable function advancement," *IEEE Circuits Syst. Mag.*, vol. 17, no. 3, pp. 32–62, 3rd Quart., 2017.



Christopher Cowan (M'18) is currently working toward the Ph.D. degree in electrical engineering at the Electrical and Computer Engineering Department, Maseeh College of Engineering and Computer Science, Portland State University, Portland, OR, USA.

In 2012, he joined the Asynchronous Research Center, Portland State University. He is a retired M.D.