

Использование встроенных PVT-сенсоров при разработке проектов на базе современных FinFET-технологий

С. Белоусов¹

УДК 62–791.2:621.3.049.774 | ВАК 05.11.01

Развитие технологий привело к появлению на рынке производств, обеспечивающих изготовление полупроводниковых устройств с топологическими нормами 7, 5, 6 и 3 нм. Это позволяет разработчикам применять в проектах новые решения, повышая вычислительную мощность устройств за счет увеличения количества процессорных ядер и повышения рабочих частот [1]. Повсеместное же внедрение облачных сервисов и бурный рост в области носимой электроники приводят к росту требований по энергоэффективности применяемых электронных компонентов. Кроме того, использование новых технологий требует более пристального внимания к вариативности технологического процесса, оказывающей существенное влияние на выход годных интегральных схем (ИС) [2]. Для удовлетворения всех вышеуказанных требований, а также для обеспечения надежности разрабатываемых ИС на всех этапах их применения инженеры используют современные методы диагностики и контроля, включающие внедрение специальных датчиков и контроллеров физических параметров на этапе проектирования будущих устройств.

В статье рассматриваются общие принципы реализации данного подхода на основе аппаратных компонентов платформы SLM (Silicon Lifecycle Management) от Synopsys, основанной в том числе на 15-летнем опыте компании Moortec, ставшей в 2020 году частью Synopsys [3], в области интеллектуальных сенсорных решений для контроля физических параметров СнК.

Каждый разработчик СнК вынужден учитывать большое количество факторов, связанных с технологическими особенностями полупроводникового производства: вариативность технологического процесса, нарушение временных характеристик, повышенное энергопотребление, эффекты старения. Каждый из этих факторов по-отдельности или в совокупности с другими факторами может привести к несоответствию заявленным техническим характеристикам или к полной не работоспособности полученного с фабрики устройства. Все эти нюансы были известны еще на заре развития полупроводниковой индустрии, однако только сейчас они стали определяющими при проектировании и производстве изделий, поскольку современные технологии работают на физических пределах плотности активных компонентов на кристалле.

Еще 30 лет назад для того, чтобы уменьшить потребление будущей микросхемы и вместе с тем увеличить ее быстродействие, достаточно было следовать правилам, описанным Робертом Деннардом из IBM [4], которые подвели базу под знаменитый закон Мура. Из этих правил, в частности, следует, что уменьшение размеров транзисторов приводит не только к повышению быстродействия за счет снижения влияния паразитных емкостей в классической КМОП-структуре, но и к сокращению энергопотребления благодаря уменьшению напряжения питания, необходимого для таких устройств. Используя этот простой эмпирический подход, многие компании достигли потрясающих успехов: подняли рабочие частоты до гигагерцового диапазона, создали устройства с несколькими десятками миллионов транзисторов. Однако, когда размеры канала достигли 100 нм, всё ярче стали проявлять себя короткоканальные эффекты, увеличивающие ток утечки в транзисторах и снижающие их быстродействие. Для преодоления этих эффектов стали использоваться новые материалы,

¹ Компания Synopsys, ведущий инженер по применению IP-блоков, sergey.belousov@synopsys.com.

повышающие подвижность носителей заряда и увеличивающие тем самым быстродействие транзисторов. Но у такого подхода были и побочные эффекты: внедрение новых материалов и усложнение конструкции транзисторов привело к тому, что вырос разброс физических характеристик активных элементов на одном кристалле. Два транзистора, находящиеся в непосредственной близости друг к другу, могут существенно отличаться по пороговому напряжению и, как следствие, по быстродействию. И причиной может стать не только неравномерное распределение легирующих примесей по пластине, но и, например, близкое расположение STI-изоляции к одному из транзисторов [5].

С внедрением современных технологий FinFET возникли дополнительные неопределенности, приводящие к технологическим разбросам параметров транзисторов, которые зачастую обусловлены их недостаточной зрелостью по сравнению с классической планарной технологией [6]. FinFET изначально был разработан для борьбы с влиянием короткоканальных эффектов в транзисторах с длиной канала менее 30 нм. Это достигается за счет того, что затвор в FinFET блокирует утечку заряда через канал, пока транзистор находится в выключенном состоянии. Вместе с тем FinFET в рабочем режиме оперирует с током гораздо большей плотности, чем планарные устройства, что в свою очередь может приводить к существенному локальному нагреву такой структуры [7]. Помимо этого, большой ток может вызвать падение напряжения в сетках питания, в результате чего могут возникнуть сбои в работе вычислительных блоков.

Кроме того, нужно учитывать, что в технологиях с длиной канала менее 30 нм наблюдается эффект «температурной инверсии». Обычно при увеличении температуры транзисторы работают хуже, поскольку чем ниже температура, тем реже носители заряда сталкиваются с узлами кристаллической решетки в канале. Однако, когда речь идет о современных технологиях, где напряжение питания в ячейках сравнимо с величиной порогового напряжения, транзисторы с повышением температуры начинают работать быстрее [8].

Проблемы, связанные с разбросом технологических параметров, могут существенно усложнить процесс разработки будущего устройства. Если не использовать специальные методики, способные это учитывать, инженерам приходится предполагать наихудшие условия работы изделия и исходя из этого проводить расчет и синтез цифровой схемы, то есть учитывать два граничных условия – best hot и worst cold.

Один из наиболее распространенных способов снизить потребляемую мощность устройства – применение динамического масштабирования частоты и напряжения питания. При использовании данного подхода при разработке у инженеров появляется возможность контролировать частоты и напряжения питания на каждом отдельном

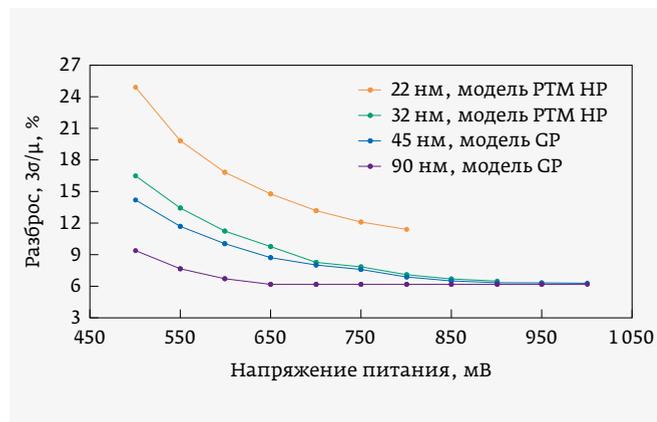


Рис. 1. Разброс скорости переключения транзисторов на пластине в зависимости от технологии

функциональном блоке ИС, обеспечивая наиболее оптимальный режим его работы и предохраняя его от избыточного старения. Но вместе с тем данный подход усложняет систему, вводя дополнительные режимы работы устройства, которые должны учитываться при разработке.

С уменьшением размеров транзисторов наблюдается все большее влияние эффекта старения на работоспособность современных ИС. Данный эффект подразумевает ухудшение быстродействия транзисторов со временем относительно начального значения – рассчитанного при проектировании и имеющего место при получении изделия с фабрики. Основными причинами старения кремния являются: температурная нестабильность смещения

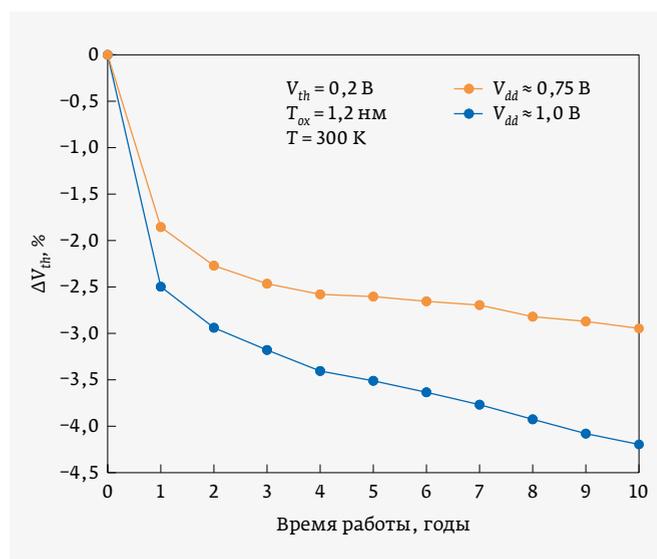


Рис. 2. Дegrадация порогового напряжения в течение времени жизни устройства вследствие отрицательного смещения напряжения затвор-исток и инжектирования горячих носителей заряда

(Bias Temperature Instability – BTI), инжектирование горячих носителей заряда (Hot Carrier Injection – HCI), многочисленные времязависимые пробои слоя подзатворного диэлектрика (Time-Dependent Dielectric Breakdown – TDDDB) и электромиграция (Electromigration – EM) [9]. Классический способ учесть старение при разработке – рассчитывать при проектировании на наихудшие условия. Однако в реальности далеко не каждое устройство, покидающее фабрику, работает в таких условиях. Следовательно, пессимизм при разработке устройства может привести к тому, что микросхема не будет работать на тех частотах, на которых она бы могла работать, если бы при ее разработке не предполагался наихудший случай. В итоге конкуренты, которые решили рискнуть, могут получить аналогичное устройство, которое в тех же условиях будет работать лучше и быстрее. Таким образом, необходим механизм, который, с одной стороны, ограничивал бы пессимизм при разработке, а с другой – гарантировал бы работу будущего устройства при любых условиях, с которыми оно может столкнуться на практике.

Отдельные области кристалла, в которых присутствует избыточный нагрев, подвержены более быстрому старению, что, в свою очередь, приводит к ошибкам в работе устройства. В качестве примера рассмотрим многоядерную систему. Операционная система для достижения максимального быстродействия часто старается использовать активное в данный момент ядро, а не пробуждать другие ядра, находящиеся в состоянии ожидания. Такая постоянная интенсивная нагрузка ядра приводит к его постепенному старению. Для того чтобы сбалансировать процесс старения модулей на кристалле, необходимы специальные динамические системы контроля и мониторинга, которые могли бы использоваться программным обеспечением для более оптимального планирования вычислительных задач, в нужные моменты переключая ядра и увеличивая тем самым срок службы устройства в целом.

Интегрированные в кристалл мониторы и датчики позволяют контролировать работу системы в режиме реального времени. Например, с их помощью можно определять текущую температуру и на основе этих данных подбирать для блока минимальное напряжение, которое бы обеспечило его работоспособность. Помимо этого, вовремя получив информацию о развитии области с повышенной температурой, операционная система может оперативно отключить активно работающее ядро и нагрузить задачами те ядра, которые не были активными до этого момента. Мониторы

и сенсоры, следящие за вариациями физических характеристик, позволяют обеспечить минимальное необходимое напряжение питания на блоке, а не максимальное расчетное, полученное при проведении анализа на наихудшие случаи работы, уменьшая тем самым общее энергопотребление системы.

СЕНСОРЫ

Существуют два основных компонента, которые необходимы для реализации встроенной системы контроля на кристалле: сенсор (монитор) и контроллер. Такая двухкомпонентная структура позволяет минимизировать площадь, необходимую для построения данной системы за счет интеграции нескольких сенсоров с одним контроллером.

Сенсоры или мониторы также подразделяются на три основных типа: датчик температуры, монитор процесса, датчик напряжения.

Датчики температуры отслеживают изменение температуры на отдельных участках кристалла. Датчики напряжения детектируют текущее падение напряжения на выбранных блоках, а мониторы процесса, построенные на базе кольцевого генератора, измеряют скорость работы активных элементов.

В самом простом варианте такой системы контроллер получает информацию от отдельной группы сенсоров и определяет характеристики выбранного блока для обеспечения заданного быстродействия. Помимо этого, он может записывать полученные значения в соответствующие регистры для дальнейшего использования программным обеспечением с целью отслеживания текущего напряжения, температуры или частоты работы блока. Данный подход отлично подойдет для небольших микросхем, однако он неприменим для крупных, выполненных с использованием современных кристалльных технологий, ИС, в которых

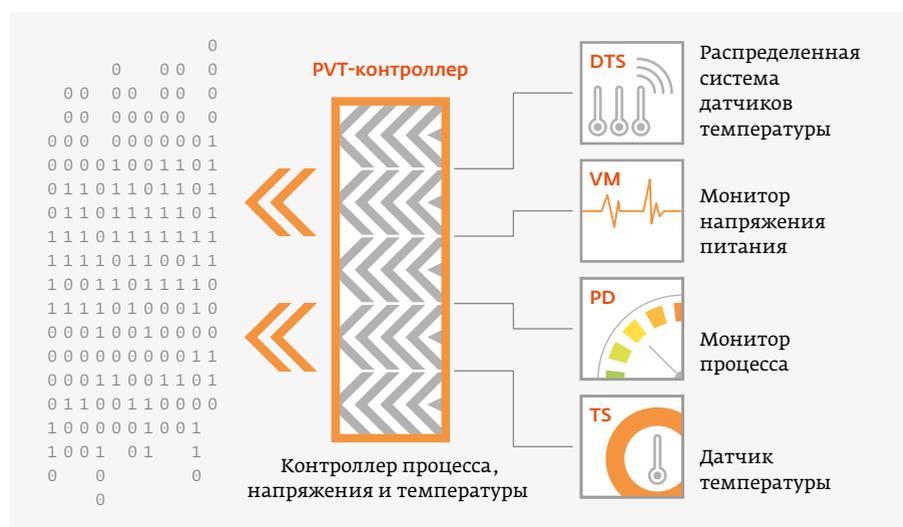


Рис. 3. Структура встроенной системы контроля на кристалле

присутствует большой разброс физических характеристик в различных областях. В таких кристаллах используют несколько подобных систем, расположенных на различных участках кристалла. Информация считывается с них центральным процессором посредством системной шины ARM APB (Advanced Peripheral Bus).

ТЕМПЕРАТУРНЫЙ ДАТЧИК

Рассеиваемая мощность с единицы площади растет при переходе на более современные технологические процессы, приводя к росту количества локальных точек нагрева и формированию на кристалле температурного градиента. Использование температурных датчиков в критически важных участках ИС на сегодняшний день является общей практикой в FinFET-технологиях.

Современные устройства, особенно те, которые используются в приложениях искусственного интеллекта, содержат в себе сотни или даже тысячи вычислительных ядер. В таких системах локальные точки нагрева или температурный градиент могут возникнуть, даже если применяются методы распределения нагрузки по отдельным ядрам на уровне ПО. Во избежание этого, необходимо балансировать нагрузки с учетом распределения температуры по кристаллу. Это невозможно без использования большого числа температурных датчиков и контроллера, который смог бы собрать и предоставить ПО данную информацию в нужном формате.

ДАТЧИК НАПРЯЖЕНИЯ

Помимо увеличения мощности, выделяемой с единицы площади кристалла, и размеров самих кристаллов, при переходе на меньшие топологические нормы происходит непрерывное уменьшение напряжений питания, увеличение числа межсоединений и сопротивления контактных площадок. На сегодняшний день в современных технологиях распространенным является напряжение питания менее 1 В.

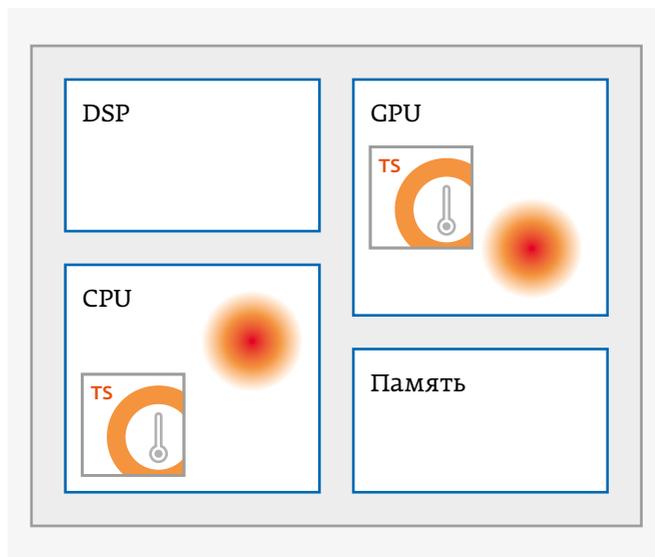


Рис. 4. Вариант расположения температурных датчиков на кристалле

Задержка на вентилях зависит от номинала напряжения питания, однако в старых технологиях его изменение на 100 мВ могло не сильно повлиять на быстродействие цифровых схем. В современных FinFET-технологиях влияние такого изменения напряжения питания на быстродействие может оказаться существенным. Для того чтобы отслеживать подобные изменения, крайне желательно иметь встроенные датчики напряжения, с помощью которых можно было бы ответить на фундаментальный вопрос: «Корректно ли работает моя цепочка питания и обеспечивает ли она необходимый уровень напряжения в ключевых функциональных блоках для их правильной работы на необходимой частоте?» Помимо этого, датчики напряжения могут отслеживать статическое падение напряжения вследствие изменения нагрузки на систему. Данные изменения происходят из-за влияния ПО,

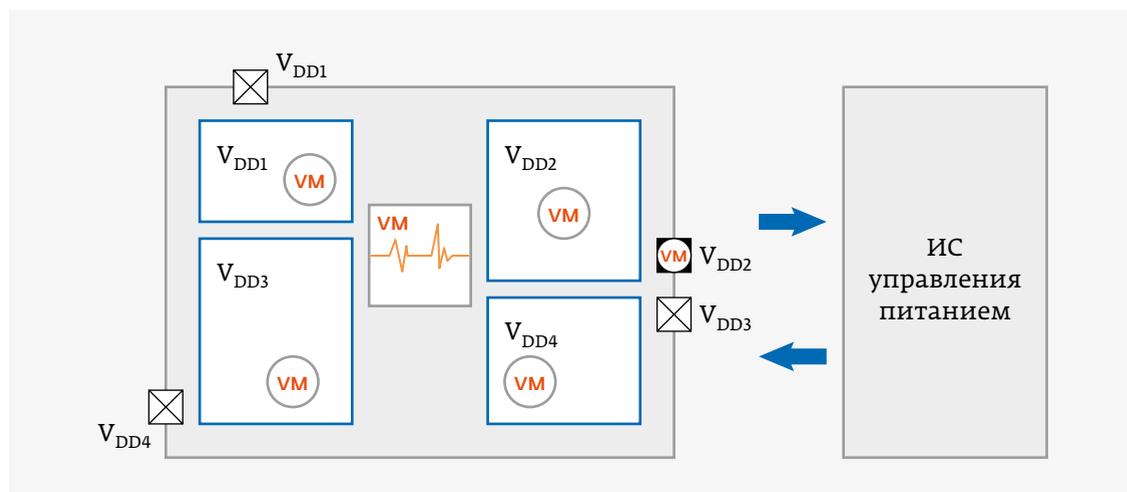


Рис. 5. Принцип измерения напряжения питания в ключевых блоках системы с помощью интегрированных в кристалл датчиков

и их сложно предусмотреть заранее. Измерение и контроль уровней напряжения на реальном кристалле являются в данном случае единственным способом учесть данные изменения.

МОНИТОРЫ ПРОЦЕССА

С переходом на технологические процессы с меньшими топологическими нормами увеличивается разброс технологических параметров на кристалле. Особенно ярко это проявилось при переходе от традиционных планарных технологий к FinFET. Специальные структуры для оценки технологических характеристик, расположенные на дорожках скрайбирования, обычно располагаются далеко от функциональных блоков. Поскольку размеры кристаллов постоянно увеличиваются, таких структур становится недостаточно для оценки реального разброса параметров в микросхеме. Именно поэтому в настоящее время разработчики стараются использовать специальные датчики, интегрированные в функциональные блоки кристалла и предназначенные для оценки физических параметров активных элементов на протяжении всего цикла применения ИС.

Другим применением таких датчиков является определение режимов работы ИС – тактовых частот и напряжений питания. Диапазон допустимых значений напряжения питания обычно определяется на фабрике в процессе тестирования микросхем в зависимости от технических характеристик полученного кристалла. Оптимальные значения напряжения обычно сохраняются в энергонезависимой памяти. В ИС, которая аттестована как SS (Slow-Slow), будет применяться повышенное напряжение питания относительно номинального, а в ИС, аттестованной как FF (Fast-Fast), – пониженное.

Масштабирование напряжения питания делится на статическое масштабирование напряжения (Static Voltage Scaling – SVS), при котором напряжение питания определяется один раз на этапе тестирования на фабрике и в дальнейшем остается неизменным, и динамическое масштабирование напряжения питания и частоты (Dynamic Voltage and Frequency Scaling – DVFS), при котором определяются две или более рабочие точки, каждая из которых включает в себя пару напряжение-частота. Впоследствии, при функционировании ИС, система может переключаться между различными режимами работы в зависимости от нагрузки в текущий момент времени. Существует и третий вариант масштабирования напряжения питания – адаптивный (Adaptive Voltage Scaling – AVS). В этом случае для достижения оптимального быстродействия и компенсации эффекта старения напряжение питания выбирается при включении питания устройства или во время его работы.

Определение рабочих частот и напряжений при проверке ИС – крайне затратная с точки зрения времени задача. Одним из способов сократить время проведения такого исследования – заранее внедрить в критически важные узлы

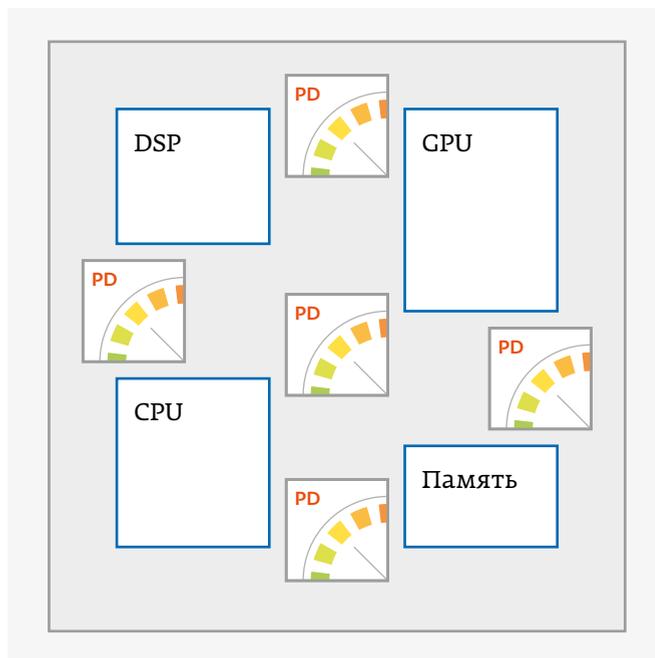


Рис. 6. Вариант расположения мониторов процесса на кристалле

кристалла датчик определения максимальной возможной скорости работы соответствующего функционального блока, представляющий собой кольцевой генератор. Поскольку тестирование кольцевого генератора – более простая задача, чем проверка всего блока, используя данный подход, можно существенно сократить время тестирования и отбраковки ИС.

Третьим вариантом использования мониторов процесса является контроль эффекта старения и адаптация к нему ИС. Примером этого могут являться рекомендации для автолюбителя заменить некоторые микросхемы во время планового технического осмотра автомобиля, поскольку система считает, что в данных ИС возможно появление сбоев из-за выработанного ресурса. Также система может использовать информацию о старении отдельных ИС для увеличения напряжения их питания с целью компенсации эффекта старения.

ЗАКЛЮЧЕНИЕ

Непрерывное развитие технологий производства интегральных схем, в особенности связанных с применением современных структур, таких как FinFET, бросает разработчикам новые вызовы. Необходимость учитывать технологический разброс параметров не только на пластине, но и на отдельных кристаллах, вероятность появления локальных точек нагрева, а также следствия применения напряжений питания низкого номинала приводят к тому, что инженерам приходится искать способы динамического отслеживания изменений характеристик устройств.

Одним из самых перспективных способов решения данной задачи является использование встроенных датчиков и контроллеров, которые позволяют реализовать на кристалле интеллектуальную сеть мониторинга параметров и предупреждения нештатных ситуаций.

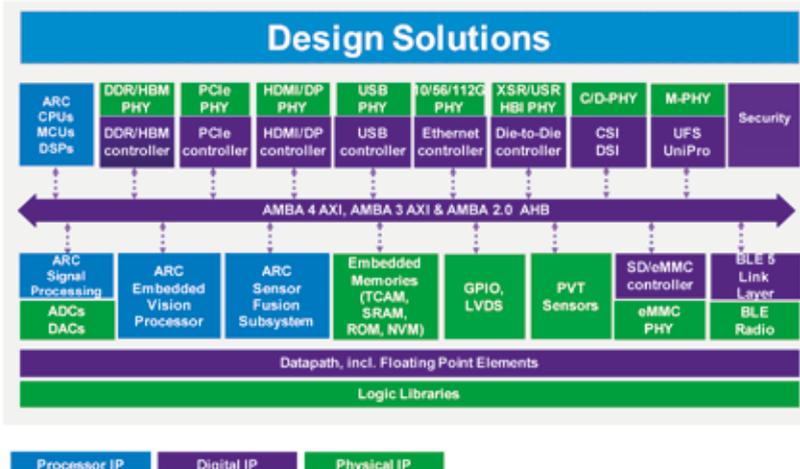
Компания Synopsys предлагает своим заказчикам широкое портфолио датчиков и контроллеров для интеграции в ИС, изготавливаемые на различных технологиях с топологическими нормами от 40 до 3 нм. Данные решения нашли применение в проектах более чем 115 заказчиков компании. Количество этих проектов превышает 320. Использование современных и надежных решений от Synopsys позволит реализовать устойчивую к эффекту старения систему, обладающую оптимальным энергопотреблением и быстродействием.

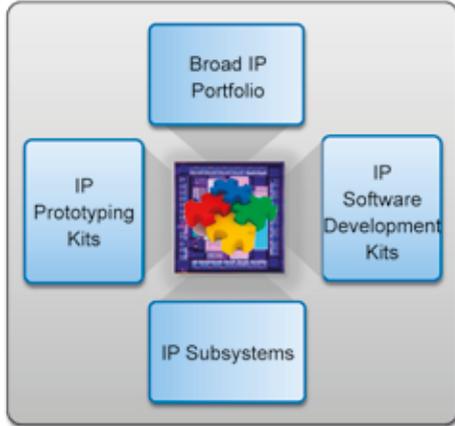
ЛИТЕРАТУРА

1. **Frumusanu A.** TSMC Details 3nm Process Technology: Full Node Scaling for 2H22 Volume Production // AnandTech. August 24, 2020. <https://www.anandtech.com/show/16024/tsmc-details-3nm-process-technology-details-full-node-scaling-for-2h22>
2. **Sperling E.** Variation Spreads At 10/7nm // Semiconductor Engineering. November 16, 2017. <https://semiengineering.com/variation-spreads-at-107nm/>
3. Synopsys Acquires In-chip Monitoring Solutions Leader Moortec // Пресс-релиз компании Synopsys. Nov. 11, 2020. <https://news.synopsys.com/2020-11-11-Synopsys-Acquires-In-chip-Monitoring-Solutions-Leader-Moortec>
4. Dennard's Law // Semiconductor Engineering Knowledge Center. https://semiengineering.com/knowledge_centers/standards-laws/laws/dennards-law/
5. **Marella S. K., Sapatnekar S. S.** The Impact of Shallow Trench Isolation Effects on Circuit Performance // 2013 IEEE/ACM International Conference on Computer-Aided Design (ICCAD). San Jose, CA, USA. Nov. 18–21, 2013.
6. **Brozek T.** Device and Process Variability // IEEE SCV-SF Electron Devices Society Seminar. Santa Clara, CA, USA. June 13, 2017.
7. **Myeong I., Song I., Kang M. J., Shin H.** Self-Heating and Electrothermal Properties of Advanced Sub-5-nm Node Nanoplate FET // IEEE Electron Device Letters. Vol. 41, No. 7. July 2020.
8. **Pratap R., Mall S., Upma** Effect of Temperature Inversion on Lower Nodes // Embedded Computing Design. October 28, 2020. <https://www.embeddedcomputing.com/technology/analog-and-power/power-semiconductors-wireless-charging/effect-of-temperature-inversion-on-lower-nodes>
9. **Kraak D., Taouil M., Hamdioui S., Weckx P., Catthoor F., Chatterjee A., Singh A., Wunderlich H.-J., Karimi N.** Device Aging: A Security and Reliability Concern // 23rd IEEE European Test Symposium (ETS'18). Bremen, Germany. May 28 – June 1, 2018.

SYNOPSYS – мировой лидер в области интерфейсных IP







Широкий набор проверенных в кремнии IP-блоков DesignWare IP:

- С наборами для прототипирования IP, готовыми к работе
- С наборами для разработки ПО для IP, позволяющими раннюю разработку ПО, отладку и тестирование
- С IP-подсистемами для сокращения издержек по интеграции, снижения рисков и времени разработки проекта

Подробности на www.synopsys.com Москва, Смоленская площадь, дом 3 Тел: +7 (495) 933 1015