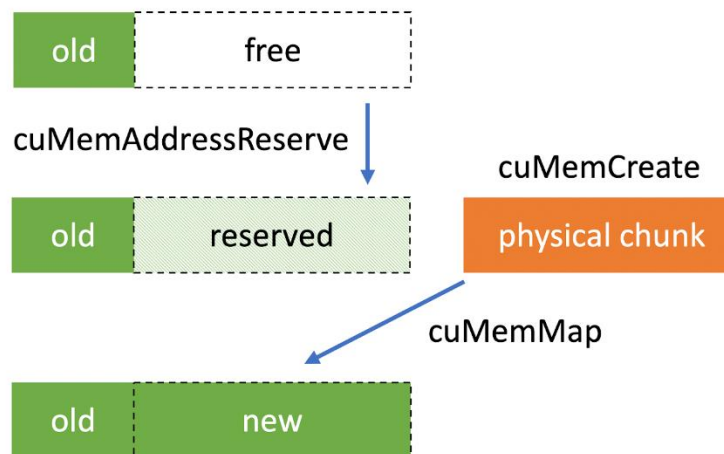


## Новое в CUDA 10.2. Низкоуровневое управление виртуальной памятью на GPU

Растет потребность в приложениях CUDA для управления памятью как можно быстрее и эффективнее. До CUDA 10.2 количество доступных разработчикам опций ограничивалось `malloc` подобными функциями, которые предоставляет CUDA.

CUDA 10.2 представляет [новый набор функций API](#) для управления виртуальной памятью. Новые функции позволяют создавать более эффективные динамические структуры данных и лучше контролировать использование памяти графическим процессором в приложениях.



**Рисунок 1.** Пример использования API-интерфейсов `cuMem*` CUDA для изменения размера буфера графического процессора.

Существует множество приложений, в которых трудно предсказать, насколько большим должно быть первоначальное распределение памяти. В этих случаях приходится выделять память с запасом, что отрицательно влияет на производительность. Иногда необходимо изменить размер под данные, хранящиеся в массивах, поддерживая при этом непрерывный диапазон адресов, который был первоначально. Фактически сейчас на CUDA добавлен функционал, подобный `void* realloc` или возможностям, имеющимся в `std::vector`.

Источник: <https://devblogs.nvidia.com/introducing-low-level-gpu-virtual-memory-management/>